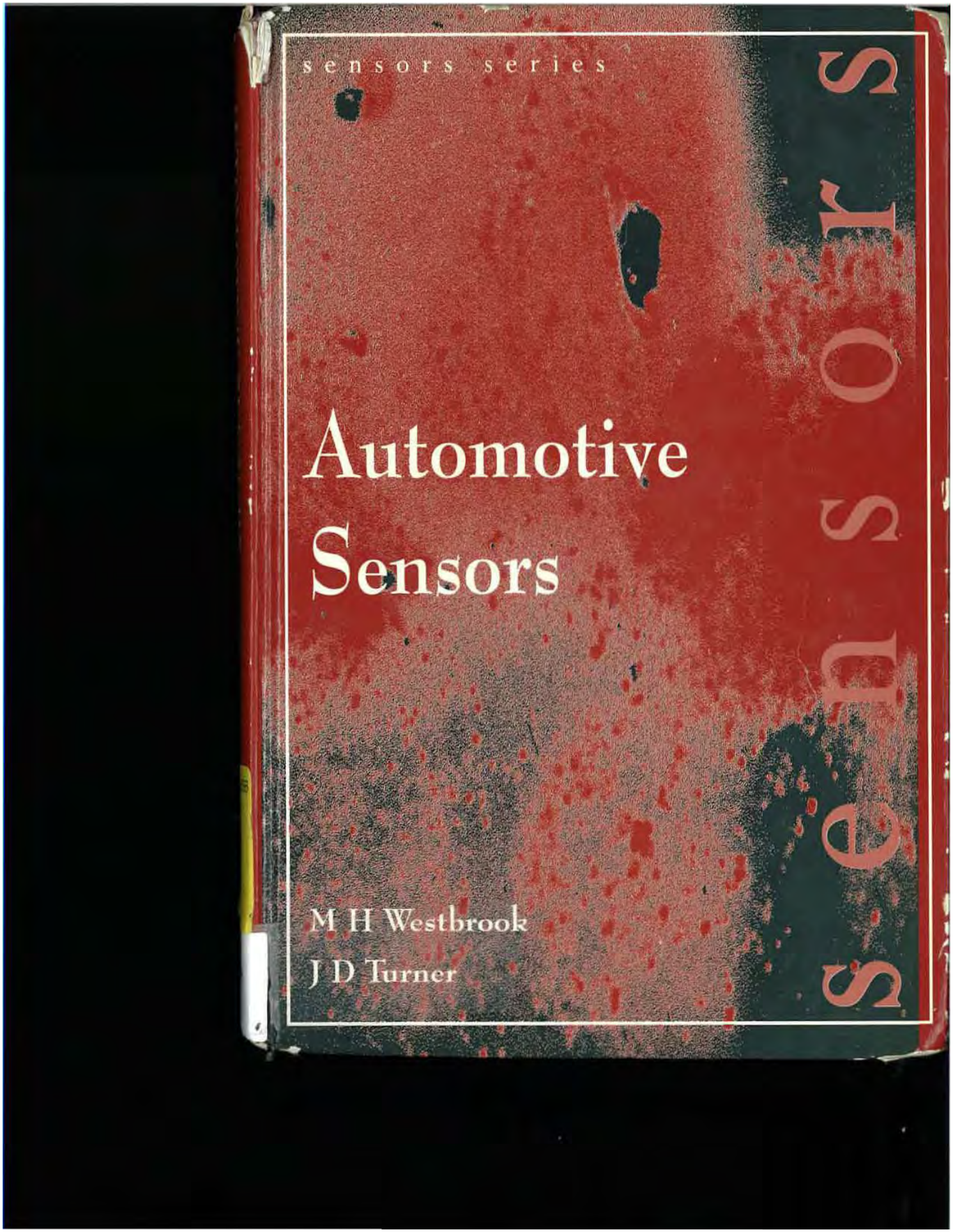

Automotive Sensors

Sensors Series

Series Editor: **B E Jones**

Other books in the series

Current Advances in Sensors
Edited by B E Jones

Solid State Gas Sensors
Edited by P T Moseley and B C Tofield

Techniques and Mechanisms in Gas Sensing
Edited by P T Moseley, J O W Norris and D E Williams

Hall Effect Devices
R S Popović

Sensors: Technology, Systems and Applications
Edited by K T V Grattan

Thin Film Resistive Sensors
Edited by P Ciureanu and S Middlehoek

Biosensors: Microelectrochemical Devices
M Lambrechts and W Sansen

Sensors VI: Technology, Systems and Applications
Edited by K T V Grattan and A T Augousti

Automotive Sensors

Professor M H Westbrook
Formerly of Ford Motor Company

and

Professor J D Turner
Mechanical Engineering Department,
University of Southampton

Institute of Physics Publishing
Bristol and Philadelphia

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 0 7503 0293 3

Library of Congress Cataloging-in-Publication Data
Westbrook, M. H. (Michael Hereward), 1926–
Automotive sensors / M.H. Westbrook and J.D. Turner.
p. cm. — (Sensors series)
Includes bibliographical references and index.
ISBN 0-7503-0293-3
1. Automotive sensors. I. Turner, J.D. II. Title.
III. Series.
TL214.C64W47 1994 94-20060
629.25'49—dc20 CIP
Series Editor: **Professor B E Jones**, Brunel University

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Techno House, Redcliffe Way, Bristol BS1 6NX, UK

US Editorial Office: Institute of Physics Publishing, The Public Ledger Building, Suite 1035, Independence Square, Philadelphia, PA 19106, USA

Typeset by Mackreth Media Services, Hemel Hempstead
Printed and bound in Great Britain by Bookcraft Ltd, Bath

Contents

Preface

The idea for this book was suggested to the authors after the 1989 Institute of Physics Sensors Conference, at which one of us had presented a paper on future developments in automotive sensors and their systems. The interest in the subject seemed high and it was certainly a critical matter for the automotive industry that low-cost, accurate sensors be developed, sensors which could be used in the mass-produced vehicles of the future with their sophisticated electronic controls on engine, transmission, suspension, braking, instrumentation and driver information systems.

Writing a book involves a significant commitment of time and effort on the part of the authors and it seemed to us important that the result should be easily accessible to a wide range of readers, both those with a general interest in the vitally important area of sensors for use in the motor vehicle as well as those involved as users, developers and researchers of automotive control systems and the sensors which make such systems practical.

We have attempted to describe in reasonable detail the whole range of sensors currently used in automotive control systems with details of their construction, operation, characteristics and method of use. We have also included a short history of vehicle sensor development since the early days of motoring, which highlights the rapidity of recent developments.

Future sensor technology is of special interest and we have made some predictions on this and how developments in conjunction with computer systems of increasing sophistication but reduced cost and size could lead paradoxically to simplified sensor and control systems with added protection against system failure and improved diagnostics.

Vehicle electronic controls would not be possible without effective, accurate, low-cost sensors, and it is with the hope of widening the understanding of these critical devices that this book has been written.

Finally, we should like to record our thanks to Daniela Hoffmann, without whose efforts the manuscript would never have reached completion.

Mike Westbrook
John Turner
November 1993

Acknowledgments

Figure 4.7 is reprinted by permission of IC Sensors Inc.

Figure 4.8 is reprinted by permission of Oxford University Press from *Physical Fluid Dynamics* (2nd edn, 1988) by Dr D J Tritton.

Figure 6.1 is reprinted with permission from *Society of Automotive Engineers Paper* 870288 © 1987 Society of Automotive Engineers, Inc.

Figures 6.2, 6.3 and 6.4 are reprinted by permission of Elsevier Sequoia from *Sensors and Actuators* A **25–27** (1991).

Figure 6.5 is reprinted by permission of the Council of the Institution of Mechanical Engineers from their publication 1985-12 (5th International Automotive Electronics Conference, 1985).

Figure 6.7 is reprinted by permission of the Institution of Electrical Engineers from their Conference Publication no 280 (6th International Automotive Electronics Conference, 1987).

Figure 6.8 is reprinted with permission from *Society of Automotive Engineers Paper* 900023 © 1990 Society of Automotive Engineers, Inc.

Figures 8.5 and 8.6 are reprinted by permission of Colvern Autosensors.

Figures 3.15, 9.5, 9.11 and 9.12 are reprinted by permission of Bruel and Kjaer.

Figures 10.3 and 10.4 are reprinted by permission of the Council of the Institution of Mechanical Engineers from their publication 1985-12 (5th International Automotive Electronics Conference, 1985).

Figure 10.6 is reprinted by permission of Elsevier Sequoia from *Sensors and Actuators* **2** (4) 1982.

Figures 10.7(*a*) and 10.9 are reprinted with permission from *Society of Automotive Engineers Paper* 850380 © 1985 Society of Automotive Engineers, Inc.

Figures 10.7(*b*) and 10.8 are reprinted with permission from the Toyota Motor Company from the paper 'Toyota lean burn engine — recent development' presented at the 13th Vienna International Motor Symposium, May 1992.

Figure 10.10 is reprinted with permission from *Society of Automotive Engineers Paper* 840141 © 1984 Society of Automotive Engineers, Inc.

Figures 13.1, 13.2 and 13.3 are reprinted by permission of W Bornhöfft, G Trenkler and VCH Verlagsgesellschaft mbH from their book *Sensors — a Comprehensive Survey* Volume 5, Chapter 5.

Figures 13.4 and 13.5 are reprinted by permission of the Editor of GEC Journals from the paper 'A new magnetic sensor technology' published in the *GEC Journal of Research* **8** (1).

Figures 13.6 and 13.7 are reprinted by permission of Dr J D Nuttall of GEC Ferranti Defence Systems Ltd from his 1991 paper 'The development of a solid-state gyroscope'.

Figure 13.8 is reprinted by permission of Dr J D Nuttall from his 1989 paper 'Review of optical gyroscopes'.

Figures 13.10, 13.11 and 13.12 are reprinted by permission of Dr R von Tomkewitsch and Siemens Plessey Controls Ltd from their paper 'Dynamic route guidance and interactive transport management with infrared beacons'.

Figure 13.16 is reprinted by permission of General Logistics plc.

Introduction

Instrumentation is a subject of fundamental importance to engineering, and it is an increasingly vital tool for automotive engineers and designers. Over the last few decades the number of automated features built into a motor vehicle has increased steadily. Sensors are essential in any automatic control system. In general, control of any automated process (such as management of the ignition timing or fuel injection in a car engine) is achieved in three stages. Sensors are used to acquire information about the process to be controlled, a microprocessor is used to decide what action should be taken, and finally actuators are required to bring about the changes required by the microprocessor. A familiar automotive example of this process is electronic engine management, in which the piston position is sensed, engine speed and load are measured, and the required ignition timing and fuel injection are determined by a microprocessor. The optimised ignition timing and fuel injection data can then be stored in a semiconductor memory as a look-up table.

The costs associated with these three stages have changed in a markedly different manner over the past 20 years. For actuators, such as the electric motors used in growing numbers by automotive designers, the price–performance ratio has improved by a factor of approximately ten. This has made possible the introduction of features such as motorised seat and mirror adjustment. For microprocessors and computing power in general the improvement is much more marked, and is estimated to be of the order of 1000. For the average sensor, however, the price–performance ratio has only improved by about a factor of three. Moreover, each kind of sensor requires its own specialised signal conditioning system, which makes the application of electronic instrumentation expensive. This lag in the progress of sensor technology hampers the process of automation and its application to motor vehicles, and it is this deficiency which is likely to be addressed within the next few years by the introduction of integrated silicon or hybrid sensors containing built-in intelligence. The implications for the automotive designer are considerable, and will probably lead to major changes in the way in which control systems are designed into a motor vehicle.

Electronic sensors have been used in road vehicles almost from their

inception. The earliest sensors were essentially switches, used to measure the crankshaft position for ignition timing purposes. An early example was Lenoir's gas engine of 1865, which used a rotary switch connecting batteries to a coil to generate the spark. Later internal combustion engines used a rigidly coupled magneto to sense the engine cycle position. In the 1920s the familiar distributor, contact-breaker and coil arrangement evolved which has persisted to the present. The cam and contact-breaker system combines the two functions of position sensing and current switching. There are always problems of wear and contact surface deterioration with such a system, and most vehicles now rely upon an electronic arrangement in which the functions of position sensing and current control have been separated, with non-contact sensing techniques being adopted.

The next form of sensor to be adopted was used for measurement of fuel level. In early vehicles the fuel tank was often placed behind the dashboard, allowing the engine to be gravity-fed. As long as the tank remained in this position mechanical sensing devices such as manometers could be used. However, the fuel tank was soon moved to its current position at the rear of the vehicle, and a pumped fuel supply adopted. This led to the introduction of electrical methods for measuring and displaying the fuel level.

Other automotive potentiometer applications followed later. In the late 1950s Bendix developed and patented their 'electro-injector' fuel injection system, in which the throttle position was sensed by a potentiometer. This was subsequently refined by Bosch to form the basis of the well known 'D-Jetronic' system, used extensively by Volkswagen and others in the 1960s.

Potentiometer sensors are now frequently used for sensing throttle and brake pedal position, steering wheel motion and suspension displacement, for automatic gearbox control, and for many other applications. Potentiometers are cheap and reasonably reliable for many applications, but suffer from the major disadvantage common to all devices which rely on a sliding contact, namely wear. While they may be adequate for, say, throttle position transduction, they may give rise to problems if used for applications such as shock absorber motion sensing. This is because a car body tends to remain close to one position relative to the wheels throughout a journey, while it undergoes large numbers of small excursions around the 'mean' position. This phenomenon is known as *dither*, and unless special precautions are taken it can cause parts of the potentiometer track to become badly worn or even destroyed locally. For this reason many manufacturers are beginning to consider alternative, non-contact forms of displacement sensor, such as inductive, magnetic, capacitive or optical types.

The need to know the speed of the vehicle arose at an early stage, and

speedometers became mandatory with the introduction of speed limits in the 1920s. The method chosen was to sense the speed of a rotating magnet driven from the gearbox, by means of the drag effect of eddy currents induced in an aluminium or copper 'cup' enclosing the magnet and working against a spring. This arrangement has survived almost unchanged for 70 years, and it is only recently that electronic systems have begun to appear in which a variable reluctance or Hall sensor is used to measure rotation rate by means of a toothed wheel.

Until the 1980s oil pressure was measured mechanically by a pressure pipe connection passed from the oil pump to the back of the dashboard. This arrangement has an unfortunate propensity to leak, which not only endangers the engine but is also injurious to the driver's trousers, as both authors can testify! This measurement is now often made by micro-machined silicon or thick-film pressure transducers. Oil pressure is not usually displayed nowadays since modern bearings are very reliable. Lubricant pressure measurements in modern vehicles are normally used simply to illuminate a dashboard warning in the event of catastrophic oil loss.

A complex electromechanical system such as a motor vehicle has to be controlled by the operator in an environment which is constantly changing, and which can generate an almost unlimited amount of input data. To successfully control a vehicle in traffic as many as possible of the mechanical functions of the vehicle need to be automated, leaving the driver free to determine the vehicle's speed and direction. The need for reliable, low-cost instrumentation in a vehicle is consequently very great. The critical quantities which have to be measured for powertrain control are ignition timing, airflow into the engine, throttle position and transmission speed, although useful supplementary information can also be gained from measurement of other quantities such as torque.

As noted above, early ignition control systems used mechanical sensing devices to control spark plug firing. Inlet manifold vacuum pressure is also measured in a mechanical system and used to infer engine load. The manifold pressure changes are used to mechanically alter the time at which a switch is closed to create the spark.

Modern timing sensors use electromagnetic, Hall effect or optical approaches to detect the motion of a projection attached to a shaft geared to the crankshaft. A certain amount of error is inevitable in these systems due to vibration and torsion (wind-up) in the geared drive. It is likely that in the future this will be eliminated by making timing measurements directly on the crankshaft. Engine load is then derived from a pressure sensor measuring inlet manifold vacuum.

Once engine load and speed have been measured the required ignition timing can be determined from a three-dimensional table relating load and speed to ignition advance. This function is implemented in a rather crude

manner by the mechanical techniques described above, but in modern vehicles the optimised data is stored in a microprocessor memory in the form of a look-up table.

The use of a three-dimensional look-up table as a means of optimising engine operation has been extended by the introduction of electronically controlled fuel injection systems. Solenoid-actuated fuel injectors are again controlled by a microprocessor, with variations in the injector opening time being used to control the amount of fuel delivered to the engine. With this system both the quantity of fuel injected and the air mass flowrate into the engine are critical. Information on airflow can be derived by measuring inlet manifold vacuum with a pressure sensor, measuring air temperature and calculating the engine swept volume per unit time. This method is known as 'speed density' measurement and has been used for some years in US and Japanese cars. A direct measurement of the airflow into the engine is, however, now preferred, and was first accomplished by the Bosch air vanemeter. This transducer was the first into service and is still widely used. It consists of a spring-loaded flap which is placed in the airstream. The flap angle is related to mass-flow rate and is transduced by a potentiometer.

An alternative form of airflow sensor which is becoming widely used is the hot-wire anemometer. Automotive versions of these were also first developed by Bosch. They have the advantage of containing no moving parts, giving increased reliability, but require correction for changes in air temperature and can be susceptible to contamination of the hot-wire surface. Further development by Hitachi, discussed in chapter 4, has produced improved, robust devices which are now in full production use.

Increasingly stringent restrictions are being placed on the amounts of polluting gas which a car exhaust can emit. To reduce these so-called exhaust emissions two approaches are adopted: first, the air–fuel ratio entering the engine is controlled to ensure complete combustion. The air–fuel ratio is inferred from measurements of the amount of oxygen in the exhaust. Secondly, a three-way catalytic converter is placed in the exhaust to remove the critical pollutants of carbon monoxide (CO), unburnt hydrocarbons (HC), and nitrogen oxides (NO_x).

Exhaust gas oxygen (EGO) sensors make use of the fact that the migration of oxygen ions across a membrane separating two gases is a function of the partial pressure of oxygen in the two gases. At the stoichiometric air–fuel ratio (when sufficient oxygen is present to burn all the fuel) the partial pressure of oxygen in the exhaust gases equals that of the atmosphere. If suitable electrodes are placed on either side of the barrier a voltage output appears only when the air–fuel ratio departs from stoichiometry.

In the future, research will undoubtedly be directed towards improving the sensors and measurement systems described in this book. The move

towards silicon will also continue, with micromachining and thick-film hybrid techniques being used to create transducer architectures on a very small scale.

The development of 'smart' sensors (in which much of the signal conditioning is carried out within the transducer housing) will provide standardised digital outputs, which are likely to be transmitted via a communications bus to the central control system. 'Smart' sensors will probably linearise their own outputs, compensate for environmental changes and include self-calibration and diagnostic functions both for themselves and for the systems to which they are applied.

Unfortunately, without special and expensive packaging silicon devices cannot cope with the highest temperatures found on a vehicle, especially around the engine, so the use of alternative semiconductor materials seems likely. One suitable candidate may be gallium arsenide, which is currently the subject of much research.

In general, an instrumentation system may be considered as falling into one of two categories. First, there are the laboratory or experimental measurement techniques used for research and development. This classification includes the instruments used to study the performance of an engineering prototype, and the laboratory devices used where high precision is required. The most important consideration faced by the designer of a measurement system intended for experimental work is its performance. In acquiring research data a high degree of repeatability, accuracy, linearity and reliability are required. The cost of this kind of system is usually of secondary importance.

The second sort of measurement system is that which forms part of a well understood device, usually a commercial product. Examples of this kind of instrument can be seen in any motor vehicle. The driver is provided with a speedometer to help control the vehicle speed, a petrol gauge to indicate when fuel is required and a milometer or other indicator to show when maintenance is needed.

For a well understood system such as a motor vehicle, a lower degree of instrument performance than that required for research is usually sufficient. For example, the accuracy of the average automotive speedometer is within a standard of −0% to +10%. However, this lack of resolution is entirely adequate to control the vehicle speed. In general the instrumentation supplied as part of a mass-produced device is of poorer quality than that used for experimental work. The principal reason for this is so that the complete system can be produced at an economic cost.

The aim of this book is to review the current state of automotive instrumentation, and to try and indicate where possible the likely course of future progress. The coverage is not restricted solely to sensors for volume application, although these predominate, but also includes a number of devices (such as torque transducers) which are at present restricted to use

in development laboratories because of their cost. The reasons for thus extending the scope of this book are twofold: first, it is primarily intended for use by those engaged in or having an interest in automotive development, and secondly, because historically the trend is for today's laboratory tool to become tomorrow's consumer electronics.

1

The Evolution of Automotive Sensors

1.1 SENSORS IN VEHICLES BEFORE THE AGE OF ELECTRONICS

Sensors have been used in vehicles since the earliest days, certainly since the time when the electric spark came to be used as the preferred method for igniting the fuel–air mixture in the cylinder of the early Otto-cycle piston engines in the late nineteenth century.

The sensing in that case was of the position of the engine cycle via the magneto drive shaft, and developed through the distributor/contact-breaker/coil of later Otto-cycle piston engines to the familiar cam and contact-breaker which has dominated ignition timing control for 60 years from the 1920s to the 1980s.

The cam and contact-breaker combined both sensing of engine cycle position and control of the charge current into the coil; in spite of problems with wear and contact surface deterioration, it was only superseded in the 1980s by electronic sensing methods and the separation from that sensing of the current control function.

Another sensor which appeared fairly early in the evolution of the car was the fuel level sensor. The use of a rod to check the amount of fuel remaining was fairly rapidly replaced by manometer tube devices which remained usable for as long as the fuel tank was under the bonnet just in front of the windscreen, but as soon as pumped fuel feed replaced gravity feed to the engine, the tank was moved to its present position at the rear and it became necessary to use electrical means for measuring fuel level and indicating it to the driver. This led to the birth of the float and potentiometer sensor with either a current sensing display device using a bimetallic device to move the needle or a balanced electromagnetic instrument display. The float and potentiometer sensor is still with us, although its accuracy leaves much to be desired in the present situation of shallow tanks of complex shape. Development has concentrated mainly on producing potentiometer tracks of variable resistance with deflection which is matched in the fuel tank characteristic. However, new methods are now becoming available and will be discussed in detail in chapter 11; but because the float and potentiometer is such a low-cost device it is

difficult to find a more accurate sensor for a similar price.

The need to know the speed of the vehicle existed as soon as the person with the red flag was sent packing, and eventually became legally required when speed limits were introduced in the 1920s. The method chosen was to sense the speed of a rotating magnet driven through a Bowden cable from the rear of the gearbox, by means of the drag effect of the eddy currents induced in a non-ferrous 'cup' enclosing the rotating magnet and connected to a sprung pointer calibrated in miles per hour. This device has survived for 70 years, and it was only during the 1980s that the Bowden cable was increasingly removed and replaced by an electromagnetic sensor on the gearbox producing electrical pulses at a rate directly related to speed, this output then being used as the input to an integrating electromagnetic speedometer display instrument.

Oil pressure level, particularly important in the days of somewhat less reliable bearings than today, was also sensed and displayed normally by a direct pressure pipe connection from the oil pump supply line to a Bourdon tube gauge giving a direct pointer indication as the tube deflected. In some cases even simpler devices were used, such as a sprung cylinder sealed into the end of the pressure pipe which protruded further from the dashboard as the oil pressure rose. It had an unfortunate propensity for leaking and dripping oil on the knee of the driver, as the authors can verify from personal experience!

Engine temperature measurement was also included, although only on the more expensive cars. The method used for this was a bulb thermocouple screwed into the engine block giving a voltage signal to a sensitive electromagnetic pointer instrument; the ubiquitous and robust thermistor has long displaced this rather delicate device.

The devices described here were the only sensors used on cars from the 1920s through to the 1960s, and it was not until the advent of electronic controls that this situation changed.

1.2 THE ADVENT OF ELECTRONIC CONTROLS

The performance of high-speed car engines had always been limited by the loss of spark energy in the ignition system at high engine speeds, owing to the limited time available to charge the coil in the short contact period available with the cam/contact-breaker ignition system. The system also suffered from a relatively short life and a continuing deterioration in timing accuracy owing to wear between the cam and cam follower on the one hand, and the build up of the deposits of arcing on the contact surfaces on the other. In fact, if these two effects had not been in opposite

directions and partially cancelled each other out the deterioration in performance with time in use would have been much more rapid than in fact it was.

It was realised during the late 1950s that if the coil could be controllably charged using a power transistor (and suitable power transistors were then becoming available), and triggered by a non-contact inductive or photoelectric sensor which had purely a timing function rather than the timing and current switching function of the contact-breaker, then the system would operate without these major disadvantages. So 'breakerless' electronic ignition was born and first used successfully in some of the racing and rally cars of the 1960s. This new ignition system was not, however, adopted for many years for mass-produced cars for a reason that is critical when any application of new technology is considered for the motor vehicle; it cost significantly more than the contact-breaker system it replaced.

So, electronic breakerless ignition was in existence as an available improvement but was not adopted outside the specialist car market until the increasing environmental pressures in the USA and, in particular, in California, on the 'smog' issue forced an agonising appraisal of what could be done to reduce the levels of carbon monoxide (CO) and the hydrocarbons (HC) emitted from the exhaust of US vehicles, ready for the first introduction of emission control regulations in California in 1966.

One of the actions quickly established was that better control of ignition timing had an important part to play in making an improvement, and that that improvement could be maintained throughout the life of the car, and certainly over the required 40 000 mile test distance, if breakerless ignition was used. It was also shown that the accurate control of the timing advance characteristic of the ignition system, which relates engine speed and load (represented by the manifold vacuum pressure level) to ignition advance angle, was critical in obtaining the lowest emission levels.

This initial introduction of electronics to ignition control triggered much more detailed studies of how engine performance, particularly in respect of exhaust emissions, could be improved by the use of electronics, and resulted in the development of electronic fuel injection in the USA, Germany and the UK. By 1967 Bosch and Lucas were in production with fuel control systems for Europe designed primarily to improve performance rather than reduce emissions. In the USA, however, in spite of the early work by Bendix, electronic fuel control was not introduced in production cars until 1975, when the increasing severity of the exhaust emission regulations made the use of catalysts in the exhaust system necessary and with this the need for precise electronic methods of controlling fuelling unavoidable. So, electronic ignition and electronic fuel injection became established very widely in the USA in the late 1970s as the only effective method, in combination with exhaust emission

regulations (which by then had added the oxides of nitrogen (NO_x) to the hydrocarbon (HC) and carbon monoxide (CO) emissions previously controlled).

1.3 STAND-ALONE AND INTEGRATED SYSTEMS

Initially separate ignition and fuelling control systems were used, but it quickly became obvious that since the same control parameters of inlet manifold pressure and engine speed were required to determine the engine operating condition, and the consequent ignition timing or fuel quantity, that these stand-alone systems could be combined to use the same sensors and to operate interactively. The subsequent addition of direct-inlet manifold mass air flow measurement instead of its derivation from the measurement of inlet manifold pressure and swept cylinder volume, did not change the move towards integrated engine control systems.

More recently, stand-alone systems for anti-lock braking control have been developed to provide the complementary facility of anti-spin control, and this has usually involved the communication of signals to the engine control system to reduce engine power output when wheel spin occurs as well as braking the offending wheel. So the communication of sensor and control signals between systems which were originally stand-alone has progressed on a rather *ad hoc* basis up to the present.

However, the advent of in-vehicle data links or multiplex systems — initially considered because of their ability to simplify increasingly complex wiring looms — offers the capability of providing sensor signals around the vehicle, and in some cases where the control time delay is acceptable, direct control signals.

This brings the possibility of fully integrated control systems much nearer and increases the need for sensors with signal processing capability built in (smart sensors), and the ability to be connected directly to an in-vehicle data link.

1.4 SMART SENSORS

The decision on how much intelligence to add to a sensor is a difficult one; as also is the definition of a 'smart' sensor. Simple impedance conversion to enable, for example, the high-impedance signal from a piezoelectric

sensor to be fed down a line to its associated electronic control unit, would certainly not qualify; but when sensors are produced with electronics which permit diagnostics, linearisation or even self-calibration, then this certainly falls within the definition. It is even more appropriate when the electronics added standardises the output, converts it to a digital data-bus format, or adds a capability to process and respond to incoming interrogation signals on the data bus.

The theme of the 1990s in automotive sensor development seems certain to move increasingly towards the world of integrated systems, where instead of being a 'stand-alone' device providing a specified output to an electronic control unit, the sensor becomes an integral part of the electronic control system, with the electronics associated specifically with processing its output and diagnosing its correct operation being found increasingly as an integral part of the sensor itself (the 'smart' sensor).

This smart sensor can be realised either by integrating the 'intelligence' onto the same chip used for the sensing function, or by using a thick- or thin-film circuit to mount a separate 'smart' circuit within the housing of the sensor. This approach is compatible with the linking of electronic systems in the vehicle by means of a data link or multiplex wiring system, since the coding and decoding functions required for such a system can, potentially, also be included in the 'intelligent' part of the sensor, so giving a compact device capable of providing standardised coded information to any electronic control unit. The control unit is then simplified because it does not now have to carry all the electronics and entry ports required for the analogue/digital conversion of a large number of sensor inputs.

The second long-term development seems likely to be the advent of smart sensors with the capability of self-calibration. Here we may expect to see low-cost smart sensors with relatively poor linearity but high repeatability being initially cycled under carefully controlled conditions through their full operating cycle, ideally *in situ* in the vehicle. The increment in sensor output per unit change of the measurand then represents the calibration of the sensor. This information is then stored in the smart sensor's memory where it can be used as the calibration curve against which future operational measurements are made. Providing the sensor has good repeatability, then wide variations in linearity and range between nominally similar sensors can be accepted, giving the opportunity for the increasing use of low-cost devices.

The third longer-term development which is likely to further change automotive control systems and the way in which sensors are used is what we call 'embedded simulation' (this will be discussed more fully in chapter 14).

The development of computer simulation of vehicle systems such as the engine, transmission, suspension etc, is currently proceeding apace, and further developments involve combining those simulations to make it

possible to represent the complete vehicle. Currently, this requires substantial computing power and memory, but it seems certain that by the turn of the century computing power of this order will be available in low-cost/small-package devices. At this point it seems probable that control systems for vehicles (and probably other self-contained real-time control systems in industry and around the home) will change in character, so that a full simulation of the system being controlled is embedded in the control system.

With the availability of this simulation it will be possible to compare the actual performance of the vehicle, as measured by suitable sensors, with the ideal as specified by the embedded simulation. This not only provides the opportunity for comprehensive adaptive feedback control of all controllable functions, but also provides a target against which the performance of all these functions can be compared, hence making full active diagnostics possible.

Another benefit of such a system is the ability to continue operating the vehicle satisfactorily even when failure of major parts of the main control system, in particular sensors, occurs. In fact, one benefit might well be the ability to dispense with many of the existing sensors, since, given information on, for example, only engine speed, torque and temperature, the simulation — particularly if it is designed to be adaptive — may well be able to specify the full operating conditions of the engine. Then, as major, slowly varying conditions such as wear, fuelling or altitude of operation change, the simulation is suitably modified to take this into account.

These three areas, integrated smart sensors, smart self-calibrating sensors and embedded simulation, seem certain to cause major changes in control systems and their associated sensors and need to be borne in mind when looking at new or improved sensor technologies for future application.

2

Sensor Applications in the Vehicle

The ability of an electronic control system to communicate effectively and accurately with the real-world system it is trying to control is proving to be the most important factor in determining the utility and reliability of that control system. We can see how important this is by looking at our own human performance as real-time control systems [1] and at how that performance is affected by any loss of operating accuracy or function of the body's sensors.

In a complex electronic/electro/mechanical system such as the modern motor vehicle, which has to be operated in intimate varying interaction with its driver and with an outside world of considerable complexity, the need for effective, accurate, reliable and low-cost sensors is very great. If we now look at the many systems to which electronics can be applied within the vehicle, and which are shown in figure 2.1, we can see that the complexity can range from the interactive control of engine and transmission to optimise economy, emissions and performance, to the

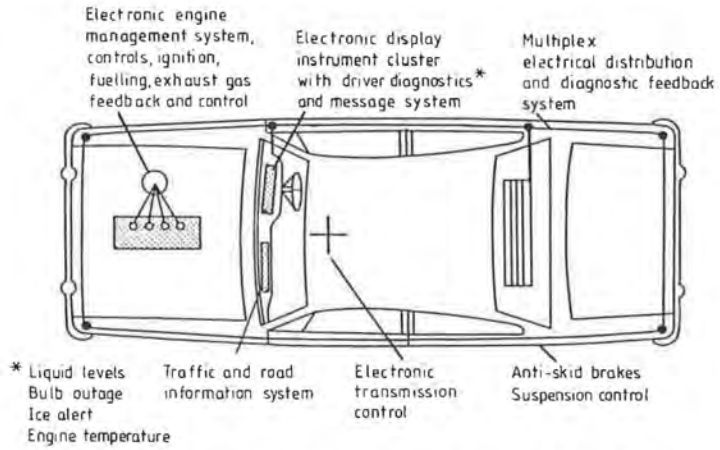

Figure 2.1 Systems to which electronics can be applied in the vehicle.

simple sensing of water temperature and level of petrol in the fuel tank — although this latter measurement is not as simple as one might think — with the result that the requirements for sensors are equally wide in scope.

Many of the sensor types mentioned in this chapter are discussed in more detail in the rest of this book.

2.1 POWERTRAIN SENSORS

In table 2.1 we have listed typical required specifications for sensors for engine and transmission ('powertrain') control. We would like to draw attention to the accuracy and temperature range over which these devices have to operate and point out that it must be remembered that they should be of minimal cost and high reliability. A fully comprehensive powertrain control system would have many of the devices listed in table 2.1, although the really critical devices are currently those which measure engine timing, inlet manifold mass air flow, manifold vacuum pressure, exhaust gas oxygen level, transmission control valve position, transmission input and output speed, and throttle and accelerator position.

2.1.1 Ignition Control

The timing sensors which are available at present are generally acceptable and normally use electromagnetic or Hall effect devices to detect the movement of a magnet or metallic projection attached to the shaft whose speed is to be measured. Almost all the cars built in recent years have electronic breakerless ignition systems fitted with such sensors. The inaccuracies in ignition timing which exist in these systems arise from mechanical vibration and torsion ('wind-up') in the geared drive to the distributor from the engine crankshaft, and will be partially overcome in future models by taking the timing from the crankshaft directly — although this then requires an additional sensor on the camshaft to determine the correct timing for each cylinder in the four-stroke cycle, and in itself suffers from wind-up due to the main engine torque and the influence of the differing operating conditions in each cylinder.

The measurement of inlet manifold vacuum pressure was certainly the first measurement made in early ignition and fuelling control systems, and it has continued to be a very important parameter. It is a relatively good measurement of engine load, since as the engine slows down under load, inlet manifold vacuum pressure (depression) moves closer to atmospheric

Table 2.1 List of powertrain sensor specifications.

Sensor/type	Proposed sensing method	Range	Accuracy (%)	Temperature operating range (°C)	Response time
Inlet manifold absolute or differential pressure sensor (petrol engines)	Piezoresistive silicon strain gauged diaphragm *or* capacitive diaphragm	0–105 kPa	±1 at 25 °C	−40 to +125	1 ms
Inlet and exhaust manifold pressure sensor (diesel engines)	As above	20–200 kPa	±3	As above	10 ms
Barometric absolute pressure sensor	As above	50–105 kPa	±3	As above	10 ms
Transmission oil pressure sensor	Differential transformer + diaphragm *or* capacitive diaphragm	0–2000 kPa	±1	−40 to +160	10 ms
Inlet manifold air temperature sensor	Metal film *or* semi-conductor film	−40 °C to +150 °C	±2 or ±5	−40 to 150	20 ms
Coolant temperature sensor	Thermistor	−40 °C to +200 °C	±2	As above	10 s
Diesel fuel temperature sensor	Thermistor	−40 °C to +200 °C	As above	−40 to +200	As above
Diesel exhaust temperature sensor	Cr/Al thermocouple	−40 °C to +750 °C	As above	−40 to +750	As above
Ambient air temperature sensor	Thermistor	−40 °C to +100 °C	As above	−40 to +100	As above
Distributor mounted timing/trigger/speed sensor/s	Hall effect *or* optical digitiser *or* eddy current/variable reluctance	Zero to maximum engine speed	±1	−40 to +125	N/A
Crankshaft mounted timing/trigger/speed sensor/s	As above			−40 to +160	N/A
Road speed sensor (speedo cable fitting)	Optical digitiser *or* reed switch *or* Hall effect	As above	±5	−40 to +125	N/A
Inlet manifold air mass flow (undirectional)	Vanemeter *or* hot wire	10 to 100 kg h^{-1} *or* 20 to 400 kg h^{-1} (two ranges)	±2 ±2	−40 to +125	35 ms for vane only others TBE but target is 1 ms
Inlet manifold air mass flow (bidirectional)	Ultrasonic *or* corona discharge *or* ion flow	±200 kg h^{-1}	±2	As above	1 ms
Accelerator pedal position sensor	Potentiometer	0–5 kΩ from min to max pedal travel	±1	−40 to +125	N/A
Throttle position sensor	Potentiometer	0–4 kΩ from closed to open throttle	±3	−40 to +125	N/A
Gear selector position sensor	Cam operated switch *or* potentiometer	8-position selection *or* 0–5 kΩ	N/A *or* ±1	−40 to +150	N/A
Gear selector hydraulic valve position sensor	Optical encoder	As above	±2	−40 to +100	N/A
EGR valve position valve position sensor	Linear displacement potentiometer	0–10 mm	±2	−40 to +125	N/A
Closed throttle/wide open throttle sensors	Microswitches	N/A	N/A	−40 to +125	N/A

Table 2.1 *cont.*

Engine knock sensor	Piezoelectric accelerometer	5 to 10 kHz 'g' range TBE	N/A	−40 to +125	Depends on resonant frequency
Engine knock + misfire sensor	Ionisation measurement in cylinder or exhaust manifold	TBE	TBE	−40 to +150 (externally) probe must meet combustion or exhaust temperatures	TBE
Exhaust gas oxygen sensor for stoichiometric operation	Zirconium dioxide ceramic with platinum surface electrodes *or* titanium discs in aluminium	Less than half one A/F ratio (used as a switch between lean & rich A/F ratios)	Not known	300 to 50 ms 850 (tip operating temperature)	
Exhaust gas oxygen sensor for lean burn operation	Zirconium dioxide oxygen pumping device with heater	14:1 to 30:1 A/F ratio	TBE	As above	

Abbreviations: EGR, exhaust gas recirculation; TBE, to be established; A/F, air–fuel ratio; N/A not applicable.

pressure and this is accentuated by the driver, who is part of the control loop, pressing the accelerator and opening the throttle further. It would really be much better, however, to measure load by measuring engine output torque, if a reliable low-cost way to do this could be found.

This manifold vacuum measurement of load has been used for controlling ignition advance from an early stage in the development of the internal combustion engine; it made use for many years of an aneroid vacuum capsule connected to the manifold to physically rotate the distributor to increase advance angle with load, while at the same time a centrifugal weight system controlled advance according to the rotational speed of the engine.

In an ignition system with electronic advance control, these functions are taken over by a pressure sensor connected to the manifold and a frequency input derived from a sensor on the engine measuring rotational speed. These pressure and speed signals then provide the input to a microprocessor which is programmed to look up the optimum advance angle from a three-dimensional table relating speed, load and advance angle and stored in memory (see figure 2.2).

By this means, significant improvements in operation and economy can be obtained. A number of designs of manifold pressure sensor have been used for this system including devices based on capacitive, inductive and potentiometric techniques, but some of the most widely used employ a silicon diaphragm with either integral silicon strain gauges [2], or a capacitive deflection sensing method [3]. In these sensors a disc of silicon is etched away to form a thin diaphragm (see figure 2.3) to which the pressure is applied, and the strain gauges are integrated onto the disc or a

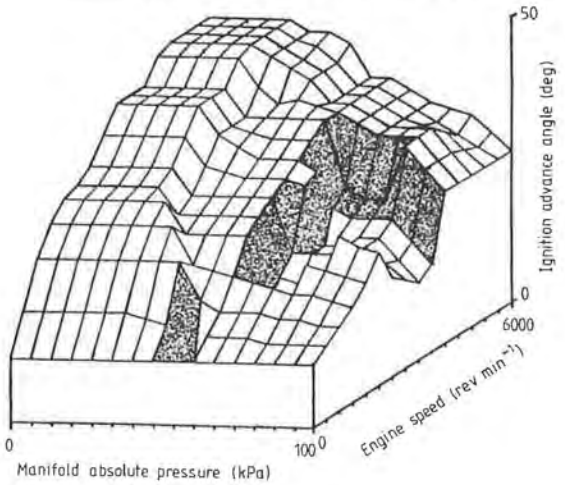

Figure 2.2 Three-dimensional table relating speed, load and advance angle before top dead centre.

capacitive plate added. This technique produces a reliable low-cost device with a good resistance to the high-temperature, high-vibration conditions under which it has to operate.

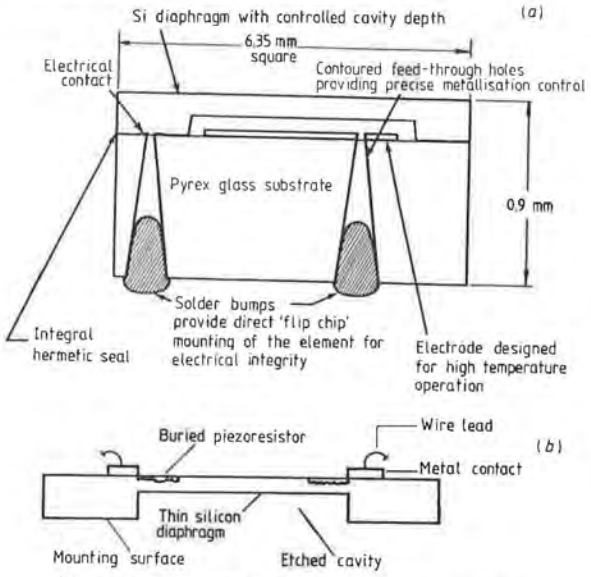

Figure 2.3 Silicon diaphragm pressure sensors.

2.1.2 Knock Sensing

When an ignition system with electronic advance control is optimised for best performance and economy, it can, under some conditions, be set sufficiently far advanced to cause a condition known as 'knocking'. Under these conditions premature high-rate combustion takes place which, because of the rapid pressure increase, can quickly cause physical damage to vulnerable structures within the combustion chamber, such as the piston crown. In an engine with conventional ignition timing the advance angle is retarded sufficiently to avoid this condition, but loss of efficiency results. For this reason it is highly desirable to operate the electronically controlled advance ignition as close to the knock limit as possible, but with the ability to retard the ignition within one or two engine cycles to a safe level. This requires a method of rapidly sensing the knock condition and up to now this has usually been achieved by the use of a piezoelectric accelerometer, known, not surprisingly, as a knock sensor (see figure 2.4). This device is usually mechanically tuned to be sensitive to the characteristic knock ringing frequency, which in normal size engines is in the region of 8 kHz, and it is placed on the engine block in a position chosen in an extensive vibration survey to give the best knock signal from all cylinders.

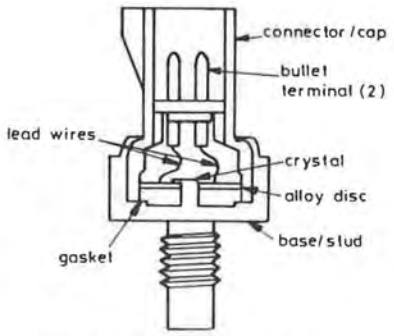

Figure 2.4 Ignition knock sensor.

It has been proposed [4] that an alternative method of obtaining this knock signal could be by measuring the ionisation current across the spark plug after normal firing. This is done by applying a small voltage to the plug sufficient to maintain an ionisation current across the plug electrodes. This current then shows a superimposed ringing signal during knock which

can be separated out and used to control ignition retard. This oscillation in ionisation current appears to be due to the variations in the gas density caused by the pressure resonance initiated by the knock condition.

2.1.3 Fuel Control

The principle of the three-dimensional look-up table as a means of describing optimised engine operation has been taken further with the electronic control of engine fuelling, usually using electromagnetic fuel injectors, where the major parameters are again load and speed, but with the fuel required being determined by injector opening time as the controlled parameter. In this system both the total quantity of fuel and the ratio of air mass to fuel injected into the engine are critical. Under these circumstances, the ideal measurement to be made is the mass air flow into the engine manifold. This can be derived by measuring inlet manifold vacuum pressure and then calculating the swept volume of the engine from the rotational speed; however, because of the compressibility of air and the significant volume of the manifold this is a very difficult calculation to make accurately, particularly under transient conditions. Many attempts have been made to develop a low-cost sensor to measure mass air flow directly, but it has always proved a difficult measurement to make. The Bosch air vanemeter, in which a pivoted vane is placed in the airstream and attached to a single potentiometer as the measuring device (see figure 2.5), was the first sensor to be used in production and has been

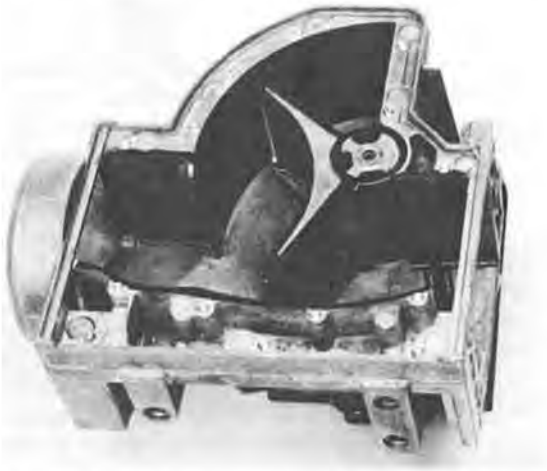

Figure 2.5 Bosch air vanemeter for manifold air flow.

(and still is) very widely used. It suffers from the fact that it measures air velocity rather than mass and therefore the signal requires processing with a signal representing air density to obtain air mass flow. It also has significant mechanical inertia under transient conditions and there is a reduction of engine efficiency caused by the flap partially blocking the air flow into the manifold, but it has proved a very valuable sensor for a whole generation of fuel-injected engines.

An alternative device which has been more recently developed, and is now also used in production, is the hot-wire anemometer (see figure 2.6), also first developed by Bosch and subsequently further developed by Hitachi. It measures mass air flow directly, is fast in response (1–2 ms), and does not significantly obstruct the manifold, but does require correction for air temperature and has some susceptibility to contamination of the hot-wire surface. This problem is dealt with by an automatic 'burn-off' operation which heats up the wire to red-heat on each occasion the vehicle is used. It also suffers from some inaccuracies under pulsed flow conditions owing to its inability to differentiate the direction of flow.

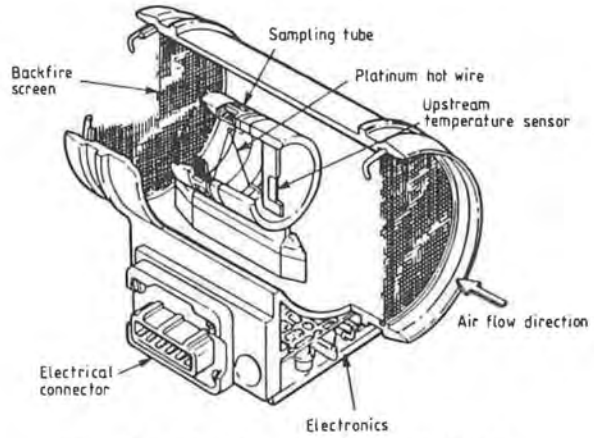

Figure 2.6 Hot-wire sensor for manifold air flow.

Other sensors for the measurement of mass air flow into the manifold have been developed, notably the vortex shedding flow meter and the ion drift flow meter, this latter having the major advantage of being able to measure direction of flow as well as its quantity. However, neither of these devices has proved entirely successful, although a vortex shedding device is used in one Japanese production vehicle.

2.1.4 Emission Control

In countries where severe restrictions on exhaust emissions exist, the

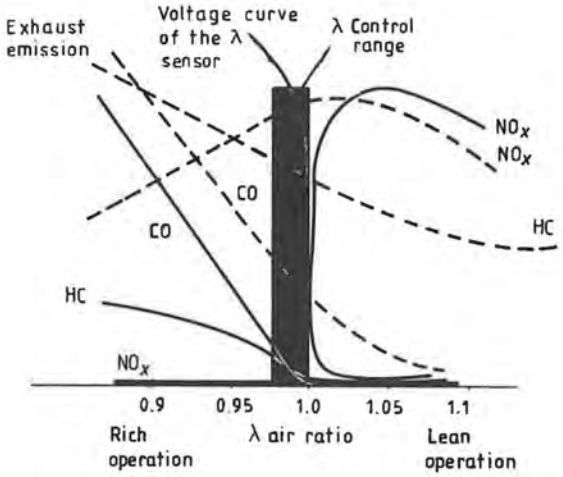

Figure 2.7 Effect of a three-way catalyst on emissions of HC, CO and NO_x above and below stoichiometry ($\lambda = 1$, λ being the ratio of air entering the engine inlet to air required for full stoichiometric operation), showing emission levels without catalyst (broken curves) and with catalyst (full curves). Also shown is the voltage curve of the EGO (λ) sensor as the air–fuel ratio passes through stoichiometry.

requirement to control the air–fuel ratio provided to the engine becomes even more critical; with current engine technology, the only way to meet these regulations is to use a so-called 'three-way catalyst' in the exhaust system to reduce the levels of the critical pollutants of carbon monoxide (CO), hydrocarbons (HC), and the oxides of nitrogen (NO_x) (see figure 2.7). For its correct operation such a catalyst requires that the air–fuel ratio fed to the engine should always be as close as possible to the optimum stoichiometric level of 14.7:1. This is only achievable by the use of a feedback control system in which this air–fuel ratio is sensed by means of an exhaust gas oxygen sensor in the engine exhaust manifold.

This sensor (see figure 2.8) makes use of the fact that the migration of oxygen ions across a suitable membrane or filter from one gas to another is only dependent on the partial pressure of oxygen in the two gases. At the stoichiometric air–fuel ratio the partial pressure of oxygen in the engine exhaust gas equals that in ambient air, so that if suitable electrodes are attached to each side of the ceramic filter used in the sensor, a positive or negative voltage is generated by the ion migration when the air–fuel ratio is below or above stoichiometry, with a rapid voltage change occurring over the transition between those two conditions (see figure 2.7). This voltage transition is ideal as a feedback signal to control the

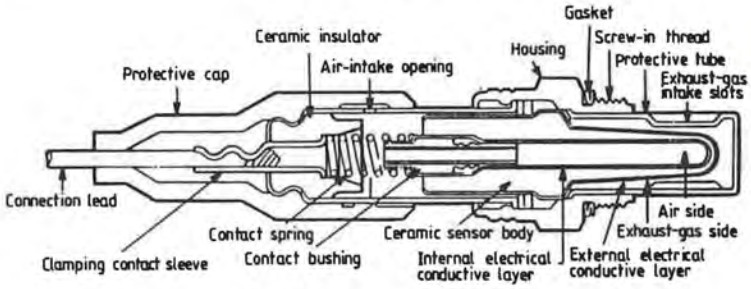

Figure 2.8 Stoichiometric exhaust gas oxygen (EGO) sensor.

amount of fuel injected and therefore the air–fuel ratio.

Unfortunately, the device requires a number of engine cycles in order to respond accurately, so that sudden changes in engine speed or load have to be compensated by open-loop adjustments based on a three-dimensional map similar to that for ignition timing but relating transient fuelling to input manifold pressure or mass air flow and engine speed.

Until recently, future developments in emission-controlled engines, particularly in Europe, were expected to be towards the use of 'lean burn' technology to obtain low levels of emissions rather than by the use of the expensive three-way catalyst with a stoichiometric engine and with consequent poor fuel economy and performance. Lean burn operation requires effective control of the engine at air–fuel ratios between 14:1 and 22:1 while still maintaining adequate driveability. If this is to be done by feedback methods then the availability of a lean burn exhaust oxygen sensor becomes very important; prototype sensors are now available using a technique shown in figure 2.9 and known as 'oxygen pumping' [5], in which the oxygen partial pressure within the sensor is positively controlled

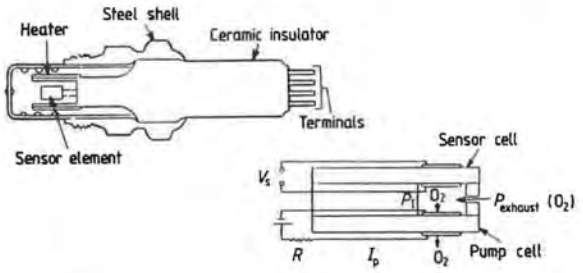

Figure 2.9 Lean burn EGO sensor. The following symbols are used in the circuit diagram: I_p = pump cell current, V_s = supply voltage, P_1 = pump cell pressure, $P_{exhaust}$ (O_2) = exhaust gas pressure.

by applying a voltage across a filter which forces oxygen ions to migrate to or from a small pumping cell. The increased partial pressure of oxygen in this cell is then compared, using a conventional oxygen sensing cell, with the increased partial pressure of oxygen in the exhaust gas stream from a lean burn engine, the resultant voltage change produced at balance being as in a conventional exhaust oxygen sensor. The current in the pumping cell required to produce this balance condition is then a measure of the increase above the stoichiometric level of the air–fuel ratio being supplied to the engine. A lean burn oxygen sensor of this type still suffers from a response time which is long compared with the rate of change of engine conditions, and the associated control system therefore continues to require open-loop compensation for transient conditions.

The recent imposition of US-level exhaust emission regulations in Europe has left lean burn unable to meet the very low emissions now required. However, proposals have been made for mixed-cycle engines which run at stoichiometry when necessary to meet instantaneous emission levels, and under lean burn conditions at other times such as in high-speed driving. These mixed-cycle engines are reputed to give excellent fuel consumption, and may bring back the need for a lean burn sensor. Such developments could also increase the need for specific exhaust gas sensors for CO, HC and NO_x levels to permit fine control of the switchover between stoichiometric and lean burn operation.

2.1.5 In-cylinder Combustion Measurement

In an ideal system it would be much better to be able to measure some meaningful property of the combustion process in the cylinder itself quickly enough to be able to control the engine operation accurately on a cycle-to-cycle basis. Three methods have been developed for meeting this requirement, the first by measuring the pressure variations in the combustion chamber, the second by detecting the arrival time of the flame front by means of an ionisation detector and the third by looking at the optical properties of the combustion by means of an optical sensor.

High-cost, laboratory-grade, instrumentation-type piezoelectric pressure sensors have been available to the engine developer for many years. The problem in applying them to production vehicles has been to reduce the price to a level which will permit the economic fitting of a pressure sensor in each combustion chamber of the engine in mass production. This has now been approached by some sensor manufacturers such as Texas Instruments [6] and a suitable piezoelectric sensor with a flush diaphragm (see figure 2.10) has been developed which is under serious consideration for production use. Good correspondence has been shown between high-cost instrumentation pressure sensors and these low-cost devices.

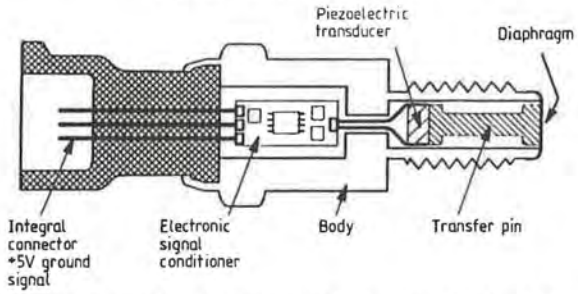

Figure 2.10 Piezoelectric flush diaphragm sensor for in-cylinder combustion pressure measurement.

(These sensors also have the capability of detecting knock conditions in the combustion chamber; this was to be expected since 'knock' is a ringing effect caused by a rapid pressure increase in the combustion chamber.)

In the case of the ionisation sensor, the idea is to detect the arrival of the flame front of the combusting air–fuel mixture on the far side of the combustion chamber, from the point of ignition. This provides two items of information. The time it takes for the flame to arrive after the firing of the spark plug gives an indication of the suitability of the ignition timing and can therefore be used as a feedback signal to correct that timing. The scatter of arrival times between successive firings gives an indication of air–fuel mixture weakness, i.e. a weak mixture produces a greater variation in arrival times. This can therefore be used as a measure of air–fuel ratio and, integrated over a number of engine cycles, can therefore be used as a feedback signal to control it. These ionisation measurements can also give an indication of knocking taking place, as can measurements of the ionisation across the spark-plug electrodes themselves after firing is complete.

The sensor itself consists of a simple insulated electrode (see figure 2.11), projecting into the combustion chamber in the appropriate location. A low voltage is applied between the electrode and the body of the engine; when the flame with its large supply of ions arrives, it reduces the resistance through the gas between the electrode and the surrounding metal and generates a large step-voltage change.

The third in-cylinder sensor is optical, and much work has been done in this area by Lucas. A quartz rod is inserted into the combustion chamber and viewed at the outer end by a photoelectric sensor. The parameters of interest are the variation in brightness and timing of the combustion; these provide similar information, including some on knock, to that supplied by the pressure sensor.

It seems probable that one of these measurements, combined with

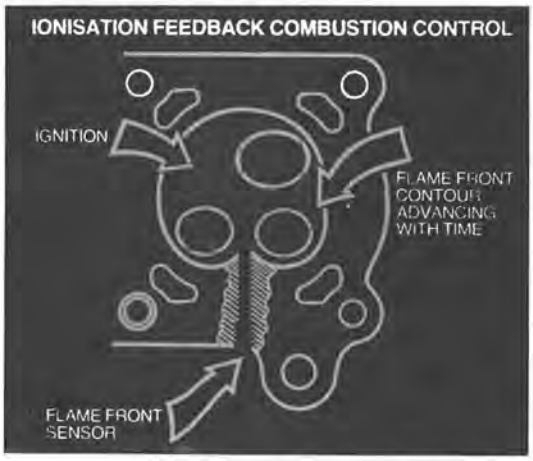

Figure 2.11 In-cylinder ionisation sensor.

feedback techniques, will be used for the adaptive and interactive control of engines as we proceed through the 1990s.

2.1.6 Speed and Torque Measurements

Finally on the engine, the measurement of speed combined with timing and of torque is of particular interest. Measurement of speed/timing has always been essential, since the control of even early engines required some events, such as ignition and valve opening, to take place at the correct time in the engine cycle. Up to the present the distributor, using the self-generating electromagnetic pick-up required for the 'breakerless' ignition systems of recent years, has been used to obtain this information. However, with the advent of distributorless ignition systems, the use of digitiser discs on the main crankshaft and electromagnetic or optical sensors matched to these is expected to increase rapidly.

Low-cost methods of measuring engine torque have not so far been available, and existing control systems have to function without this important parameter. However, new developments in low-cost torque measurement and its telemetry from the moving to the stationary parts of the vehicle have been made [7] and the measurement may become an important one for the future of engine control. This would be of particular importance if our predictions on the use of embedded simulation come to fruition, since the accurate measurement of engine torque and speed and the comparison of these measurements with a good computer simulation

should provide all the information necessary to control the engine effectively.

The consideration of engine torque leads us naturally on to the device which transmits that torque and output power to the road wheels; this is the transmission or gear box.

2.2 TRANSMISSION CONTROL

It is important first to understand that the function of the transmission is purely that of a power-matching device between power source (the engine) and load (the driving wheel/road system). With a conventional manual transmission the driver acts as the feedback loop sensing speed and load and adjusts the transmission ratio within the mechanical limitations to what he or she perceives as the best operating conditions for what he or she is trying to achieve. One of the main feedback parameters is, however, engine speed in the form of the pitch and noise level of the engine which the driver hears, but unfortunately this is rather a bad representation of engine power output; therefore the driver's gear changing, although it may optimise subjective acceleration and driveability, does not give anything like optimum operation for economy and performance. In fact if we study the torque/speed curves for a typical engine (figure 2.12) we see that optimum economy would be obtained by keeping the engine at the lowest possible speed as long as possible during acceleration and changing the gear ratio to give increased vehicle speed, only increasing engine speed to produce more power output when a wide-open throttle condition is reached. By this means the best possible economy is obtained. Good acceleration performance will require some modification to this strategy.

Operating the transmission in this way requires the use of either an automatic stepped transmission with electronic control and smooth changes and sufficient steps [8] to give an adequate range of gear ratios, or an electronically controlled continuously variable transmission (CVT) with an adequately wide ratio.

The transmission ideally needs to be variable in ratio throughout the operation range and, in a fully integrated powertrain (engine and transmission) control system, to be controlled interactively with the engine. The sensors required, therefore, are for engine speed, transmission output speed or vehicle speed, and assuming that the transmission ratio is hydraulically controlled, hydraulic valve position and hydraulic oil pressure. The first of these is normally already available from the existing sensor on the engine for engine speed, vehicle speed then being available

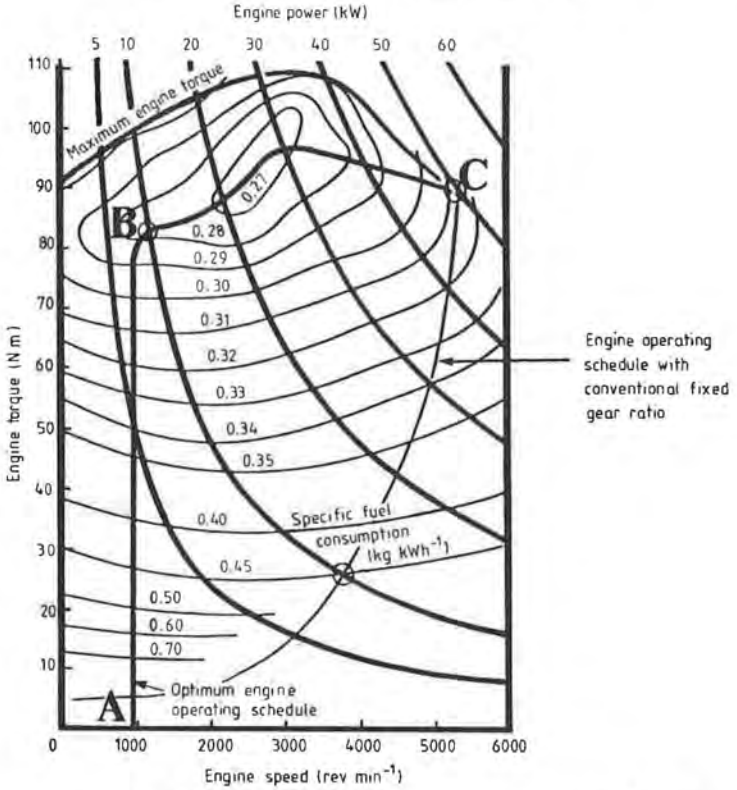

Figure 2.12 Engine torque/speed curves with specific fuel consumption islands (circles). Curve AC shows engine operating schedule with a conventional fixed gear ratio, while curve ABC shows engine operating schedule with optimised gearing.

from the inductive sensor used on the transmission to provide a signal for the electronic speedometer. In the case of the position of the hydraulic control valve, if the actuator is a stepping motor a separate sensor may not be required; if a sensor is required, an encoder disc and electromagnetic or optical sensor is attached to the end of the valve shaft. For hydraulic system pressure, for which accurate control is critical for some types of transmission, diaphragm-operated linear variable differential transformers (LVDT) have been used in experimental systems. In future systems a pressure sensor derived from etched-silicon technology seems likely to be the best device from the standpoint of cost and reliability.

Another major system which is about to appear in production cars is that of suspension control.

2.3 SUSPENSION CONTROL

There are two major types of suspension control. In one, known as active suspension and developed particularly by Lotus, the vehicle springs are replaced or substantially augmented by hydraulic jacks, these jacks then being electronically controlled to maintain the body of the vehicle as level as possible irrespective of the road surface being traversed by the vehicle. In the other (and much cheaper) system, known as adaptive damping, the stiffness of the shock absorbers is controlled either continuously or in a series of steps by an electronic control system which can react to bumps in the road, turning movements of the steering, and speed changes. This gives a soft ride for lower-speed straight-line driving, with a progressively harder ride for improved handling during high-speed driving and lane-change or turning movements.

The diagram in figure 2.13 shows the layout of an experimental adaptive-damping system and the most noticeable feature is the large number of sensors required (a similar range of sensors would be required for a full active-suspension system).

To take them in turn, the gearbox speed sensor is the same device as already described and is used to provide information about the vehicle's speed for the transmission control system; the throttle-position sensor is also already used in the engine control system. Omitting the switch inputs, a unique sensor is required for steering-wheel position and velocity to inform the control system that the driver has initiated a turn. The device currently uses a digitiser disc with an optical sensor mounted on the steering wheel, the digital information obtained being processed electronically to give information on both position and rotational velocity.

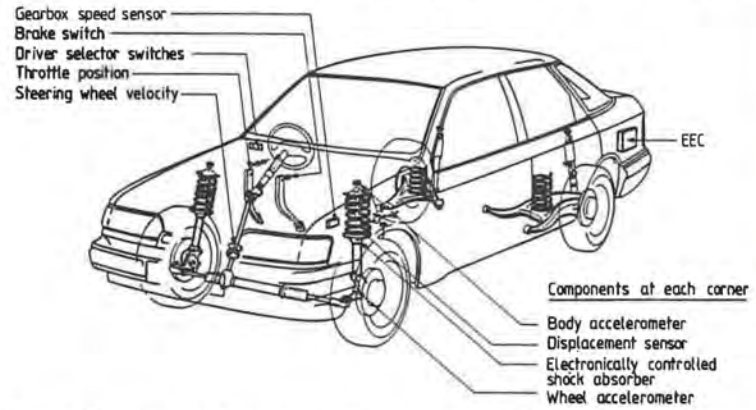

Figure 2.13 Layout of adaptive-damping system. (EEC indicates the electronic engine controller module.)

The body accelerometer, which is likely to be constructed using piezoelectric or etched-silicon devices combined with thick-film technology in future production, provides information on vertical body acceleration at each corner of the vehicle and may, in some systems, be augmented by a further lateral accelerometer at the centre of the vehicle. The wheel-to-body displacement sensor, because of the large movement to be measured, is likely to be of an inductive type in which the shock-absorber piston moves vertically inside a coil of wire. Finally, the wheel accelerometer, which will again probably be made in future production using etched-silicon technology, measures vertical wheel acceleration.

It can be seen that these diverse sensors are required to provide sufficient input information to actively and effectively control the vehicle suspension in response to all the varied inputs from road to vehicle. In the longer term, embedded simulation might, however, allow a reduction to only one major vertical and lateral acceleration measurement together with the steering wheel input to give predictive information.

2.4 ANTI-LOCK AND ANTI-SPIN SYSTEMS

We shall now look at one other important system within the vehicle, part of which (anti-lock braking) is already in production, and the other part of which (anti-spin traction control) will soon be seen in vehicles on the road.

Anti-lock brakes work by sensing whether, on braking, one wheel starts to slow down towards a locked condition faster than the other wheel. If this happens, braking pressure is taken off that wheel which is then allowed to speed up until it is again rotating at the speed of the other wheels, braking then being reapplied.

Anti-spin or traction control is the exact reverse of anti-lock. It is intended to retain traction when a wheel spins on a slippery surface when power is applied. In this case, as the wheel starts to speed up towards spin, braking is applied to that particular wheel and engine power is reduced by backing off the (electronic) engine throttle through the engine control system.

It is clear that if a vehicle already has both electronic engine control with electronic throttle and electronic anti-lock, then the provision of anti-spin can be made at a very small additional cost (see figure 2.14). It is a classic example of the benefits of integrating different electronic control systems around the vehicle. The critical sensors involved are those for sensing wheel rotation speed. Those currently used in production are electromagnetic, using variable reluctance or Hall effect devices, detecting

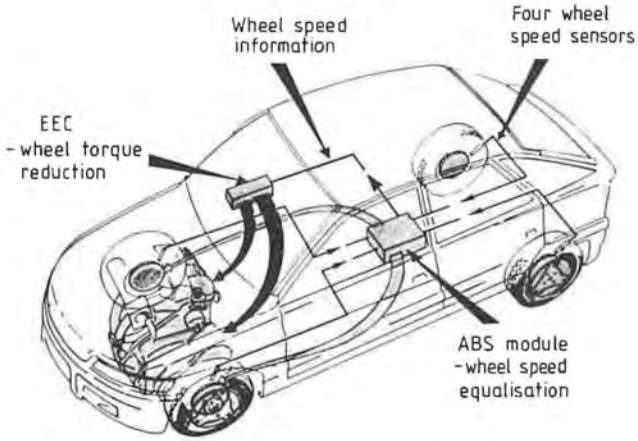

Figure 2.14 Integrated anti-lock and anti-spin system.

the teeth on a steel digitiser ring mounted on each wheel. Because of the exposed position and the risk of contamination from salt, water and mud, inductive or magnetic sensing seems essential, so that these sensors are unlikely to change dramatically in the near future; however, they are likely to be some of the first sensors to repay the addition of intelligence at the sensor, since what is of real interest is not transmitting the complete train of relatively high-frequency pulses from each wheel through the vehicle to the central electronic control unit, with the risk of electromagnetic interference always present in the severe environment of the vehicle, but transmitting the change in speed which takes place, and this is essentially a very low-frequency event in electronic terms. This could be easily handled by a suitable smart sensor and would potentially be much more suitable for transmitting over the multiplexed data busses that we shall see in cars in the near future.

2.5 DRIVER INFORMATION AND DIAGNOSTICS

Moving away from major systems of engine, transmission and suspension, we come to a wide range of electronic systems associated with other vehicle functions. One function most visible to the driver is that of information and diagnostics. In this category come two measurements which have been made and displayed to the driver almost since cars were first driven in significant numbers on the road. These are the quantity of

fuel remaining in the fuel tank and engine water jacket temperature. Measuring water temperature is a straightforward operation, usually done by a simple thermistor with some linearisation using resistors. Design requirements are usually for the temperature to be displayed on a gauge, although a simple red warning light display would be perfectly adequate and create less of the frequent confusion and concern which exists among drivers when a reading, which is perfectly acceptable from an engine operating viewpoint, is away from the mid-point on the gauge.

2.5.1 Fuel Quantity

In the case of quantity of fuel remaining a gauge reading of good accuracy is clearly important, particularly as the measurement is often used today as the basis for a trip computer calculation and display of the miles remaining to be travelled before the tank is empty.

Finding a low-cost accurate method of making this measurement is surprisingly difficult, and this difficulty is currently being compounded by the trend towards wider tanks of shallower depth and complex cross section. Currently, the time-honoured float and arm controlling an immersed potentiometer slider is used, albeit with a much improved composite track with a characteristic matched to the tank shape. Nevertheless, as most people reading this book will no doubt have experienced, the accuracy of the present fuel gauges leaves much to be desired.

Many alternative measuring devices have been tried, including the measurement of the capacitance change between two narrow plates dipping into the fuel and the measurement of the change in optical transmission between two light guides projecting into the fuel.

The most recent, and probably the most promising, method is one which uses a vibrating rod, whose natural frequency is changed as the level of immersion in the fuel changes. This device, which is described in chapter 11, is of greater accuracy than the existing float and potentiometer, has an inherently digital output and is potentially of low cost.

2.5.2 Fuel Flow

Another seemingly intractable problem is that of measuring fuel flow in a vehicle. The demand for the measurement of 'instantaneous' and trip fuel economy, usually displayed on a trip computer, is steadily increasing but our ability to measure fuel flow accurately on carburetted engines remains poor. The major problem is the very large fuel flow 'turn down' ratio from $50\,l\,h^{-1}$ to $1\,l\,h^{-1}$, with the result that a device which works well at the high

flow rate is almost useless at the low rate. The device usually used is a small low-cost turbine with a inductive/magnetic sensor which produces a simple digital signal where frequency represents flow rate; however, at very low flow rates turbine meters become inaccurate owing to the small amount of energy available to turn them. The increasingly widespread use of fuel injection, largely as a result of stricter exhaust emission regulations, is, however, causing this problem to disappear, since it is possible to obtain accurate fuel flow information by measuring the pulse length used to control the injector opening period, and this is directly related to fuel flow.

2.5.3 Oil Level

A further driver information measurement that one would like to make more effective is that of engine oil level. Current devices usually operate on the hot-wire principle using a thermistor as the heat generating and sensing device and detecting its heating, and therefore change in resistance, when the oil level drops and it is exposed to the air. The problem here is that this only works when the engine is stopped, and in fact has been standing for some time, so that the oil can drain back into the sump from the other parts of the engine. This has the result that the oil level warning currently available on some cars only detects the oil level at first switching on the engine, and even then measures incorrectly if the engine has just been run for a short time, and does not measure at all during running; so that all the oil may be lost during running without the low-level oil warning being activated.

The problem is that in a running engine the oil and air form an emulsion and so far no effective way has been found of measuring the oil content of this emulsion while also making allowance for the oil being pumped around the other parts of the engine. It does, however, seem possible that a smart sensor of the future could be developed to do this.

2.6 OTHER SYSTEMS

There are, of course, many sensors already being used, and which will be used in vehicles of the future, other than those that we have discussed in this chapter. But we think that this is a reasonable overview and should give the reader a feel for the wide range of sensors required in the vehicle. However, one area where many advances will be seen over the next decade is that in which the driver receives information and interacts with

the world outside the vehicle. Within this category are a wide range of driver information systems and, in particular, vehicle navigation and headway control, which the Europe-wide PROMETHEUS and DRIVE research programmes have been set up to investigate.

2.6.1 Vehicle Navigation

Vehicle navigation is going to become of considerable importance during the coming decade. The increasing congestion in our cities is already putting great pressure on local and national authorities to make more effective and efficient use of our urban roads. Vehicle navigation provides — with the right system — a means of doing this by providing directional information to drivers which interactively takes into account traffic hold-ups and congestion. The Autoguide/ALI-SCOUT system currently being operated as a pilot scheme in London and Berlin is an example, and is interesting from the sensor point of view, because it requires communication between the vehicle and a central computer via receiver/transmitters at the roadside and on the car; the sensing system used in the receivers involved is critical to the success of the system.

The inductive loops controlling traffic lights have been used in conjunction with suitable inductive receiver/transmitters on the car, as have high-frequency short-range radio receivers. The preferred system for Autoguide/ALI-SCOUT, however, uses an infrared receiver/transmitter mounted on a convenient traffic light standard at each major junction (see figure 2.15) and a similar receiver/transmitter mounted on the inside of the car windscreen behind the central rear-view mirror. The infrared sensor and radiator used has proved to be highly effective for this vital function, and we can expect to see these in widespread use within five years.

Also with all in-vehicle navigation systems there is a requirement for a basic dead-reckoning capability which requires the use of sensors for the measurement of distance travelled and of geographical heading direction. The first of these is usually provided by wheel rotation sensors on the non-driven wheels of the vehicle and the second by means of a magnetometer sensitive to the earth's magnetic field mounted on the vehicle which can be electronically recalibrated at frequent intervals. Differential wheel speed sensing of the rear wheels is also used to measure turning rates on corners. By these means dead-reckoning of vehicle movement is possible over short distances, but must be corrected by either receiving information on the vehicle's position from external roadside transmitters or by matching against a stored map in the vehicle.

By means of such 'navigation sensors' infrastructure-based navigation systems such as Siemens' Autoguide/ALI-SCOUT (now known as APPLE in the UK) and autonomous systems such as Philips' CARIN and Bosch's

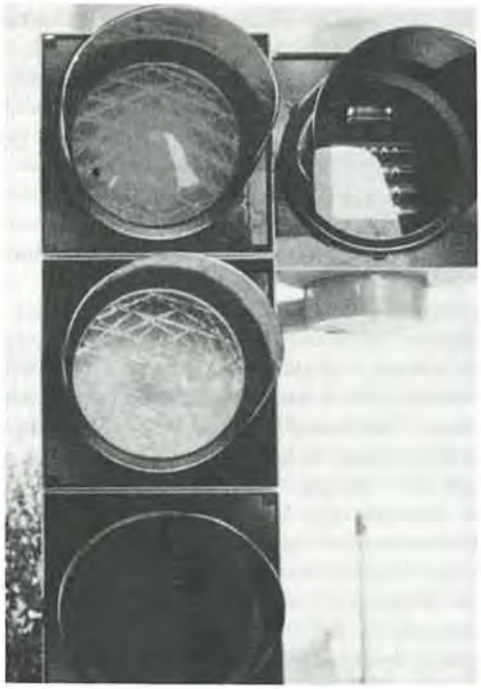

Figure 2.15 Infrared transmitter/receiver for vehicle navigation, shown at top right of traffic light cluster.

Travelpilot may be made to give accurate and effective vehicle navigation and route guidance.

The availability of low-cost satellite location systems is expected to result in rapid ongoing development in this technology.

Vehicle navigation systems are discussed in greater detail in chapter 13.

2.6.2 Headway Control

A major issue for traffic control in the future is how the driver may be provided with information on the distance and rate of approach to the vehicle in front. Vehicle radar has been investigated in many laboratories over many years, but so far the problems caused by false warnings have not been fully overcome. If they can be overcome, however, the use of a recently reported GaAs chip, which has all the dipoles and receive/transmit modules for a phased x-band radar on the single 100 mm

GaAs wafer, may well make such a system a viable proposition by reducing the size and cost of vehicle radar to an acceptable level. Once again the receiving/sensing method used is critical for success.

Other sensing systems for visibility, the presence of passing vehicles and road-surface conditions such as the presence of pot-holes are all part of the need to provide the driver and control systems with more information to make the vehicle operate more efficiently and effectively within its total environment.

2.7 SUMMARY

The important issues for the next few years are not so much how we can use sensors based on new technology, even those which are 'smart' and self-calibrating, but how we can effectively transmit information from the sensors around the vehicle so that it is only necessary to make a particular measurement with one device. If this problem is solved all the necessary control systems can then be supplied with the required data from just a single sensor for each parameter. Engine and vehicle speeds would be typical of sensed parameters which need to be widely available because they are required by a number of different control systems.

The methods by which the vehicle communicates with the outside world are also going to become vitally important, and sensors have a major role to play there.

The most distant step is the introduction of what we called, in chapter 1, the technique of 'embedded simulation', which offers the possibility of a dramatic reduction in the number of sensors required, combined with a greatly enhanced ability to adapt the system to its operating conditions and to diagnose faults.

REFERENCES

[1] Westbrook M H 1984 The human body and its transducers: a real-time real-life computing and control system *Proc. IEE* **131A** 10–16

[2] Oakes J A 1981 A pressure sensor for automotive application *Proc. 3rd Int. Conf. on Automotive Electronics* (London: IMechE) pp 143–9

[3] Behr M E, Bauer C F and Giachino J M 1981 Miniature silicon capacitance, absolute pressure sensor *Proc. 3rd Int. Conf. on Automotive Electronics* (London: IMechE) pp 255–60

[4] Blauhaut R B, Horton M J and Wilkinson A C N 1983 A knock detection system using spark plug ionization current *4th Int. Conf. on Automotive Electronics* (London: IEE) (late paper not included in the proceedings)

[5] Fleming W J 1982 Engine sensor: state of the art *Society of Automotive Engineers Paper* 820904

[6] Anastasia C M and Pestana G W 1987 A cylinder pressure sensor for closed loop engine control *Society of Automotive Engineers Paper* 870288

[7] Turner J D 1988 Development of a rotating-shaft torque sensor for automotive applications *Proc. IEE* **135** 334–8

[8] Richardson R M, Main J J and Lindre J 1983 A five-speed microprocessor controlled economy transmission *Proc. 4th Int. Conf. on Automotive Electronics* (London: IEE) pp 32–8

3

Pressure Sensors

When a fluid (liquid or gas) comes into contact with a surface it produces a force perpendicular to the surface. The force per unit area is called the pressure. The SI unit of pressure is the pascal (Pa). 1 pascal is equivalent to 1 newton per m^2 ($N\,m^{-2}$). Other units still in use in automotive engineering include pounds per square inch (psi or $lb\,in^{-2}$), atmospheres (atm), millimetres of mercury (mmHg) and bars. Conversion factors between the various units of pressure are given below:

1 Pa	=	$1\,N\,m^{-2}$	=	$1.45 \times 10^{-4}\,lb\,in^{-2}$
$1\,lb\,in^{-2}$	=	$6895\,N\,m^{-2}$	=	$0.0703\,kg\,cm^{-2}$
1 atm	=	$101\,325\,N\,m^{-2}$	=	$14.7\,lb\,in^{-2}$
1 bar	=	$100\,000\,N\,m^{-2}$	=	$14.5\,lb\,in^{-2}$
1 mmHg	=	$133.3\,N\,m^{-2}$	=	$1.93 \times 10^{-2}\,lb\,in^{-2}$.

Pressure sensors are important for a number of automotive applications, such as measuring inlet manifold and cylinder pressures, sound measurement (when the pressure sensor is called a microphone), or for monitoring lubrication and hydraulic systems. A number of other applications are also emerging, such as monitoring tyre pressure.

Pressure measurements may be divided into three categories, namely *absolute pressure, gauge pressure* and *differential pressure*. The *absolute pressure* is the difference between the pressure at a particular point in a fluid and the absolute zero of pressure, i.e. a complete vacuum. A mercury barometer is an example of an absolute pressure sensor, since the height of the column of mercury measures the difference between atmospheric pressure and the 'zero' pressure of the Torricellian vacuum that exists above the column of mercury.

If a pressure sensor measures the difference between an unknown pressure and local atmospheric pressure, the measurement is known as *gauge pressure*. The circular dial-type pressure gauges fitted to steam boilers normally indicate gauge pressure. This convention presumably arose because the amount of work obtained from a steam engine is a

function of the amount by which the supply steam pressure exceeds that of the atmosphere.

If the pressure transducer measures the difference between two unknown pressures, neither of which is atmospheric, then the measurement is known as *differential pressure*. The difference between absolute, gauge and differential pressure measurement is shown in figure 3.1 using mercury manometers.

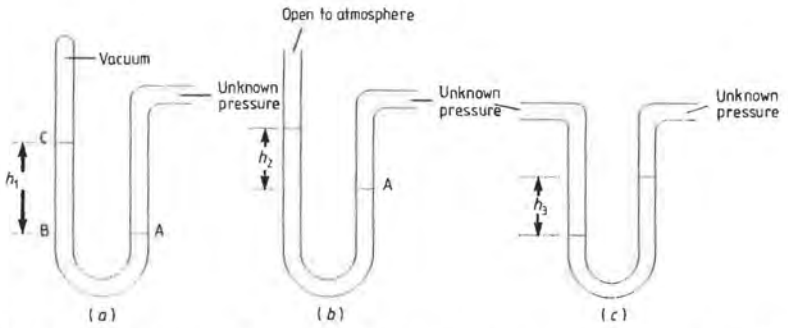

Figure 3.1 Comparison of types of pressure measurements: (*a*) absolute; (*b*) gauge; (*c*) differential.

There are three fundamental means by which a pressure may be measured. The simplest approach involves balancing the unknown pressure against the pressure produced by a column of liquid of known density. Instruments using this principle are called *manometers*. The analysis of a manometer is straightforward. Consider a simple U-tube containing a liquid of density ρ, as shown in figure 3.1. If we consider figure 3.1(*a*), the points A and B are at the same horizontal level when the device is in equilibrium. The liquid at C stands at a height h_1 above B. Then the pressure at A is P_A, where

$$\begin{aligned} P_A &= P_B \ (P_B \text{ is the pressure at B}) \\ &= \text{pressure due to column of liquid BC} \\ &= h_1\rho. \end{aligned}$$

In the case of figure 3.1(*b*), the unknown pressure at A is

$$P_A = h_2\rho + \text{atmospheric pressure}$$

since the manometer is open to atmospheric pressure at one side, rather than being sealed as was the case in figure 3.1(*a*). The analysis is similar for figure 3.1(*c*), except that in this case the atmospheric pressure has been replaced by a second unknown pressure.

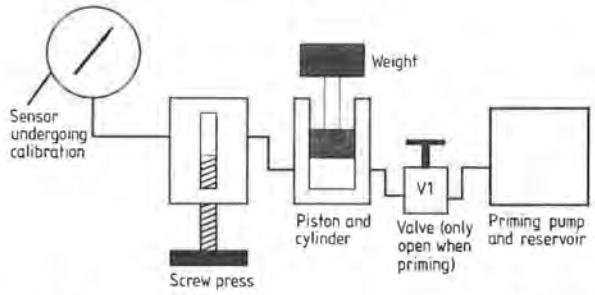

Figure 3.2 Dead-weight pressure measurement system for sensor calibration.

The second method of pressure measurement involves allowing the unknown pressure to act on a known area. The resulting force is measured either directly or indirectly. Devices of this type are called *dead-weight testers*, and they are normally only used for calibrating other forms of pressure sensor. The basic arrangement is shown in figure 3.2. It consists of a priming pump and reservoir, an isolating valve, a weighted piston, a screw press and the pressure sensor under test. The sequence of events is as follows. First, the screw press is extended to its zero position. Weights representing the desired pressure are applied to the piston. The priming pump is operated to pressurise the system, and the priming valve (V1 on the diagram) is closed. The screw press is then adjusted until the pressure in the system has increased sufficiently to just raise the piston off its end-stops. Neglecting any friction forces, if the pressure on the piston is $P\,\mathrm{N\,m^{-2}}$, and its area is $A\,\mathrm{m^2}$, then the resulting force on the piston is PA N, which will support a weight $W = PA$ N. The accuracy of the calibration depends on the precision with which the piston and its associated cylinder are made, and on the degree to which friction has been eliminated from the system.

In the third approach the unknown pressure is allowed to act on an elastic structure of known area and properties. Most commercial pressure sensors adopt this approach, and this includes the majority of those used in automotive engineering. The resulting stress, strain or deflection is measured in a variety of ways.

The most common form of pressure sensor is probably a diaphragm fitted with strain gauges, although capacitive systems are used almost as frequently. Figure 3.3 shows diaphragm-type pressure sensors configured to measure differential, gauge and absolute pressure.

Section 3.1 contains a general introduction to the design theory associated with elastic pressure sensors. The emphasis is placed on diaphragm and membrane types since these are the most usual in automotive applications. Section 3.2 discusses fabrication techniques.

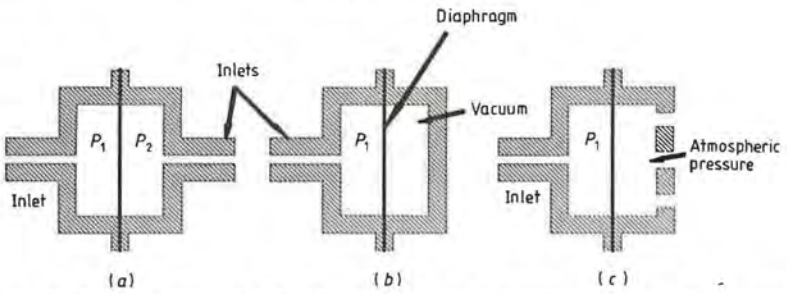

Figure 3.3 Differential (*a*), absolute (*b*) and gauge (*c*) pressure sensors (diaphragm type).

Obviously most automotive engineers will be concerned with the selection rather than the design of pressure sensors. However, understanding how a sensor operates, and what limitations the designer has had to cope with, is an essential prerequisite to its selection and proper use. A detailed discussion of diaphragm-type pressure sensors with piezoresistive and capacitive primary sensing follows, together with a discussion of the signal conditioning techniques employed for capacitive transducers (section 3.4.2). Environmental considerations are discussed in section 3.6, and a summary of the current 'state of the art' in automotive pressure sensing is given in section 3.7.

3.1 ELASTIC PRESSURE SENSORS

In theory a very wide range of elastic structures can be used for pressure measurement. The literature is full of ingenious devices employing novel forms of architecture. However, in practice almost all those which have reached commercial production use one of three approaches, and are based on diaphragms, bellows or Bourdon tubes. Diaphragms are probably the most common for automotive applications, although bellows have been used in the past for measuring manifold absolute pressure (MAP). Bourdon tubes are mainly confined to laboratory measurement systems.

It should be noted that pressure sensors based on an elastic structure have two components; the elastic structure, which forms the primary sensing element and deflects in response to an applied pressure, and a secondary transducer, which senses this deflection and converts it to

(usually) an electrical signal. The most common form of secondary transducer is the strain gauge, although potentiometers, differential transformers, variable capacitors and variable inductors have also been used.

3.1.1 Bourdon Tubes

The Bourdon tube is the basis of many mechanical pressure sensors (particularly the familiar circular moving-pointer type). Bourdon tubes are also used in some electrical transducers, where the output displacement is sensed by potentiometers or differential transformers. The basis of all forms of Bourdon tube is a tube of non-circular (and usually flat-sided or oval) cross section as shown in figure 3.4. One end is sealed, and pressure is applied. The flat sides bulge outward, as though the tube were trying to attain a circular cross section. The resulting distortion tends to straighten the tube. The end of a C-type Bourdon tube undergoes a curved displacement as shown in figure 3.4(*a*), while the spiral and twisted types produce angular motion. Theoretical analysis of the behaviour of a Bourdon tube is difficult, and usually some form of finite-element analysis is needed. Bourdon tube pressure sensors can be markedly nonlinear, and often display an unwanted thermal sensitivity. They also suffer from hysteresis errors, which are usually of the order of 1–2% of full-scale deflection.

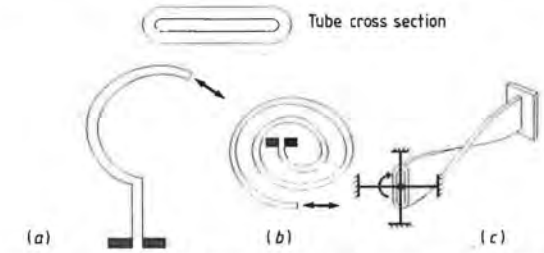

Figure 3.4 Bourdon tubes: (*a*) C-type; (*b*) spiral; (*c*) twisted tube.

C-type Bourdon tubes have been used for pressures up to about 7×10^8 $N\,m^{-2}$ (100 000 psi). The sprial and twisted versions produce larger displacements, and are mainly used below 7×10^6 $N\,m^{-2}$ (1000 psi). The best accuracy that can be achieved is usually around 0.1%.

The free end of a twisted Bourdon tube is usually supported by an arrangement such as that shown in figure 3.4(*c*), which is stiff in all radial directions but soft in rotation. This helps to protect the device from damage due to shock loads and vibration.

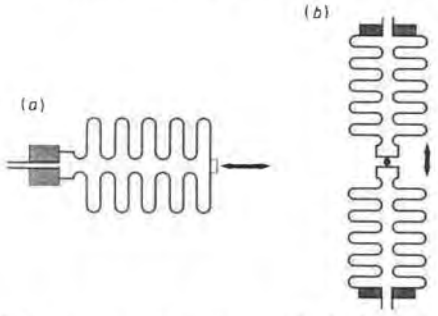

Figure 3.5 Bellows pressure sensors: (*a*) single bellows (gauge pressure); (*b*) double bellows (differential or absolute pressure).

3.1.2 Bellows

Figure 3.5 shows two configurations of bellows. The deflection of a bellows is usually more linear than that of a Bourdon tube. They are reversible with low hysteresis, and are often found in pneumatic systems where they act as pressure/displacement transducers. However, the most common application is undoubtedly in the production of low-cost aneroid barometers for measurement of atmospheric pressure. An evacuated bellows has been used to form an early MAP sensor [1] as shown in figure 3.6. However, it was found that the bellows MAP sensor was relatively expensive to produce, and its large-scale automotive use has been discontinued.

Bellows are manufactured in a variety of materials. The spring rate (modulus of compression) is proportional to the modulus of elasticity of the material from which the bellows is formed, and to the cube of the wall thickness. It is also inversely proportional to the number of convolutions and to the square of the outside diameter of the bellows [2].

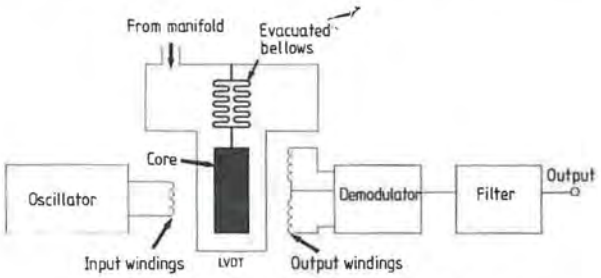

Figure 3.6 MAP sensor using evacuated bellows.

3.1.3 Diaphragms and Membranes

Diaphragms are probably the most popular elastic structure used in pressure sensors. They can be subdivided into two types; thin membranes under radial tension, which form part of an inductive or capacitive pressure sensor, and thicker diaphragms or plates, used in conjunction with resistive or piezoelectric transducers. The membrane type is the most sensitive, and can be used for applications such as measuring low-pressure fluctuations. The most familiar example of a membrane pressure sensor is undoubtedly the capacitor microphone, discussed in detail below (p 48). Where larger pressures are to be measured a thin circular plate is used which is strong enough to carry strain gauges. The plate is either clamped around its circumference by a pair of solid rings, or alternatively the whole assembly may be machined from a solid block of material.

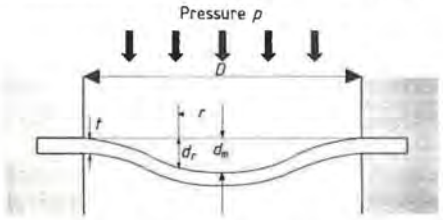

Figure 3.7 Flat circular diaphragm pressure sensor: Y, Young's modulus of the diaphragm; ρ, density (SI units); ν, Poisson's ratio; D, t, d_m all in millimetres.

In order that the relationship between applied pressure and deflection is reasonably linear, the centre deflection, d_m, of the plate must not exceed half its thickness. The approximate design equations for a flat circular diaphragm of this form are as follows (see figure 3.7 for nomenclature and units)

$$d_m = \frac{3(1-\nu^2)D^4p}{256\,Yt^3} \tag{3.1}$$

(d_m is linearly related to pressure p if $d_m \leq 0.5t$). The maximum circumferential stress s_m is

$$s_m \simeq \frac{3D^2p}{16t^2}\ \text{(Pa)}. \tag{3.2}$$

Since a diaphragm has stiffness and mass it will resonate. The lowest natural frequency (for an air or gas medium) is given by

$$f_0 \simeq \frac{10^4 t}{\pi D^4}\sqrt{\frac{Y}{3\rho(1-\nu^2)}}\ \text{(Hz)}. \tag{3.3}$$

For example, using mild steel with a Young's modulus of 210 GPa,

Poisson's ratio of 0.3 and density 7800 kg m^{-3}, a 10 mm diameter diaphragm 0.5 mm thick will deflect about 0.01 mm when the pressure difference is 2.5 MPa (about 250 atm). The resonance frequency will be about 500 Hz.

Equations (3.1), (3.2) and (3.3) are adequate for design purposes if there is no possibility of the centre of the diaphragm deflecting by more than 0.5t. Larger deflections will produce nonlinearity since a stretching action is added to the basic bending of the diaphragm, causing a stiffening effect. For large deflections (i.e. where $d_m \geqslant 0.5t$) equation (3.4) should be used [3]. (Unlike equations (3.1)–(3.3), where the units are those specified on figure 3.7, equations (3.4)–(3.6) require SI units.)

$$p = \frac{16Yt^4}{3R^4(1-\nu^2)}\left[\frac{d_m}{t} + 0.488\left(\frac{d_m}{t}\right)^3\right] \tag{3.4}$$

where R is the diaphragm radius = D/2.

A diaphragm such as that shown in figure 3.7, clamped at the edges, and subjected to a uniform differential pressure p, has at a radius r from the centre on the low-pressure surface, a radial stress s_r and a circumferential stress s_t given by the equations

$$\begin{aligned} s_r &= \frac{3pR^2\nu}{8t^2}\left[\left(\frac{1}{\nu}+1\right) - \left(\frac{3}{\nu}+1\right)\left(\frac{r}{R}\right)^2\right] \\ s_r &= \frac{3pR^2\nu}{8t^2}\left[\left(\frac{1}{\nu}+1\right) - \left(\frac{1}{\nu}+3\right)\left(\frac{r}{R}\right)^2\right]. \end{aligned} \tag{3.5}$$

The deflection d_r at any radius r is given by equation (3.6)

$$d_r = \frac{3p(1-\nu^2)(R^2-r^2)^2}{16Yt^3}. \tag{3.6}$$

Equations (3.5) and (3.6) all give linear relations between stress and pressure, and are sufficiently accurate when $d_m \leqslant 0.5t$. Equation (3.4) can be used to estimate the degree of nonlinearity in any given application.

Figure 3.8 shows the form of the radial and tangential stress distributions on the diaphragm surface. Since regions of both positive and negative stress exist a bridge containing four active sensors may be used. This has two benefits. First, it provides first-order temperature compensation, which makes the pressure sensor output almost immune to thermal changes [4]. Second, the resistance changes experienced by the individual strain gauges are additive in a four-arm bridge, which increases the sensor output. Figure 3.9(a) shows how gauges 2 and 4 are placed as close to the centre of the diaphragm as possible, and are oriented to read tangential (tensile) strain. Gauges 1 and 3 are placed radially, close to the edge of the diaphragm, where the stresses are compressive. Note that equations (3.5) cannot be

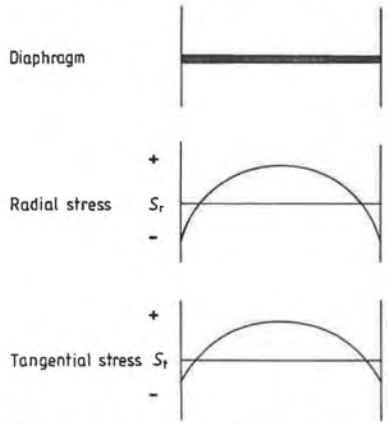

Figure 3.8 Diaphragm stress distributions: +, tensile stress; −, compressive stress.

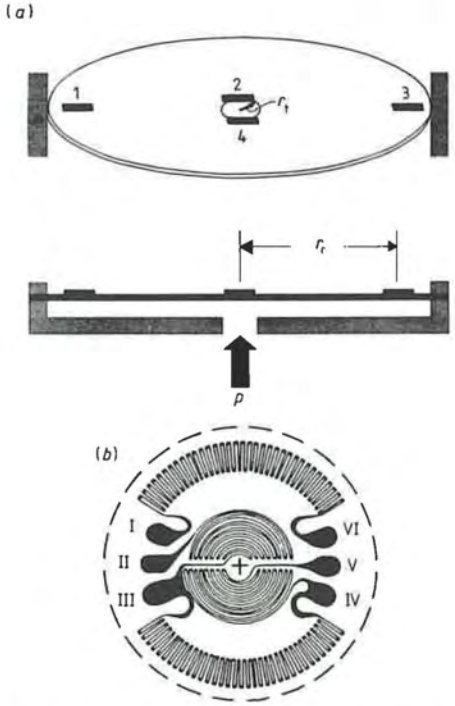

Figure 3.9 (*a*) Diaphragm and strain gauge pressure transducer. (*b*) Strain gauge rosette for pressure diaphragm.

used directly to determine the strains experienced by the gauges, since the diaphragm is in a state of biaxial stress, and both the radial and tangential stress contribute to the radial and tengential strain at any point. The general biaxial stress–strain relation gives

$$\epsilon_r = \frac{s_r - \nu s_t}{Y}$$

$$\epsilon_t = \frac{s_t - \nu s_r}{Y} .$$

Once the gauge strains are known the individual gauge resistance changes ΔR can be obtained from the gauge factors.

To aid in the construction of miniature pressure transducers the discrete strain gauges shown in figure 3.9(*a*) can be replaced by a pressure-diaphragm rosette such as that shown in figure 3.9(*b*). Rosettes of this form are configured to take advantage of the radial strains at the diaphragm edge and the tangential strains at the diaphragm centre. The solder connection points (I–VI) are placed in a low-strain region to avoid damaging the joints.

3.2 FABRICATION TECHNIQUES]

Most small commercial pressure sensors are of the diaphragm type. Three manufacturing techniques are used, two of which are based on stainless steel diaphragms and one which uses silicon. Clearly stainless steel has the advantage of ruggedness and is able to maintain a pressure seal even after electrical failure has occurred. For high-quality sensors where small size and low cost are not prime requirements, foil strain gauges are bonded to a stainless steel diaphragm. The diaphragm is usually encapsulated in a welded stainless steel enclosure, which may also contain the signal conditioning circuitry. In this case the assembly is usually referred to as a pressure *transducer* rather than a pressure sensor. In automotive engineering this type is normally only used for experimental and prototype work.

A cheaper approach is to use thick-film piezoresistors on a stainless steel diaphragm. (A description of the thick-film fabrication process is given in chapter 9.) This results in low-cost sensors which are physically smaller than the traditional bonded-foil strain gauge type, although not as small as silicon versions. It is difficult to apply thick-film strain gauges to diaphragms less than about 3 mm in diameter. The thick-film approach has

a further advantage, in that the cost of setting up a thick-film production plant is considerably less than that of a comparable silicon facility. If the required number of sensors runs to only hundreds or a few thousand per year, thick film will probably prove to be cheaper than silicon. The pressure sensor (i.e. the membrane and its associated piezoresistors) can be made to form part of a *thick-film hybrid*, which allows signal conditioning circuits to be included within the sensor housing.

If very large numbers of sensors are required, or if small size is essential, silicon architecture is almost always adopted. The predominance of silicon as the material used for creating low-cost sensors is partly due to the wealth of experience on manipulating and using silicon built up by the semiconductor industry. Its mechanical properties also make it well suited for sensor use. Silicon has a density less than that of aluminium, and an elastic modulus approaching that of steel. It obeys Hooke's law over a large strain range than steel, and has a higher ultimate tensile strength. Silicon's main disadvantage as a sensor material is that the complex technology used to create sensor architectures is very expensive, and large production runs are needed before the initial costs are recovered. The other disadvantages of silicon diaphragms are that they are easily damaged by water and other chemicals, and that they tend to shatter if struck by small gas or liquid-borne particles. Since silicon is a brittle crystalline material it does not yield plastically and will crack or shatter if exposed to shock loads. Protective screening techniques are available for shock protection, but these cannot always be used conveniently. Silicon pressure sensors are available using both piezoresistive and capacitance transducers.

3.2.1 Silicon Micromachining

Micromachining can be defined as the collection of processes used to manufacture small mechanical components and structures from silicon. Some of these processes were originally developed for integrated circuit production, while others have been evolved specifically for sensor production. Micromachining techniques can be considered under four headings:

(i) Abrasive techniques such as sawing and grinding, which are used to cut coarse features into silicon.

(ii) Methods used in the semiconductor industry. These include lithographic and deposition techniques. Photolithography is used to transfer patterns onto a silicon wafer to act as a mask for etching. Deposition processes allow multiple layers of materials to be built up. These layers can have mechanical or electrical functions. Oxidation,

diffusion and ion implantation are also used to create masks for etching, etch stops or circuit elements such as piezoresistors.

(iii) The most important micromachining process is etching. Wet etching involves the use of liquids which dissolve silicon. There are isotropic etchants available, which etch the material equally fast in all directions, but in general selective etchants are more useful. These show a very high selectivity (several hundred to one) in the rate at which they attack different crystal planes. This feature can be used to produce highly accurate angles and dimensions, which can be arranged to self-align and self-stop as dictated by the crystal planes. The most common wet etchants are potassium hydroxide (KOH) and ethylenediamine pyrocatechol (EDP). KOH is the safest to use, but EDP has the advantage that oxide† can be used as a mask material to define the etch pattern.

A useful adjunct to the use of the wet process is the ability to inhibit etching in certain areas, usually at a certain depth in the Z direction, by pre-doping the silicon with boron. This 'etch stop' technique is often used to determine the thickness of a micromachined component such as a diaphragm.

Dry etching involves the use of gaseous etch species which react with the material to be etched, producing gaseous products. These waste gases are extracted by a vacuum system. Dry etching is used in the production of very fragile components, where the structure being created may be damaged by the presence of a liquid. The most common dry etch processes are plasma etching, in which the substrate to be etched is enveloped in an RF plasma, and reactive ion etching (RIE), in which the substrate is given an electrical bias so that it attracts ions which then physically remove material.

(iv) Bonding techniques have been developed to allow further variations in sensor architecture and packaging. Fusion bonding is achieved by heating wafers in good flat contact, and gives either a silicon–silicon bond or one with an oxide layer between the wafer faces. Bonding to metals and glass is achieved by the use of an electric field, and is often termed anodic bonding.

3.3 STRAIN GAUGES AND PIEZORESISTANCE

The background to conventional foil strain gauges of the type used in pressure sensors is well known [4], and will not be discussed in detail here.

†'Oxide' is the conventional term for silicon dioxide. It is used as an electrical insulator, a mechanical component or as a mask for etching.

In essence an applied strain changes the geometry of the strain gauge, and hence its resistance.

The operating principle of the piezoresistive strain gauges used in silicon and thick-film pressure sensing is probably less familiar, however. The piezoresistive effect occurs in silicon and thick-film materials. As the name implies, a change in electrical resistance occurs in response to changes in the applied stress. Piezoresistive (PR) sensors are formed by placing stress-sensitive resistors on highly stressed parts of a suitable mechanical structure such as a diaphragm. The PR transducers are connected in a Wheatstone bridge circuit. When the diaphragm is exposed to a differential pressure it undergoes strain changes. These strain variations are converted into an electrical output which is proportional to pressure (see equations (3.5)) by the PR transducers. The advantage of using piezoresistive transducers is that very high gauge factors may be achieved.

The drawbacks of PR devices are that the output signal level is moderate (typically 100 mV full scale for a 10 V bridge excitation), and the sensitivity can be temperature dependent, although this can largely be compensated for by proper bridge design.

If a rectilinear resistor has length l, width w, thickness t and a bulk resistivity ρ its resistance R will be

$$R = \frac{\rho l}{wt} . \tag{3.7}$$

The gauge factor or strain sensitivity is defined as k, where

$$k = \frac{\mathrm{d}R/R}{\varepsilon} .$$

ε is the relative change in length of the resistor (the strain) due to a stress σ, applied to the substrate parallel to its length. Figure 3.10 shows the consequences of the applied stress. The length increases by an amount $\mathrm{d}l$, while the width and thickness decrease by $\mathrm{d}w$ and $\mathrm{d}t$ due to Poisson's ratio ν. It is clear that $\mathrm{d}w = -\nu w\varepsilon$ and $\mathrm{d}t = -\nu t\varepsilon$.

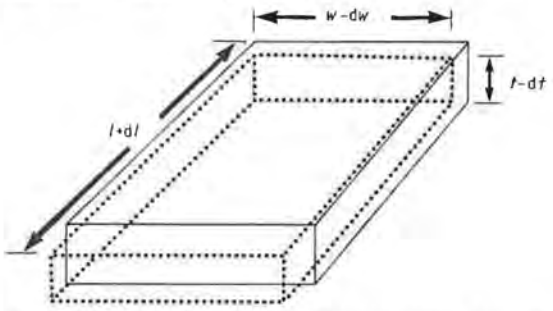

Figure 3.10 Consequences of applying stress to a piezoresistor.

The original cross section was

$$A = wt.$$

Owing to the strain ε, the new cross-sectional area is

$$\begin{aligned} A' &= (w - dw)(t - dt) \\ &= wt + 2\nu wt\varepsilon + \nu^2 wt\varepsilon. \end{aligned}$$

The term $\nu^2 wt\varepsilon$ is very small compared with the other two terms in the equation and can be neglected. We can therefore write the change in cross-sectional area as

$$A - A' = dA = -2\nu\varepsilon A$$

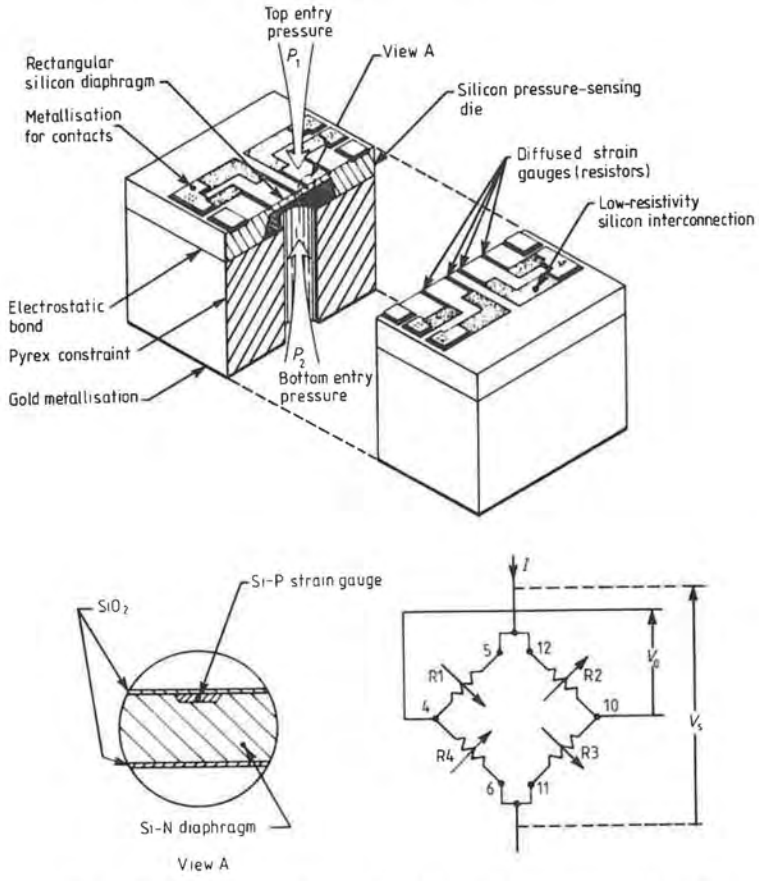

Figure 3.11 (Above and opposite) Micromachined pressure sensor (courtesy of IC Sensors).

giving

$$dA/A = -2\nu\varepsilon.$$

Differentiating equation (3.7) gives

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dl}{l} - \frac{dA}{A}$$

and hence the gauge factor k is

$$k = \frac{d\rho/\rho}{\varepsilon} + (1 + 2\mu). \tag{3.8}$$

Typically ν will be between 0.2 and 0.3. Equation (3.8) shows that the longitudinal gauge factor is a function of changes in both longitudinal

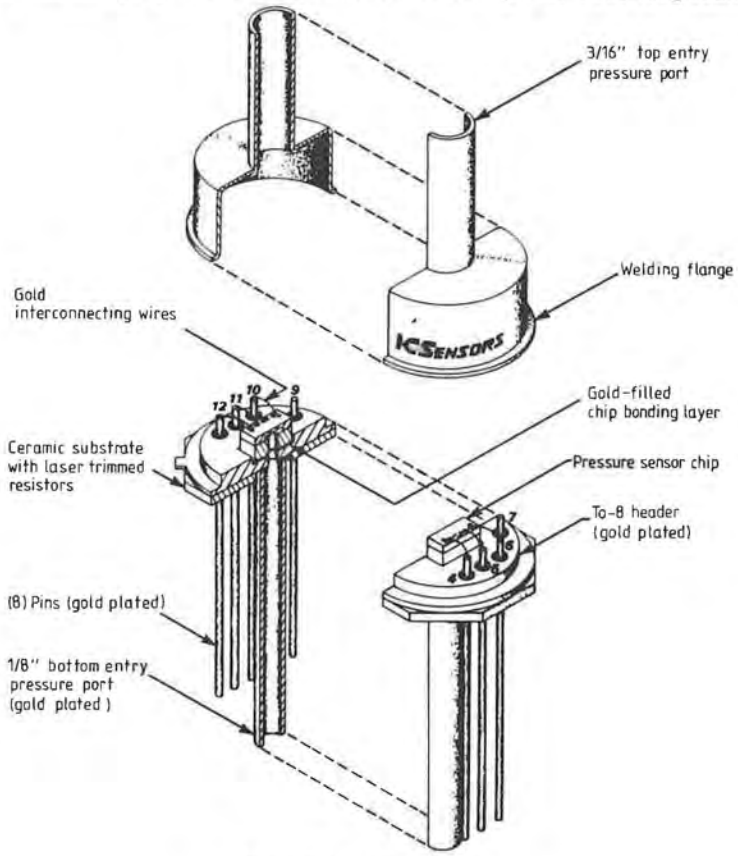

Figure 3.11 (Continued)

resistivity and geometry. In a conventional foil or wire strain gauge the piezoresistive effects are negligible, and the variations in resistance are mainly a function of dimensional changes. For a foil gauge k is approximately 2. For piezoresistive strain gauges the first term in equation (3.8) is significant, and higher gauge factors (typically around 10) can be achieved, giving enhanced sensitivity. It should be noted, however, that the resistivity of most PR materials is strongly temperature dependent, and that as a result PR strain gauges generally have a higher thermal sensitivity than other types. This difficulty can be partially overcome by the use of temperature compensation circuits.

Figure 3.11 shows a typical micromachined pressure sensor using piezoresistive strain gauges. The diffused strain gauges are interconnected within the device to form a Wheatstone bridge.

3.4 CAPACITANCE PRESSURE SENSORS

The majority of silicon pressure sensors are piezoresistive and use diffused or ion-implanted piezoresistors. However, a number of successful capacitive pressure sensors have been used, notably by the Ford Motor Company with their silicon capacitance absolute pressure (SCAP) sensor [6].

In piezoresistive pressure sensors diffused or implanted piezoresistors are arranged in single, half or full bridge forms. The differential strains sensed by the bridge resistors are used to achieve a push–pull effect to increase the sensitivity as discussed earlier. However, the limiting value of $\Delta R/R$ (i.e. change in resistance over unstrained resistance) is only about 10^{-2}, due to hysteresis and linearity considerations. PR devices are sensitive to transverse stressing of the sensor package, and the base strain sensitivity can be considerable. This increases the chance of interference, especially when the measurand is small and the sensor is subjected to vibration.

Capacitive pressure sensors are less prone to error. The diaphragm deformation caused by an applied pressure is turned directly into a change in capacitance, which can then be converted into an electrical form such as frequency, charge or voltage. Since the total change in capacitance is the integral of the changes produced by each part of the diaphragm the susceptibility to side stress is lower than that of a comparable PR pressure sensor.

Capacitance pressure sensor (CPS) designs can be grouped into three categories as shown in figure 3.12. Figure 3.12(a) shows a pedestal-and-ring arrangement, originally designed for a medical application [7]. The circular diaphragm is 610 μm in diameter and 10 μm thick. The

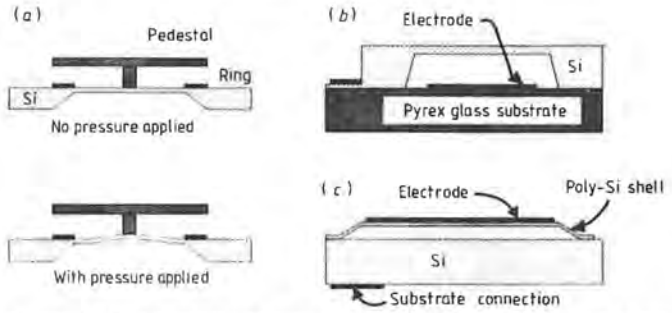

Figure 3.12 Capacitive pressor sensor structures: (*a*) pedestal-to-ring type; (*b*) clamped-edge diaphragm type; (*c*) thin-shell chamber type.

pedestal/ring gap is 5 μm and the zero-pressure capacitance 1 pF. At 300 mmHg full scale $\Delta C/C = 0.2$ (cf 0.01 for PR strain gauges). The device is somewhat nonlinear (around 10% maximum), but is highly repeatable.

Figure 3.12(*b*) shows a different approach in which a silicon diaphragm is etched from a 300 μm thick wafer. The diaphragm is bonded to a Pyrex glass substrate. The gap chamber is either evacuated and sealed for absolute pressure measurement, or is supplied with a vent for gauge or differential pressure measurement. The Ford SCAP sensor discussed earlier is of this type. Some recent transducers have used silicon rather than glass as the substrate to reduce the thermal drift of the sensor and simplify the processing.

Figure 3.12(*c*) shows a thin-shell type pressure sensor. Micromachining techniques are used to deposit material on the silicon wafer and selectively etch off part of the material to construct a microchamber or cavity as shown in the diagram. The capacitor electrodes are on the top of the chamber and on the surface of the silicon substrate. With this approach no bonding is needed. However, care has to be taken to control built-in stress in the shell material.

In a diaphragm CPS the diaphragm deforms in response to the differential pressure p across its sides, as shown in figure 3.13. This deformation changes the capacitance C between the electrodes, and a signal conditioning circuit is used to convert the output to a voltage V_0. If the supply voltage is V_s, the overall sensitivity of the CPS is S where

$$S = \left(\frac{\Delta C/C_0}{p}\right)\left(\frac{\Delta V_0/V_s}{\Delta C/C_0}\right) = G_c A_v = \frac{\Delta V_0}{pV_s}.$$

C_0 is the zero-pressure capacitance. The first term

$$G_c = \frac{\Delta C/C_0}{p}$$

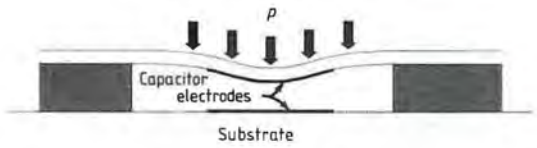

Figure 3.13

is similar to the gauge factor of a piezoresistive sensor, and A_v is the capacitance-to-voltage gain of the signal conditioning circuit.

3.4.1 Capacitor Microphones

Microphones are increasingly finding automotive applications for ultrasonic theft alarms, in-car telephones and as part of active noise reduction systems. The capacitor microphone (also called a condenser or an electrostatic microphone) is essentially a parallel-plate capacitor. One of the plates is a thin membrane exposed to the medium in which the sound is to be measured, so that pressure fluctuations alter the capacitor plate spacing. The resulting capacitance changes cause fluctuations in the voltage across the capacitor. The output signal consists of a varying voltage V_{out} as shown in figure 3.14.

The charge on the capacitor may be generated by an externally applied voltage, or by the properties of the material used to manufacture the capacitor. In the latter case the sensor is often called an electret microphone. An electret is a permanently charged insulating material made by allowing a molten plastic to solidify under the influence of a strong electric field.

The components of a typical capacitor microphone are shown in figure 3.15. A diaphragm made from metal foil is fixed close to an insulated rigid metal back plate. These two form a parallel-plate capacitor. A polarisation

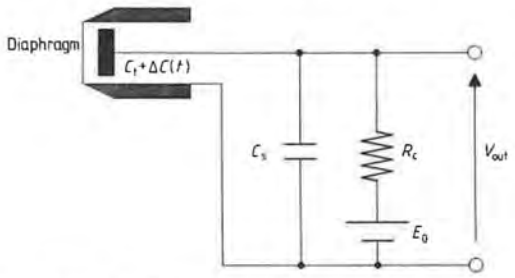

Figure 3.14 Equivalent circuit of a capacitor microphone.

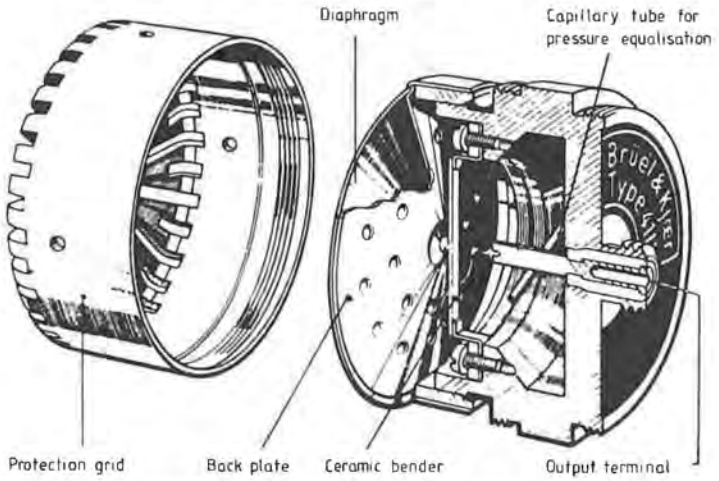

Figure 3.15 Capacitor microphone construction (courtesy of Bruel and Kjaer).

voltage E_0 is applied across the plates as shown in figure 3.14. The polarisation voltage source has a high impedance R_c, so that the time constant $R_c(C_t + C_s)$ is long compared with the lowest sound frequency to be measured. A further reason for making the time constant long is that this ensures that the charge stored on the capacitor is approximately constant. If the charge is constant and the capacitance varies, a varying voltage will appear across the plates. The capacitance C_s is due to unavoidable stray capacitance within the sensor, rather than a deliberate design feature.

If the sound pressure acting on the diaphragm produces a capacitance change ΔC, then the output of the microphone will be a voltage V_{out} where

$$V_{out} = \frac{\Delta C E_0}{C_t + C_s}$$

since $C \gg \Delta C$. Note that the microphone sensitivity S is proportional to the polarisation voltage E_0 but inversely proportional to the total capacitance $C_t + C_s$. Capacitor microphones can be manufactured with sensitivities as high as 100 mV Pa^{-1}.

If we want the microphone to have a flat response then the capacitance change ΔC and hence the deflection of the diaphragm for a given sound pressure must be independent of frequency. In other words, the diaphragm must be 'stiffness controlled' [5], and it will have a natural frequency well above that of the highest-frequency sound to be measured.

The voltage output from a capacitor microphone is proportional to the applied sound pressure. However, when the frequency of the sound

increases to the point where the wavelength is of the order of the diaphragm diameter, the diaphragm acts as a high-impedance obstacle in the sound field, from which the wave will be reflected. When this happens the pressure sensed by the diaphragm is incorrectly high. This is a manifestation of the 'pressure doubling' effect, discussed further in [4]. In such a situation a microphone with a flat pressure response would give an incorrect reading.

3.4.2 Capacitance Measurement Circuits

A variety of circuits can be used to convert a change in capacitance into a usable electrical signal. The techniques described here are reasonably simple, although much more complicated arrangements are described in the literature [7].

The most common form of capacitive sensor is the parallel-plate capacitor with a variable air gap. Most pressure sensors are of this form. If two plates, each of area A, are separated by a perpendicular distance y, field lines pass from one to the other as shown in figure 3.16. Except at the edges, the magnitude of the field is V/y, where V is the potential difference between the plates. If y is small compared with the dimensions of the plates edge effects are unimportant, and we can assume the field to be uniform between the plates and zero elsewhere. Under these conditions the capacitance C and the stored charge Q are related by

$$C = \frac{Q}{V} = \frac{\varepsilon_0 A}{y}. \tag{3.9}$$

ε_0 is the permittivity of free space, which is $8.85 \times 10^{-12} C\,N^{-1}m^{-2}$. The value for air will be within 0.1% of this value. Equation (3.9) can be used to calculate the nominal capacitance of any capacitive sensor. For example, a pressure sensor based on a diaphragm 5 mm in radius, 0.1 mm away from a fixed electrode, will have a nominal capacity of C_0 where

$$C_0 = \frac{8.85 \times 10^{-12} \times \pi \times (5 \times 10^{-3})^2}{1 \times 10^{-4}} \simeq 7 \times 10^{-12}$$

(i.e. 7 pF). The impedance of a capacitor is $1/j\omega C$. Thus the magnitude of the impedance of a device based on a 5 mm diaphragm will be around 1.4 MΩ at 10 kHz. The high impedance of capacitive sensors is the cause of some of the difficulties associated with their use. Spurious noise

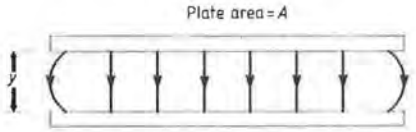

Figure 3.16 Cross section through a parallel-plate capacitor.

voltages can appear, and it is essential to use high-input-impedance electronics to avoid undue sensitivity to the length and position of connecting cables.

Equation (3.9) shows that the variation of capacitance C with plate separation d is nonlinear (in fact it is hyperbolic). Thus the percentage change in d from a chosen neutral position must be small if reasonable linearity is to be achieved. However, ratios of $\Delta C/C$ up to 0.2 are common, and as mentioned earlier this is at least an order of magnitude better than anything which can be achieved with strain gauges.

The sensitivity of capacitance to changes in plate separation can be seen if equation (3.9) is differentiated

$$\frac{dC}{dy} = -\frac{\varepsilon_0 A}{y^2}. \tag{3.10}$$

The sensitivity dC/dy increases as y decreases. However, the *percentage* change in C is equal to the *percentage* change in y for small changes about any neutral position, since

$$\frac{dC}{dy} = -\frac{C}{y},$$

$$\frac{dC}{C} = -\frac{dy}{y}.$$

Capacitor microphone circuit The simplest circuit used for capacitive sensors is shown in figure 3.14, and is that generally employed with capacitor microphones. When the capacitor plates are at rest with a separation y_0 no current flows and $V_{out} = 0$. If there is then a relative displacement y_d from the equilibrium position y_0, a capacitance change ΔC is produced which is related to V_{out} by

$$V_{out} = \frac{\Delta C(t) E_0}{C_t + C_s} \tag{3.11}$$

since $C_t \gg \Delta C(t)$. It should be noted that the sensor sensitivity is proportional to the polarisation voltage E_0 but inversely proportional to the total capacitance $C_t + C_s$.

From equation (3.11) the output V_{out} is clearly zero when $\Delta C(t) = 0$, i.e. when the capacitor plates are at rest in any equilibrium position. Thus the circuit of figure 3.14 does not allow static pressure measurements to be made. If y_0 is varying, however, V_{out} will measure the motion. Microphones do not usually need to measure sound frequencies below about 20 Hz, for which this circuit is entirely adequate.

Differential capacitance sensors If a 'push–pull' or differential capacitive sensor is to be used the arrangement shown in figure 3.17 will allow the measurement of static deflections and hence of static pressures. A thin

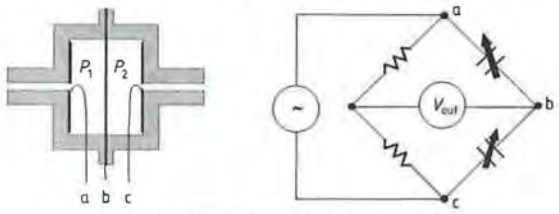

Figure 3.17 Push–pull capacitor sensor.

diaphragm acts as the movable plate and is held between a pair of fixed electrodes. With equal pressures applied to the two ports the diaphragm is centred and the out-of-balance voltage V_{out} is zero. If the pressures are different the diaphragm deflects, one capacitor increasing in value and one decreasing. V_{out} is then a sinusoidal signal with an amplitude which is proportional to the applied pressure. If the pressure difference is reversed V_{out} undergoes a 180° phase change. A direction-sensitive DC output can therefore be obtained by conventional phase-sensitive demodulation and filtering.

Linear feedback method A useful method of circumventing the nonlinearity inherent in equation (3.9) is shown in figure 3.18. A capacitive sensor is connected as the feedback component in an operational amplifier. Op-amp circuits may be analysed by assuming the following [4]:

(i) The inputs draw such a small current that it may be assumed to be negligible.

(ii) Since the open-loop gain is very high, the inputs can be assumed to be at the same voltage.

Thus we can write

$$\frac{1}{C_f}\int I_{in}dt = E_0 - E_{AB} = E_0$$

$$\frac{1}{C_s}\int I_s dt = V_{out} - E_{AB} = V_{out}$$

$$I_{in} + I_s - I_{AB} = 0 = I_{in} + I_s.$$

Manipulation then gives

$$V_{out} = \frac{1}{C_s}\int I_s dt = -\frac{1}{C_s}\int I_{in}dt = -\frac{C_f}{C_s}E_0 \quad (3.12)$$

$$V_{out} = -\frac{C_f y E_0}{\varepsilon_0 A} = Ky \quad (3.13)$$

where K is a constant. Equation (3.13) shows that the amplitude of V_{out} is directly proportional to the plate separation y. Linearity is therefore

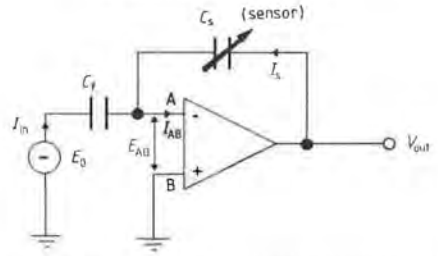

Figure 3.18 Feedback capacitive sensor circuit.

achieved for both small and large displacements. E_0 is usually between 10 and 100 kHz, with 50 kHz being typical. The output V_{out} is also a 50 kHz sinusoid, which is often rectified to provide a DC voltage.

Oscillator circuits A common approach is to use a capacitive sensor to determine the frequency of operation of an oscillator. Figure 3.19 shows one approach. The circuit is known as a *relaxation oscillator*. The op-amp is assumed to be ideal and to be supplied from equal and opposite DC voltages V_s. The circuit is essentially a differential amplifier, so if the differential input V_{AB} is positive, V_{out} will limit at $+V_s$, and vice versa.

When the circuit is switched on V_{out} will go to one of the supply voltages V_s. (It is impossible to predict which one, and in any case it does not matter.) Assume that V_{out} equals $+V_s$. The non-inverting terminal A is connected to the mid-point of the potential divider formed by R_1 and R_2. If we define the potential divider ratio as β, where

$$\beta = \frac{R_2}{R_1 + R_2}$$

it is obvious that the voltage at the inverting terminal V_A must always equal βV_{out}. So at switch-on $V_A = +\beta V_s$. Figure 3.20 shows the voltages V_{out} and V_A at switch-on when time $t = 0$. At switch-on the capacitor C (which is the sensor) is uncharged, so V_B equals zero as indicated.

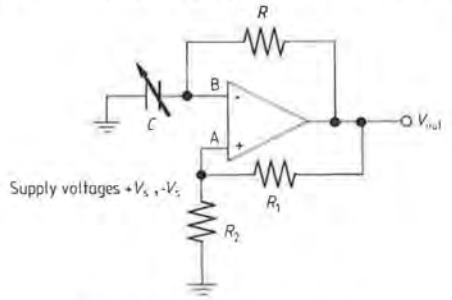

Figure 3.19 Relaxation oscillator.

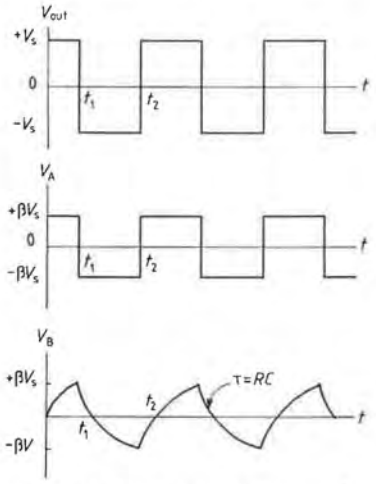

Figure 3.20 Relaxation oscillator waveforms.

The initial voltage across the feedback resistor R is V_s. A current i flows through R, and begins to charge C. As a consequence V_B begins to rise exponentially towards V_s, with a time constant $\tau = RC$. However, when V_B passes through βV_{out} the inverting input becomes more positive than the non-inverting input A. This means that the differential input V_{AB} is now negative, so the circuit switches at t_1 to $V_{out} = -V_s$. We now have

$$V_{out} = -V_s$$
$$V_B = +\beta V_s$$

and

$$V_A = \beta V_{out} = -\beta V_s.$$

The current i through R and C is now reversed, and noting the voltages at either end of R we have

$$i = (V_s + \beta V_s)/R.$$

V_B will now fall exponentially towards $-V_s$ (see the graph of V_B from t_1 to t_2 in figure 3.20). As soon as V_B passes through $-\beta V_s$, V_{AB} becomes positive, and the cycle begins again.

To calculate the output frequency refer to figure 3.21. At t_1 we have

$$i = \frac{V_s + \beta V_s}{R}.$$

Immediately before the transition at t_2, V_B has fallen to $-\beta V_s$. Therefore the voltage across R is $V_s - \beta V_s$, and i has fallen to

$$\frac{V_s - \beta V_s}{R}.$$

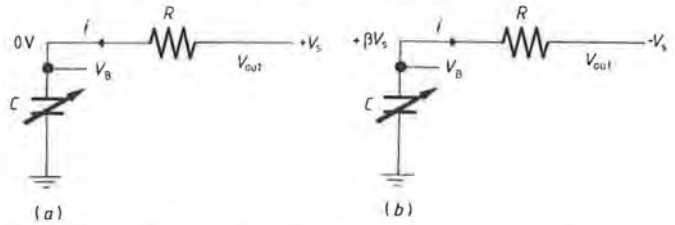

Figure 3.21 Relaxation oscillator timing network: (*a*) at switch-on; (*b*) immediately after switching to $-V_s$.

Using these two current values and the fact that a capacitor discharge is exponential, we have

$$\frac{V_s}{R}(1-\beta) = \frac{V_s}{R}(1+\beta)\,e^{-T/RC}.$$

Therefore

$$e^{+T/RC} = \frac{1+\beta}{1-\beta}$$

and

$$T = RC\ln\left(\frac{1+\beta}{1-\beta}\right)$$

where T is the time taken by half a cycle of the complete waveform. The output frequency (which is a function of C if R is fixed) is

$$f_{out} = \frac{1}{2T}.$$

An alternative to the relaxation oscillator is to use the sensor as the capacitive element in an LC oscillator of the Hartley or Colpitts type [8]. In the latter case it may be necessary to add extra parallel capacitance to the transducer, which will reduce the fractional change in frequency $\Delta C/C$.

A possible disadvantage of the LC oscillator is the fact that the period T depends on the square root of C. However, if the fractional change in C is small, the relationship between T and C will be reasonably linear. For good accuracy it is obviously essential that the inductor L is stable in value and has a stable temperature coefficient.

An advantage of the LC circuit as compared with the relaxation oscillator is that because of the high circuit Q phase jitter is reduced. However, the disadvantage of a high Q is that the system responds more slowly to changes in C.

3.5 PRESSURE SWITCHES

Pressure switches are sometimes required in automotive engineering.

These normally have the function of indicating when a preset pressure threshold is passed. Examples include engine oil pressure sensors, brake hydraulic system sensors and simple tyre pressure monitoring systems where a minimum pressure is detected. Most pressure sensors are diaphragm-based and rely upon the diaphragm deflection being sufficient to close (or open) a contact. The threshold pressure is dictated by the diaphragm geometry, and is not usually adjustable by the user. Silicon micromachined pressure switches are available which can be used in applications where size or cost restricts the use of conventional pressure sensors. Pressure switches are generally cheaper than pressure sensors, since no strain gauges or interconnections are needed. A typical device is shown in figure 3.22. The diaphragm is typically 1 mm across and is fabricated from silicon as described in section 3.2.1. Instead of diffusing resistors into the silicon to measure strain, a metal pattern is deposited in an etched shallow well.

The setpoint (i.e. the pressure at which the contacts close) is adjusted during manufacture by a process known as electrical trimming. The metal pattern consists of a regular series of raised metal contacts as shown in the diagram. Clearly, when pressure is applied the contacts at the centre of the diaphragm will close first. As the pressure is increased contacts further away from the centre will close in succession. Each metal contact is manufactured to include a micromachined fuse, which can be blown by applying a voltage pulse. Thus if a switch is required to operate at low pressure, no fuses are blown. If a high-pressure switch is required some of the central fuses are blown, ensuring that a larger force has to be applied before a valid switch connection is made.

Pressure switches are generally more robust than pressure sensors, since the glass substrate provides an integral overforce protection for the silicon diaphragm.

The main disadvantage of this type of pressure switch is that the current through a pair of contacts must be smaller than that which is required to blow a fuse. For this reason commercial devices often include a transistor switching stage.

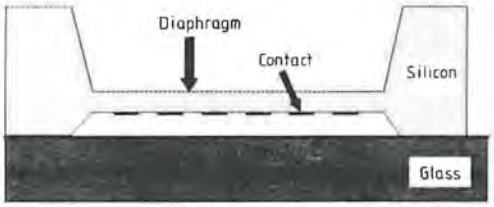

Figure 3.22 Cross section of a silicon pressure switch.

3.6 ENVIRONMENTAL CONSIDERATIONS

Automotive pressure sensors have to function in a particularly hostile environment. They are subjected to extremes of temperature (typically from −40 to +140 °C), vibration, electromagnetic interference and chemical attack. Pressure measurements may be required in water, oil, hydraulic fluid and air. Some of these fluids will attack some types of diaphragm, and special precautions may be necessary as a consequence. The sensor may be required to operate in a high-frequency, high-intensity electromagnetic environment. Electrical connections can also cause problems if small signals are to be transmitted. For these last two reasons it is desirable to include as many of the signal conditioning functions as possible within the sensor housing. The intention of this section is to indicate some of the likely areas in which problems may occur, and to outline some of the solutions.

3.6.1 Chemical Attack

The diaphragm material which is most immune to chemical attack is undoubtedly stainless steel. Stainless steel diaphragms can readily be made with diameters as small as 3 mm and with a minimum thickness of 0.01 mm or less. Steel diaphragms may be electron-beam welded to the sensor housing to provide an impermeable and rugged seal. However, they are often strain-gauged manually, and this leads to a high unit cost which is often unacceptable in automotive applications.

Silicon diaphragms may be made cheaply by micromachining (see section 3.2) and can have piezoresistors diffused onto the surface as part of the manufacturing process. The unit cost of a silicon pressure sensor in mass production is very low, and they can be made extremely small with active diaphragm diameters down to 0.5 mm. The sensor package can include sophisticated signal conditioning functions, which are fabricated by the same processes used to form the diaphragm and its associated piezoresistors. However, silicon is attacked by water and many other fluids, and it is generally unwise to expose silicon directly to the fluid in which pressure is to be sensed.

Thick-film piezoresistors on stainless steel offer a compromise between the approaches described above. The strain-gauging is to some extent automated, since a printing process is used (see chapter 9), and the strain gauges are of lower cost than the bonded foil type. Signal conditioning circuits may be included by the creation of a thick-film hybrid device. However, the size of a thick-film-on-steel pressure sensor is larger than that of a comparable silicon device.

A number of techniques are adopted which enable the user to profit from the benefits of silicon while also enjoying the ruggedness and immunity to chemical attack provided by stainless steel. The first of these involves the use of two diaphragms and silicone oil as a pressure-transmitting medium. Figure 3.23(*a*) shows the arrangement. Silicone oil is chemically inert, and a silicon device can safely be exposed to it with no possibility of damage. The silicon sensor is mounted within a stainless steel shell, sealed at the front by a welded stainless steel diaphragm. Pressure in the working fluid causes this diaphragm to deflect, which causes similar pressure variations to appear in the silicone oil. The silicon diaphragm senses the pressure in the silicone oil, rather than that in the (potentially damaging) working fluid. The electrical connections to the device are sealed by means of a glass feed-through.

An alternative approach is to use a flexible silicone rubber barrier (often called a 'sock') to contain the silicone oil as shown in figure 3.23(*b*).

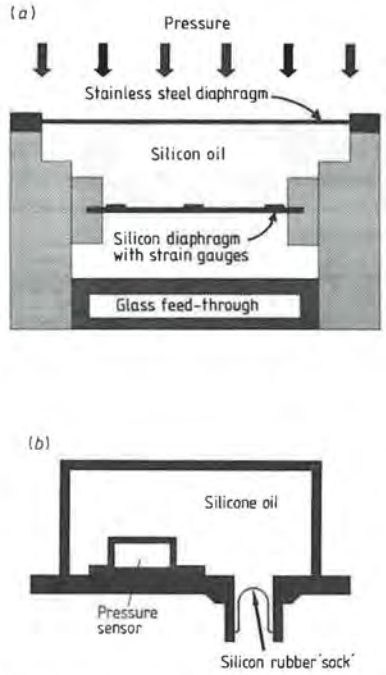

Figure 3.23 (*a*) Double diaphragm for chemical protection. (*b*) Silicone rubber sock to protect the sensor.

3.6.2 Overrange Effects

Most pressure transducers will survive exposure to pressures beyond their measurement limit without damage. Manufacturers' data sheets usually give overrange values, although these are generally very conservative and most devices will survive considerably more than the suggested limits. However, the effects of transducer resonance must be borne in mind when considering overrange limits. If dynamic pressures are to be measured at frequencies more than about 30% of the resonance frequency, failure can occur as a result of the amplifying effect of mechanical resonance.

3.6.3 Acceleration Sensitivity

If a diaphragm is clamped around its edges and the whole assembly is accelerated, the diaphragm will deflect. The deflection is a function of the diaphragm mass and stiffness. In automotive applications any sensor acceleration which causes erroneous readings is likely to result from engine or other high-frequency vibrations. The acceleration levels resulting from cornering, braking etc are normally too low to affect the output of a pressure sensor. An exception is in the area of tyre pressure sensing, however, since any sensor attached to a tyre will experience centrifugal accelerations up to about 800*g*, plus shock loadings from irregularities in the road surface.

Well designed commercial pressure sensors have acceleration sensitivities of the order of 0.000 05% of full-scale per *g* in the pressure sensitive direction. Cross-axis acceleration sensitivities are generally less than 20% of that in the sensitive direction.

3.6.4 Thermal Sensitivity

All pressure sensors have characteristics which are to some extent thermally dependent. The sensitivity, linearity and zero offset can all change with temperature.

Thermal sensitivity changes The manufacturer of a pressure sensor will normally provide details of its thermal sensitivity as part of the device's datasheet. Typical variations for silicon and stainless steel devices are of the order of 1–2% per 100 °C. This means that the sensitivity of the sensor will change by 1–2% away from the calibrated value for every 100 °C by which the sensor temperature departs from that at which it was calibrated. For example, a sensor with 1% thermal sensitivity and a nominal sensitivity of 1 mV Pa^{-1} will shift to between 0.99 and 1.01 mV Pa^{-1} if its temperature is changed by 100 °C.

Zero offset shift with temperature Almost all sensors have some 'null

pressure' or 'baseline' offset. This means that when the device is switched on the output will not be exactly zero when no pressure is applied. Zero offset values are usually specified in the datasheet, and are typically 10–20 mV for strain gauge bridge devices. Zero offsets can arise from incorrect balancing of the strain gauge bridge during manufacture, or from incorrect transducer mounting which results in the diaphragm being placed under stress. For example, over-torquing of screw-coupled pressure sensors can result in the appearance of an unexpected zero offset. It is common for large zero offsets to be present at the transducer output on switch-on, which reduce as the sensor 'warms up'.

Typical thermal changes in zero offset are of the order of 1–2% of full-scale output per 100 °C. It should also be borne in mind that any thermal zero shift data supplied with a pressure sensor will only be valid for equilibrium temperature conditions. If large thermal gradients are present, or if rapid temperature changes occur, unpredictable changes in output may take place.

Thermal linearity changes The nonlinearity and hysteresis of a pressure sensor's output vary over the full-scale pressure range. Normally the manufacturer will provide 'worst-case' figures in the form of a percentage deviation from the ideal (straight line) output graph. It is rare for a manufacturer to specify the extent of any thermally induced linearity change. However, these undoubtedly occur, and if extreme accuracy is required the user is well advised to carry out calibration checks at intervals across the required operating temperature range.

3.6.5 Electromagnetic Susceptibility

Pressure sensors, like all other transducers, are sensitive to radio-frequency interference (RFI). Most RFI is carried by the power supply connections, although some is also radiated, especially close to the high-voltage ignition circuits. The best solution to this is to screen or remove the source of the electromagnetic field. However, if this is not feasible the next best alternative is to encapsulate the transducer in a container made from a conducting material such as stainless steel. This has the subsidiary benefit of providing mechanical protection, and can be combined with protecting the diaphragm from chemical attack as discussed in section 3.6.1.

3.7 SUMMARY OF CURRENT AUTOMOTIVE PRACTICE AND CONCLUSIONS

The first automotive pressure measurement was almost certainly of the degree of vacuum in the inlet manifold. Mechanical vacuum-operated

'advance and retard' mechanisms were introduced in the 1920s, and dominated the design of ignition timing control systems until the early 1970s. A number of early manifold absolute pressure (MAP) sensors were based on diaphragms or bellows, and used potentiometers to transduce motion into an electrical form. However, these systems were often unreliable, and they were rapidly superceded by silicon devices as soon as these became widely available.

A number of designs of MAP sensor have been used in the last two decades, including devices based on capacitive, inductive, potentiometric and strain gauge techniques. The most widely used devices employ a silicon diaphragm, with either diffused piezoresistive strain gauges [9] or capacitance deflection sensing [6]. Silicon micromachining techniques produce rugged devices which stand up well to the rigours of an automotive environment. If large numbers are to be produced the unit costs can be quite low. However, it is a common fallacy to suppose that automotive products are manufactured in large numbers. For example, in 1991 Ford produced about 1.5 million vehicles in Europe. If a pressure sensor is required for use in a particular model, the resulting annual demand of a few hundred thousand devices, although large by automotive standards, is seen as small by electronic component makers, who are used to producing devices by the million.

A technology which seems likely to grow in importance as a result of the figures given above is the fabrication of thick-film devices. As explained in section 3.2, the costs of setting up a thick-film production facility are much lower than those for silicon, and the initial outlay involved in designing a new device is greatly reduced. Thus, the unit costs are reasonable even if only a few thousand devices are required. In many automotive applications space is not a real constraint, so the larger size of a thick-film pressure sensor is not too inconvenient. Thick-film can, in general, match the environmental performance of silicon.

The automotive applications of pressure sensors have broadened from the control of ignition timing to include a wide range of other parameters. Pressure switches are common to warn of loss of engine lubricant or hydraulic failure. In pressurised cooling systems pressure switches are often used to detect the fact that a leak has occurred, before all the coolant is lost. Pressure measurements are often used to control turbo boost pressure and for altitude compensation. In vehicles fitted with active-suspension systems a pressure sensor may be required to control variable-stiffness air springs or possibly adjustable dampers. Some forms of engine knock sensor are pressure based [10]. It appears likely that some form of tyre pressure monitoring will be introduced in the near future.

Efforts will undoubtedly continue to develop better sensors with more inbuilt intelligence. It may be that optical sensors for automotive use in severe electromagnetic environments will appear. However, before this

can happen improvements in the temperature capability of fibre optics are required. In the meantime the trend toward silicon in both its conventional and thick-film forms will undoubtedly accelerate, leading to the appearance of 'smart' sensors, in which the output appears in digital form.

REFERENCES

[1] Ribbens W B 1988 *Understanding Automotive Electronics* (Howard Sams) ISBN 0 672 27064 6

[2] Noltingk B E (ed) 1988 *The Instrumentation Reference Book* (London: Butterworth) ISBN 0 408 01562 4

[3] Doebelin E O 1990 *Measurement Systems Application and Design* 4th edn (New York: McGraw Hill) ISBN 0 07 100697 4

[4] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 47295 0

[5] Turner J D and Pretlove A J 1991 *Acoustics for Engineers* (London: Macmillan) ISBN 0 333 52142 0

[6] Behr M E, Bauer C F and Giachino J M 1981 Miniature silicon capacitance absolute pressure sensor *Proc. Int. Automotive Electronics Conf.* (London: IMechE)

[7] Ko W H 1986 Solid state capacitive pressure transducers *Sensors and Acutators* **10** 303–20

[8] Beards P H 1987 *Analogue and Digital Electronics* (Englewood Cliffs, NJ: Prentice-Hall) ISBN 0 13 032962 2

[9] Oakes J A 1981 A pressure sensor for automotive application *Proc. 3rd Int. Conf. on Automotive Electronics* (London: IMechE)

[10] Mock R and Meixner H 1990 A miniaturised high-temperature pressure sensor for the combustion chamber of spark ignition engines *Eurosensors IV (Karlsruhe, Germany)* (Amsterdam: Elsevier) pp 103–6 ISBN 0 444 75078 9

4

Air Flow Sensors

Information on the air flow into an internal combustion engine is essential if the engine air–fuel ratio and therefore the combustion are to be effectively controlled.

In the older carburettor-type engines, the air flow across the carburettor jet was sufficient to induce fuel into the air stream, and by using jets with different size holes it was possible to adjust the air–fuel ratio to be approximately correct over the hot working range of the engine; additional or variable jets, as in the SU carburettor, being used for operation under conditions of acceleration or cold start.

With the advent of emission controls it rapidly became clear that by using carburettors alone it was not possible to control the air–fuel ratio accurately enough to meet the critical requirements of the exhaust catalyst. The only solution was to use fuel injection together with a method of separately and continuously measuring the mass of air entering the engine.

The first method to be tried made use of a sensor to measure inlet manifold vacuum pressure (described in chapter 3), from which air mass flow can then be calculated from a knowledge of the volumetric flow rate of the engine, the rotational speed and the air density. This method is known as 'speed density' measurement. It is necessary to measure air temperature to calculate air density; the measurement accuracy obtained is adversely affected by pressure variations caused by resonances in the manifold, by the effect of reverse flow conditions in some types of engine and by engine variability in manufacturing and subsequent wear.

For these reasons the direct measurement of air mass flow into the engine manifold has for many years been considered to be a better and more accurate approach, and this chapter will describe five of the air mass flow sensors which have been developed. Of the sensors considered only the hot-wire device measures air mass flow directly, the other devices described measure volume flow or velocity and require the calculations shown in each section to obtain a mass flow measurement.

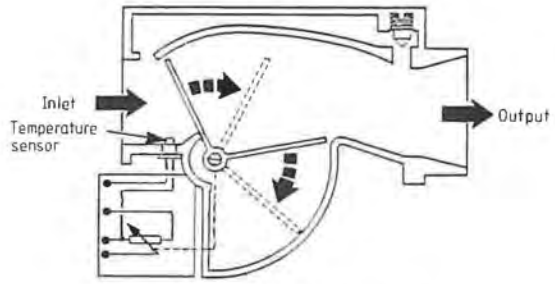

Figure 4.1 Bosch air vane sensor.

4.1 VANE AIR METER

The first and still probably the most widely used device is the vane air flow meter developed by Bosch. In this device a pivoted vane is placed in the airstream (see figure 4.1) and deflects against spring pressure according to air flow. It is coupled to a potentiometer which produces an analogue voltage dependent on vane deflection.

The operation of the vane-type air flow meter is based on the Bernoulli equation and produces an output signal which is proportional to air flow volume and the square root of air density. It is necessary to measure inlet air temperature to enable air density to be determined and air mass flow to be calculated. Two major problems exist with the device; the first of these is that the vane obstructs the air inlet to a degree dependent on air flow rate which causes a reduction in engine efficiency. The second problem is that the device has significant mechanical inertia which leads to slow response of the order of 100 ms to step transients with the result that air–fuel ratio calculation during transients has to be done without adequate information about air mass flow. It has, however, proved to be a very valuable sensor for a whole generation of fuel-injected engines. By 1993, however, the vane-type air flow meter was being progressively replaced by the developed hot-wire mass air flow meter which is described in the next section.

4.2 HOT-WIRE MASS AIR FLOW METER

The hot-wire mass air flow meter relies on the fact that a heated body in an air flow will lose heat at a rate related to the mass and velocity of the air flowing over it and the temperature difference between the wire and

the air. In its original form as developed by Bosch, a wire, usually of platinum, is suspended in the air flow to be measured and heated electrically to a temperature significantly higher than the air temperature. If the temperature difference between the air and the hot wire is maintained constant by controlling the current flow in the hot wire then the following relationship exists

$$I^2 = A + B\sqrt{G_a}$$

where A and B are constants, I is the heating current through the hot wire and G_a is the mass air flow rate. This direct measurement of mass air flow rate is one particular advantage of this method over the vane meter, the other is its fast response, of the order of 1–2 ms for a freely suspended wire and less than 10 ms when the wire is wound or a conductive film deposited on a bobbin as in the version developed by Hitachi.

The first mass air flow meter using the hot-wire principle for engine control [1] used a single platinum wire suspended in the centre of the inlet manifold with an upstream temperature sensor for incoming air (see figure 4.2). The wire was found to be susceptible to surface contamination which modified its heat transfer function and therefore its measurement accuracy, and a strategy was devised in which the wire was briefly heated to a much higher temperature than normal when the engine was started or stopped to 'burn-off' any contamination. This action was found to be effective, but the single exposed wire was also vulnerable to mechanical damage, particularly from engine backfire, and a new arrangement in which the hot wire was wound on a small ceramic bobbin and placed in a by-pass passage in the inlet manifold (figure 4.3) was developed by Hitachi

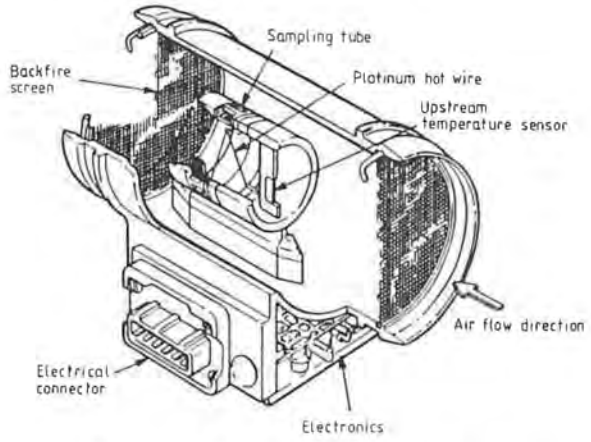

Figure 4.2 Hot-wire sensor (in main inlet)

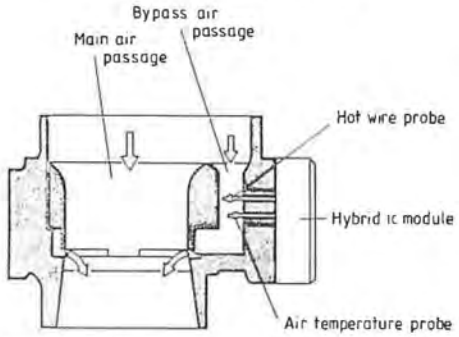

Figure 4.3 Hot-wire sensor (in by-pass passage).

and found to give similar results to the exposed wire, although with some reduction in speed of response to the order of 10 ms.

More recent sensors have been in the form of platinum deposited in a spiral on a tiny ceramic bobbin with a corrosion-proof coating on the surface with the adjacent air temperature sensor made in identical form to give a closely comparable thermal response under transient conditions.

The original circuit arrangement for the mass air flow sensor and the temperature sensor is shown in figure 4.4. The hot-wire resistance R_H forms one arm of a Wheatstone bridge circuit and its temperature is maintained at 100 °C above the ambient temperature of the intake air flow. For a typical hot wire the necessary heating current to maintain this temperature difference varies from 500 to 1200 mA depending on the quantity of intake air flowing. This heating current is then a measure of mass air flow and is measured as a voltage drop across the precision low-temperature-coefficient thin-film resistor R_3. The other half of the bridge consists of the temperature sensor R_K and the relatively high-resistance potential dividers R_1, R_2. Changes in ambient air temperature then change the voltage at point 1 and bridge balance is maintained by using an operational amplifier to compare the voltages at points 1 and 2 and using

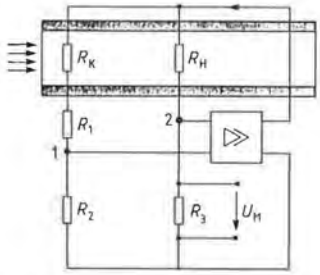

Figure 4.4 Basic bridge circuit used with hot-wire sensor.

the output to control the bridge current through the arms R_H, R_3 and R_K, R_1, R_2. During installation the bridge resistor R_2 is laser trimmed during final testing to calibrate the sensor at a preset air flow level.

In more recent systems the circuit has been further developed by the use of an additional operational amplifier as shown in figure 4.5. With this arrangement, the current through the air temperature sensor can be suppressed so that measurement of air temperature is not affected by self-heating of the sensor.

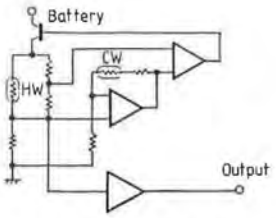

Figure 4.5 Circuit diagram for hot-wire (HW) sensor with air temperature sensor current suppressed (CW = cold wire).

The response curve for a hot-wire system of this type is shown in figure 4.6, and it can be seen that there is some reduction in sensitivity at higher air flow rates. It is important to maintain as near laminar air flow as possible over the hot-wire sensing element if accuracy is to be maintained.

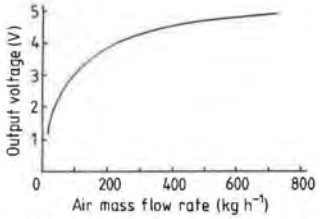

Figure 4.6 Hot-wire sensor operating characteristic.

A further development used in some situations is the use of a voltage–frequency converter to give a digital output from the system. The sensor is designed to operate over an air flow range of 0.0035 to 0.14 kg s^{-1} and an operating temperature range of −40 to +125 °C with an accuracy of better than ±4%; in contrast to the vane air meter this sensor can effectively respond to transient engine conditions.

Future development of the sensor, which is being used in increasing quantities, is towards the use of matched etched silicon devices for air flow and temperature sensing. An example of the configuration is shown in figure 4.7.

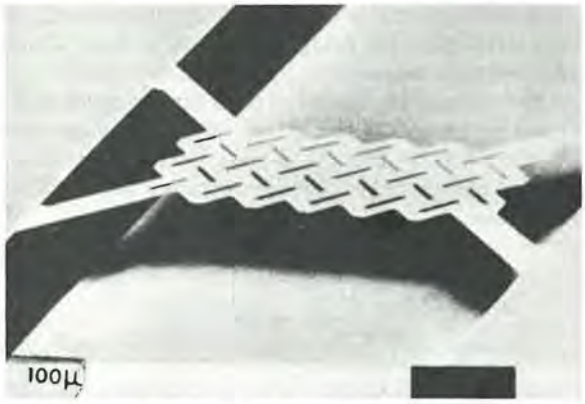

Figure 4.7 Etched silicon hot-wire sensor (courtesy of IC Sensors).

4.3 VORTEX SHEDDING MASS AIR FLOW SENSOR

Other than the vane air and hot-wire sensors, the only other mass air flow sensor used in production engine control systems is that using vortex shedding. Although widely used for plant instrumentation (primarily for liquids), its use as an automotive mass air flow sensor was first described in 1976 [2,3].

The operation of this device depends on the fact that holding a bluff object in a moving air or liquid stream results in the production of vortices downstream from the object. This is illustrated in everyday life by the motion in wind of a flag attached to a flagpole. In this case the vortices are produced alternately behind the flagpole and on either side of the flag and cause the usual flapping motion. As the wind velocity increases the number of vortices produced in unit time also increases and the flapping frequency increases correspondingly. In a vortex air flow sensor the bluff body obstruction produces the characteristic Von Karmen vortex stream shown in figure 4.8. These vortices are usually detected by means of an ultrasonic beam which is transmitted directly across the vortex stream (see figure 4.9), although with a suitable design of bluff body with a trailing edge it is possible to detect them by strain gauging of the bluff body.

The signal received at the ultrasonic sensor of figure 4.9 measures the varying density of the air between itself and the ultrasonic transmitter, and this signal includes not only the signals caused by the passage of the vortices, but also information from which air density can be derived. This parameter is required because the output from a vortex shedding flow meter is a volumetric measure of the air flow and must be corrected to mass flow by calculation using the measured air density. The rotational

Figure 4.8 Vortex street in the wake of a cylinder (hidden by bracket) at Reynolds number of of 5.5. (Reproduced with permission from Dr D J Tritton.)

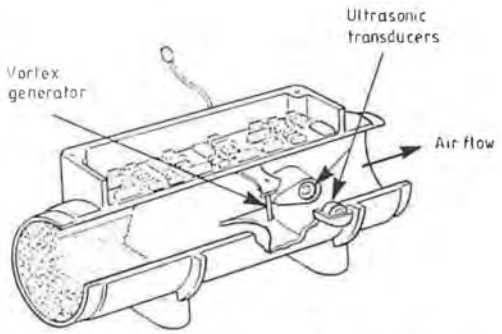

Figure 4.9 Ultrasonic sensor system for vortex measurement.

velocity of the vortices results in dispersion of the ultrasonic beam so that the ultrasonic energy received is modulated in amplitude at the rate at which the vortices are passing through the ultrasonic beam, so giving a direct measurement of air velocity.

The relationship between air velocity and the rate at which vortices are shed, and therefore passing the ultrasonic detectors, is given by the relationship $f = SV/d$, where d is the width of the bluff body and S is the Strouhal number. For a Reynolds number of greater than 700, the Strouhal number is sensibly constant and of the order of 0.21, and it is desirable to operate in this region where the frequency versus velocity response is linear and independent of environmental parameters. In a practical vortex air flow sensor, however, it is necessary to operate down to relatively low flow rates, so that some nonlinearity in the response is unavoidable. A typical response curve is shown in figure 4.10.

In tests by Barnicoat [4], response rates of 10 ms have been measured for a step air flow change of 200 $ft^3\ min^{-1}$ (5.7 $m^3\ min^{-1}$). Barnicoat and co-workers also established that changes in atmospheric pressure have only a minor effect on the accuracy of measurement, although air temperature has some effect at high flow rates due to the change in air viscosity.

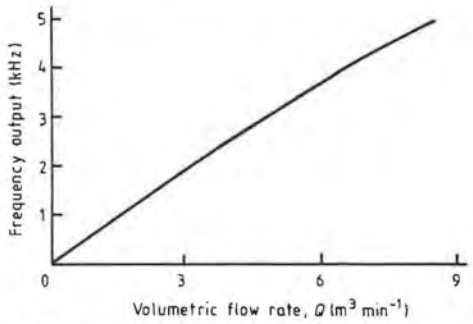

Figure 4.10 Vortex flowmeter response (frequency versus flow rate).

To summarise the characteristics of the vortex shedding flowmeter: it is fast in response, produces a semi-digital (frequency) output signal, does not measure reverse flows, has a fixed and known zero point and is potentially of high reliability and low cost. It has never, however, fulfilled the promise it seemed to offer back in the late 1970s, although it has continued to be used by one Japanese motor manufacturer on production engine control systems.

4.4 CORONA DISCHARGE MASS AIR FLOW SENSOR

A method for measuring mass air flow which causes no obstruction in the flow pipe, has no moving parts, has a very fast response rate and can measure and differentiate flow in either direction, sounds like the answer to an engine control technologist's prayer. The corona discharge mass air flow sensor meets all these criteria, but for other reasons has never achieved the widespread use originally envisaged.

The principle of operation is shown in figure 4.11 [5]. A wire carrying a high voltage of between 5 and 10 kV is suspended along the axis of a cylindrical tube carrying the air flow to be measured. The wire is at a high enough voltage and of a small enough diameter to produce an ionised corona around itself. Surrounding the wire on the inner surface of the tube are cylindrical ring electrodes at ground potential, one upstream and one downstream of the centre of the wire and separated by a small gap on the centre line. If no air is flowing through the tube the corona ions will drift towards these collector ring electrodes under the influence of the potential difference between them and the wire. Because these collector electrodes are equally divided either side of the centre line of the wire, an equal ion

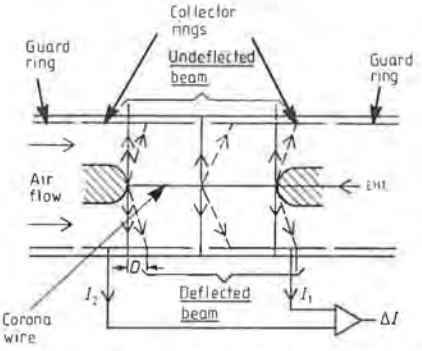

Figure 4.11 Corona flowmeter showing principle of operation.

current will flow in both electrodes if an external circuit is attached. If there is air flow then the beam of ions will be deflected in the direction of flow as shown in figure 4.11 and a differential current $I_1 - I_2$ will exist which can be measured to give a difference current ΔI.

The operating equation of the device relating the mass air flow M to the output difference current I is

$$\Delta I = \frac{I\dot{M}}{Vk}$$

where I is the total current $(I_1 + I_2)$ flowing through the two collector electrodes and V is the voltage applied to the central wire. The factor k in the equation relates the air density to the mobility of the ions, taking into account the fact that the deflection of ion flow is inversely proportional to ion mobility. The value of k therefore varies under the influence of the composition of the air, particularly the presence of water, and the device therefore requires correction for humidity. It should also be noted that other impurities in the air can affect k.

This sensitivity to air composition is the reason that the device has not been acceptable for normal vehicle use; it has, however, found a niche in test cell instrumentation where it is possible to control the air quality used and to make full use of the fast response of the device and its ability to measure reverse as well as forward air flow, this being an important factor in the study of inlet air flow in engine development, particularly at low engine speeds.

The total current $I_1 + I_2$ is normally about 50 μA and the difference current ΔI may be as small as 0.5 μA at low air flow rates. This requires sophisticated high-impedance electronics to make reliable measurements and provides the opportunity to arrange a control loop to maintain the corona wire EHT voltage so that V/I is maintained constant and the corona discharge operates at a constant resistance of about 200 MΩ.

This control system stabilises the operating conditions and can also be used to produce a voltage proportional to total current which can be used to drive the two guard rings shown in figure 4.11 outside the two measuring ring electrodes. This arrangement minimises the leakage of current from the collector ring electrodes.

In summary, therefore, the corona discharge mass air flow sensor has all the features required of an accurate, fast-response, non-obstructive sensor, but its susceptibility to air quality (humidity in particular) makes it unacceptable for in-vehicle use although it fulfils a useful role in development and testing.

4.5 ULTRASONIC MASS AIR FLOW SENSOR

The ultrasonic mass air flow sensor uses the differential effect of air velocity on the time delay of an ultrasonic signal propagated with and against the direction of air flow as its basic principle of measurement.

The configuration of the sensor is shown in figure 4.12. The air is passed through a tube in which are mounted two ultrasonic transducers capable of both transmitting and receiving ultrasonic waves. The two transducers are pulsed with a very short drive pulse which triggers them to produce a short wave train of a period significantly less than the time for sound to travel between the two transducers. Immediately the wave train is complete, both transducers are switched to receive, they each then detect the received signal from the distant transducer and their associated electronics compares the time taken for the two simultaneous pulse trains to travel in opposite directions along the tube.

If there is no air flow through the tube, then the transmission times will be the same. If there is air flow, the velocity of this will add to the transmission time in one case and subtract from it in the other. The value of mass air flow G at the discrete sampling instant k is then described by the equation

$$G_k = C_\beta P_k \Delta t_k \tag{4.1}$$

C_β being a proportionality constant depending on the geometrical configuration of the flow meter modified to allow for the velocity profile across the tube and calculated from the relationship

$$C_\beta = \frac{Sg\alpha_\beta}{2L}$$

where S is the cross-sectional area of the tube, g is the acceleration due to gravity, α is the ratio of specific heat at constant pressure versus specific

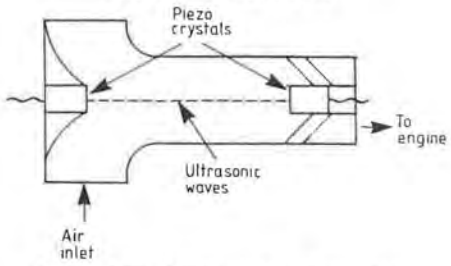

Figure 4.12 Ultrasonic air flowmeter.

heat at constant volume for air (c_p/c_v), L is the distance between ultrasonic transducers and β is a shape coefficient which accounts for variations in velocity profile across the tube and which must be evaluated experimentally. P_k is the mean value of air pressure during the measurement interval and Δt_k is the delay time calculated from the relationship

$$\Delta t_k = \frac{2LW_k}{U^2 - W_k^2}$$

where W_k is the mean value of air velocity in the measurement interval and U is the velocity of sound in air. The linearity of the relationship of equation (4.1) is good when turbulent flow is fully established. In the laminar flow region the variation of the velocity profile causes significant variations in C_β.

Claims have been made for measurement accuracies of air velocity of ±1% over a typical inlet mass air flow range of 0.009 to 0.18 $kg\,s^{-1}$ [6]. Below this bottom limit the onset of laminar flow introduces major inaccuracies and above the top limit measurement becomes unstable because of high turbulence.

Operation is possible over a temperature range from −40 °C to +105 °C provided that the air temperature does not exhibit rapid variations. The measuring tube and its air inlet and outlet arrangements require careful design for satisfactory operation as does the associated electronics, particularly if, in the engine environment, electromagnetic interference with the high-impedance ultrasonic transducers is to be avoided.

4.6 THE FUTURE

The five types of mass air flow sensor described cover the methods used for measurement of automotive mass air flow up to the present. The most

widely used method is probably still the vane air meter but this is being replaced by the hot-wire sensor which is now operating in a significant proportion of internal combustion engine vehicles in Europe, Japan and the US as the emission regulations for vehicles become progressively more stringent. The use of etched silicon sensors, initially as hot-wire devices such as the device shown in figure 4.7, can be expected to increase and micromachining seems likely to be able to produce low-cost, fast-response devices in the future using the deflection of silicon microbeams or the rotation of microturbines for this important measurement.

It is certain that for as long as internal combustion engines continue in use and require to be accurately controlled, the need for mass air flow measurement will continue and the mass air flow sensor will be a critical part of the engine control system.

REFERENCES

[1] Manger H 1981 LH — Jetronic — a new gasoline injection system with a hot wire air mass meter and a microprocessor controlled ECU *Int. Automotive Electronics Conf. (IMechE Conference Publication C180/81)* (London: IMechE) pp 137–41

[2] Cartnell B C and Zeisler F L 1976 Engine mass air flow meter *Society of Automotive Engineers Paper* 760017

[3] Joy R D 1976 Air flow for engine control *Society of Automotive Engineers Paper* 760018

[4] Barnicoat G D, Thorne G A and Joy R D 1979 An automotive mass airflow sensor *Int. Automotive Electronics Conf. (IEE Conference Publication 181)* (London: IEE) pp 150–4

[5] Cockshott C P, Vernon J P and Chambers P 1983 An air mass flowmeter for test cell instrumentation *Int. Automotive Electronics Conf. (IEE Conference Publication 229)* (London: IEE) pp 20–6

[6] Rinolfi R 1977 Engine control system with ultrasonic mass air flowmeter *Society of Automotive Engineers Paper* 770855

5

Temperature Sensors

Temperature is an important parameter for automotive engineers. The first temperature sensors to be applied to motor vehicles were probably used to monitor engine coolant water. However, in modern vehicles they are also used in connection with electronic fuel control systems, to measure the temperature of inlet air and exhaust gas. They are also used for environmental control, to regulate heating or air conditioning in the passenger compartment, and for ice warning systems. The majority of automotive temperature sensors are thermistors, although other forms are used as well.

In addition temperature sensors can be used for liquid level sensing, to control diesel fuel dewaxing, and for de-icing systems such as those supplied to heated exterior mirrors and windscreen washer jets.

Most of the electrical temperature sensors used in automotive engineering are either resistive or thermoelectric. Resistive devices may be either metallic or semiconductor, and require some form of bridge circuit for signal conditioning since they are modulating transducers. Thermoelectric sensors or thermocouples are self-generating, but their very low output means that an amplifier is always needed in practice. In the past bimetallic temperature sensors were very common and are still sometimes used. The thermal expansion of a solid is used as a primitive temperature transducer to control thermostat opening in water-cooled engines. Infrared emission or pyrometry has been proposed a number of times for automotive temperature sensing, and may be particularly useful for monitoring inaccessible or rotating components such as driveshaft joints (where the temperature of the Hooke's joint bearings may give an early warning of failure). However, pyrometric sensors have yet to appear on a production vehicle.

Another temperature-based sensor which has yet to get beyond the research stage as far as automotive applications are concerned is the heat flux gauge. These measure the rate at which heat is being transferred to or from a body, rather than its temperature. These devices have been included and are discussed in section 5.8, since it seems likely that they may find automotive applications in the near future.

The temperature of any system is a measure of its energy. This is best described by the principle of equipartition, which states that for a system in thermal equilibrium with its surroundings the mean kinetic energy per degree of freedom of the particles which go to make up the system is $\frac{1}{2}kT$. In this equation k is Boltzmann's constant and T is the absolute temperature in kelvin (K).

The most direct realisation of a scale of temperature using this approach is based on the ideal gas equation, which can itself be derived directly from the principle of equipartition by means of the kinetic theory of gases [1]. The gas equation $pv = RT$ gives the pressure p of a specific volume v (specific volume = (density)$^{-1}$) of gas at a temperature T, with a constant of proportionality R which is equal to the number of molecules in the specific volume divided by k. The constant-volume gas thermometer, in which the pressure of a fixed mass of an inert gas is measured as a function of temperature, uses this approach. It forms the basis of most temperature measurement schemes. The only fixed point used is the triple point of water, defined as 273.16 K or 0.01 °C by international agreement.

Absolute measurements with gas thermometers are infrequently made, and are usually only undertaken by standards laboratories. For practical purposes a set of fixed points have been agreed which form the basis of the International Practical Scale of Temperature. Table 5.1 shows some of the fixed points. They are chosen for their ease of replication. Interpolation between them can be achieved by a temperature transducer once it has been calibrated with respect to the fixed points.

Although temperature is very apparent to our senses it is less easy to define a meaningful scale for temperature than it is for, say, length. This is because temperature cannot be measured directly but has to be detected

Table 5.1 Reference points used to define a temperature scale.

Temperature (°C)	Reference point	Measurement technique used
−182.97	Boiling point of oxygen	Platinum resistance thermometer
0.01	Triple point of water	
100.0	Boiling point of water	
444.6	Boiling point of sulphur	Platinum thermocouple
960.8	Melting point of silver	
1063.0	Melting point of gold	

by observing the effect it produces, such as the expansion of a liquid, and assumptions have to be made about the linearity of the effect observed.

5.1 RESISTIVE TEMPERATURE TRANSDUCERS

Resistive temperature detectors are probably the most common type for automotive use. They may be either metallic or semiconductor based. The semiconductor versions are the commonest and are probably cheapest. They are sometimes known as *resistance temperature detectors* or RTDs. Metallic resistive temperature sensors offer better performance than semiconductor RTDs and may be preferred if high accuracy is required.

5.1.1 Metallic Resistive Temperature Sensors

These transducers are similar in appearance to wire-wound resistors and often take the form on a non-inductively wound coil of a suitable metal wire such as platinum, copper or nickel. They may be encapsulated within a glass rod to form a temperature probe, which can be very small in size. An alternative form is shown in figure 5.1, in which a rectangular matrix of platinum is deposited on a ceramic substrate. Connections within the matrix are laser-trimmed to give a precise value of nominal resistance.

The variation of resistance R with temperature T for most metallic materials can be represented by an equation of the form

$$R = R_0(1 + a_1T + a_2T^2 + \ldots + a_nT^n) \tag{5.1}$$

where R_0 is the resistance at temperature $T = 0$. The number of terms

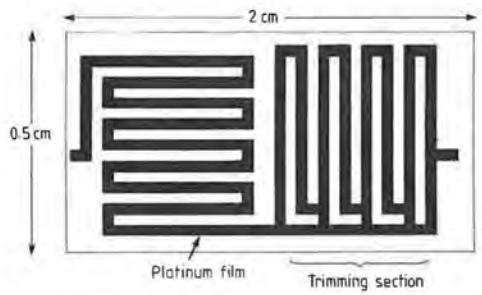

Figure 5.1 Miniature platinum resistance transducer.

necessary in the summation depends on the material, the accuracy required and the temperature range to be covered. Platinum, nickel and copper are the most commonly used metals, and they generally require a summation containing at least two of the a_n constants for accurate representation. Tungsten and nickel alloys are also frequently used. In automotive applications it is often possible to model a metallic RTD using only constant a_1. With platinum for example $a_1 \approx 0.004$ (R in Ω and T in K). If this value is substituted into equation (5.1), and a_2, a_3 etc are set to zero, the resulting nonlinearity is only around 0.5% over the temperature range −40 to +140 °C.

The nominal resistance (R_0) of a metallic RTD can vary from a few ohms to several kilohms. However, 100 Ω is a fairly standard value. The resistance change of a metallic RTD can be quite large, and is typically up to 20% of the nominal resistance over the design temperature range.

5.1.2 Thermistors

Thermistors are small semiconducting transducers, usually manufactured in the shape of beads, discs or rods. They are made by combining two or more metal oxides. If oxides of cobalt, copper, iron, magnesium, manganese, nickel, tin, titanium, vanadium or zinc are used the resulting semiconductor has a negative temperature coefficient (NTC) of resistance. This means that as the temperature rises the electrical resistance of the device falls. Most thermistors used in automotive engineering are of this type, and they can exhibit large resistance variations. Typical values are 10 kΩ at 0 °C and 200 Ω at 100 °C. This very high sensitivity allows quite small temperature changes to be detected. However, the accuracy of a thermistor is not as good as that of a metallic RTD, owing to unavoidable variations in the composition of the semiconductor which occur during manufacture. Most thermistors are manufactured and sold with tolerances of 10 or 20%. Any circuit using semiconductor thermistors must therefore include some arrangement for trimming out the error.

Thermistors are markedly nonlinear (unlike metallic RTDs). The resistance–temperature relation is usually of the form

$$R = R_0 \exp[\beta(1/T - 1/T_0)] \tag{5.2}$$

where R is the resistance (in Ω) at temperature T, R_0 is the nominal resistance at temperature T_0 and β is a constant which is characteristic of the thermistor material. The reference temperature T_0 is usually taken to be 298 K (25 °C), and β is of the order of 4000.

If equation (5.2) is differentiated and rearranged the temperature coefficient of resistance α at temperature T_0 can be found

$$\alpha = \frac{dR/dT}{R} = -\beta/T_0^2 \qquad (5.3)$$

where β is the characteristic temperature constant for the material and T_0 is in K. The units of the temperature coefficient of resistance α are ΩK^{-1}. If β is taken as 4000, a typical value for α for a thermistor at 298 K (25 °C) is $-0.045\ \Omega K^{-1}$.

Thermistors can be used within the temperature range from −60 to +150 °C, which comfortably covers most of the temperature measurements required in automotive engineering. The accuracy can be as high as ±0.1%. The main problem associated with thermistors is their nonlinearity, as expressed by equation (5.2).

Positive temperature coefficient (PTC) thermistors can also be made using compounds of barium, lead or strontium. PTC thermistors are usually only used to provide thermal protection for wound equipment such as transformers and motors. The characteristics of a PTC device have the form shown in figure 5.2. It can be seen that the resistance of a PTC thermistor is low (and reasonably constant) below the switching temperature T_R. Above this point the resistance rises spectacularly. In use PTC thermistors are often embedded in the windings of the equipment to be protected, and are connected in series with the power supply. If the temperature becomes too high the resistance rises and power is effectively disconnected from the load. PTC thermistors are better than NTC types for this sort of thermal protection task since they are fail-safe: if a connection to a PTC sensor fails, the resulting high impedance will disconnect the power. If the same happens to a thermal protection circuit containing an NTC thermistor, a false 'low temperature' indication will be given, and full power will be applied.

'Conventional' thermistors are manufactured as discrete components. However, it is also possible to print thermistors onto a suitable substrate using thick-film fabrication techniques (see chapter 9 for a description of

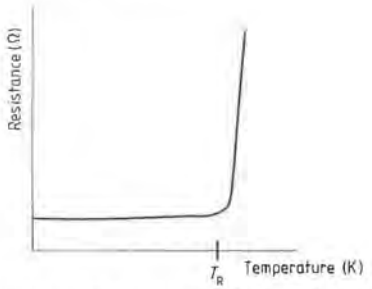

Figure 5.2 Resistance–temperature characteristic for a positive temperature coefficient thermistor.

the thick-film fabrication process). Thick-film thermistors are very low cost and physically small, and have the further advantage of being more intimately bonded to the substrate than a discrete component. It has been shown [2] that thick-film thermistors can have as good if not better performance than a comparable discrete component. No large changes in thermal characteristics are found after trimming.

5.1.3 Resistance Temperature Sensor Bridge Circuits

Thermistors are modulating transducers producing a resistance change. For laboratory work they are normally used in a Wheatstone bridge circuit, which must include some form of bridge balancing arrangement similar to that shown in figure 5.3 (see [3]). Bridge circuits may be dispensed with in less demanding applications. While the resistance changes exhibited by a metallic RTD are reasonably linear, those shown by semiconductor thermistors are markedly nonlinear. In both cases the resistance changes are large. Even if the sensor output is linear, the out-of-balance voltage measured using a bridge circuit is not necessarily linear for large changes in sensor resistance. Take for example the case of a 500 Ω platinum resistance thermometer which exhibits a 100 Ω resistance change over its design temperature range. If the sensor is included in a bridge with four equal arms, the out-of-balance voltage will be very nonlinear as a function of temperature. However, if the fixed resistors R_1, R_2 in figure 5.3 are of considerably higher resistance (about 10 times higher is normal) than R_3 and R_4, and if care is taken to balance the bridge at the middle of its design temperature range rather than at one end, reasonable linearity may be achieved.

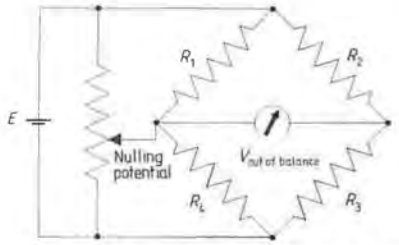

Figure 5.3 Bridge circuit with balancing adjustment.

Resistance thermometer bridges may be excited with either AC or DC voltages. The current through the sensor is usually in the range from 1 to 25 mA. This current causes I^2R heating to take place, which raises the temperature of the thermometer above that of its surroundings and causes a so-called *self-heating error* to occur. The magnitude of this error depends

on the heat transfer conditions (for instance, the conductivity of the surface to which the sensor is attached, and the presence or otherwise of a fluid flow). However, it is not often significant in automotive engineering where an accuracy of ±1 °C is generally sufficient.

An alternative to the classical bridge techniques for conditioning RTD sensors is the four-wire 'ohmmeter' technique shown in figure 5.4. This is widely used with digital data acquisition systems, where any sensor nonlinearity is corrected in the computer software. A precision current source is used, so resistance changes in the two connecting wires L3 and L4 have no effect on the sensor current I_{ex}. A high-impedance voltmeter is used, typically with a FET input of >200 MΩ. This ensures that the currents through L1 and L2 are negligible, as are the lead-wire resistance errors.

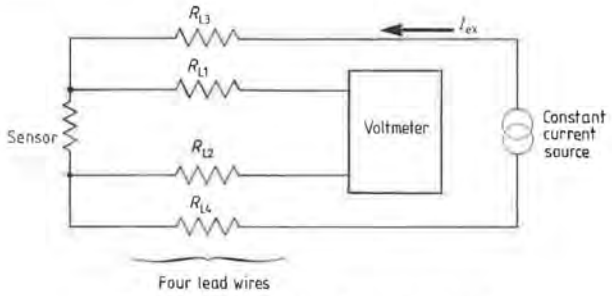

Figure 5.4 Four-wire ohmmeter circuit.

5.2 THERMOCOUPLES

A thermocouple is a self-generating transducer comprising two or more junctions between dissimilar metals. The conventional arrangement is shown in figure 5.5(*a*), and it will be noted that one junction (the 'cold junction') has to be maintained at a known reference temperature, for instance by surrounding it with melting ice. The other junction is attached to the object to be measured.

Thermocouple materials are broadly divided into two arbitrary groups based upon cost. The groups are known as the *base metal* and *precious metal* thermocouples. The most commonly used industrial thermocouples are specified by type letters as shown in table 5.2.

The arrangement of figure 5.5(*a*) is inconvenient because of the layout

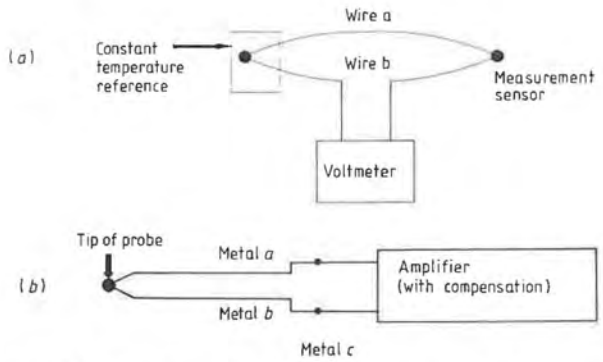

Figure 5.5 Thermocouple circuits: (*a*) standard thermocouple arrangement; (*b*) thermocouple arrangement with cold junction compensation.

Table 5.2 Thermocouples

Type	Conductors (positive conductor first)	Accuracy	Output (mV) for indicated temperature (cold junction at 0 °C)	Service temperature range (°C)
B	Platinum: 30% rhodium alloy Platinum: 6% rhodium alloy	0–1100 ± 3 °C 1100–1550 ± 4 °C	1.24 at 500 °C	0–1500
E	Nickel: chromium/constantan	0–400 ± 3 °C	6.32 at 100 °C	−200–850
J	Iron/constantan	0–300 ± 3 °C 300–850 °C ± 1%	5.27 at 100 °C	−200–850
K	Nickel: chromium/nickel: aluminium (chromel/alumel)	0–400 ± 3 °C 400–1100 °C ± 1%	4.1 at 100 °C	−200–1100
R	Platinum: 13% rhodium/ platinum	0–1100 ± 1 °C 1100–1400 ± 2 °C 1400–1500 ± 3 °C	4.47 at 500 °C	0–1500
S	Platinum: 10% rhodium/ platinum	As type R	4.23 at 500 °C	0–1500
T	Copper/constantan	0–100 ± 1 °C 100–400 ± 1%	4.28 at 100 °C	−250–400

Type B: best life expectancy at high temperatures.
Type E: resistant to oxidising atmospheres.
Type J: low-cost, general-purpose.
Type K: general-purpose, good in oxidising atmospheres.
Type R: high-temperature, corrosion resistant.
Type S: as R.
Type T: high resistance to corrosion by water.

of leads and the need for a reference temperature. A more practical scheme is shown in figure 5.5(*b*). The two wires are laid out side by side, and are connected to a voltage measuring circuit. The junctions between the two wires and the voltmeter do not cause any error signal to appear so long as they are at the same temperature. Since there is no proper reference junction with this approach, the system is liable to give an erroneous output if the temperature of the surrounding environment changes. This is avoided by the use of so-called *cold junction compensation* (see later) in which the characteristics of the signal conditioning amplifier are modified by including a thermistor (see section 5.2) in the circuit.

The arrangement shown in figure 5.5(*b*) is almost universally applied whenever thermocouples are used. The two wires are often enclosed within a tube or flexible sleeve of stainless steel or copper for protection, although this increases the time constant of the system.

The main advantages of thermocouples are their wide temperature range, nominally from −180 to +120 °C for a chromel/alumel device, and their linearity. They are more expensive than thermistors and are therefore mainly used for prototype and experimental work rather than for production vehicles. Table 5.2 gives the characteristics of some of the most common commercially available devices.

If a short section of tubing is made of butt-welded thermocouple materials and inserted into a pipeline, the temperature of a fluid flowing inside the pipe may be measured non-intrusively. Ready-made sensors of this type are available commercially [4], and are often used in automotive engineering for experimental work involving the measurement of fuel and coolant temperatures.

5.2.1 Thermocouple Compensation

As noted earlier, it is not normally practical to have thermocouple cold junctions maintained at a controlled reference temperature. However, with the cold junctions at ambient temperature, which may change, some form of cold junction compensation is required. Consider the arrangement shown in figure 5.6, which shows a thermocouple with its measuring junction at t °C and its cold junction at ambient temperature. The thermocouple output is $E_{(t-i)}$, but what is required is the output that would obtain if the cold junction were at 0 °C, i.e. $E_{(0-i)}$. Thus a voltage $E_{(0-a)}$ must be added to $E_{(t-i)}$ to correct the output signal

$$E_{(0-i)} = E_{(t-i)} + E_{(0-a)}. \tag{5.4}$$

The voltage $E_{(0-a)}$ is called the cold junction compensation voltage, and it

is provided automatically by the circuit of figure 5.6 which includes a thermistor R_t. R_1, R_2 and R_3 are temperature-stable resistors. The bridge is first balanced with all the components at 0 °C. As the ambient temperature is changed away from 0 °C an imbalance voltage will appear across AB. This voltage is scaled by selecting R_t such that the imbalance voltage across AB equals $E_{(0-a)}$ in equation (5.4).

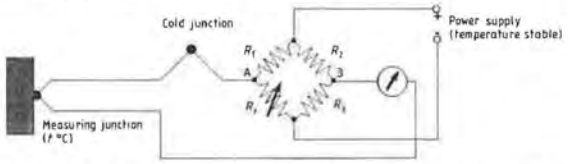

Figure 5.6 Bridge circuit with cold junction compensation.

5.2.2 Multiple Thermocouple Arrangements

Several thermocouples may be connected in series or parallel as shown in figure 5.7 to achieve useful functions. The series arrangement (figure 5.7(*a*)) is used mainly as a means of enhancing sensitivity. All the measuring junctions are held at one temperature, and all the reference junctions at another. This arrangement is often called a *thermopile*, and for n thermocouples gives an output n times as great as that which can be obtained from a single couple. A typical commercially available chromel/constantan thermopile has 25 junctions and produces about 0.5 mV °C^{-1}.

The parallel arrangement shown in figure 5.7(*b*) generates the same temperature as a single couple if all the measuring and reference junctions are at common temperatures. If the measuring junctions are at different temperatures and the thermocouples all have the same resistance, the output voltage is the average of the individual voltages. The temperature

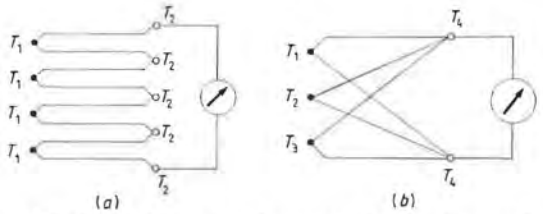

Figure 5.7 Multiple-junction thermocouples: (*a*) thermopile; (*b*) averaging arrangement.

corresponding to the output voltage is the mean temperature (as long as the thermocouples are linear over the measurement range).

5.3 BIMETALLIC TEMPERATURE SENSORS

Bimetallic strips are made by bonding together two metals with different coefficients of thermal expansion. Typical materials used are brass and Invar. As the temperature of the bimetallic component changes the brass side expands or contracts more than the Invar, resulting in a change of curvature.

If two metal strips A and B with coefficients of thermal expansion α_A and α_B are bonded together as shown in figure 5.8, a temperature change will cause differential thermal expansion to occur. If it is unrestrained the strip will deflect to form a uniform circular arc. Analysis [5] gives the radius of curvature ρ as

$$\rho = \frac{t[3(1+m)^2 + (1+mn)(m^2 + 1/mn)]}{6(\alpha_A - \alpha_B)(T_2 - T_1)(1+m)^2} \tag{5.5}$$

where ρ is the radius of curvature, t_A, t_B are the thicknesses of strips A and B, t is the total strip thickness, n is the ratio of elastic moduli, E_B/E_A, m is the thickness ratio, t_B/t_A, and $T_2 - T_1$ is the temperature rise. In most practical cases $t_B/t_A \approx 1$ and $(n+1)/n \approx 2$, giving

$$\rho \approx \frac{2t}{3(\alpha_A - \alpha_B)(T_2 - T_1)}. \tag{5.6}$$

Equations (5.5) or (5.6) can be combined with the appropriate strength of materials expressions to calculate the deflections of most shapes of bimetallic component, or of the forces developed by partially or completely restrained structures.

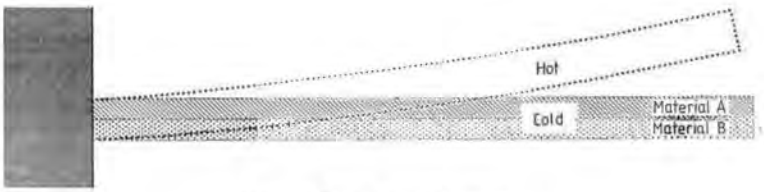

Figure 5.8 Bimetallic strip.

Bimetallic devices are used for low-cost temperature sensing, and as 'on–off' elements in temperature control systems. They are often used as overload cut-out switches for electric motors. Their use in combination with a conductive plastic potentiometer has been reported for an automotive temperature sensing application [6].

5.4 PN JUNCTION SENSORS

PN junctions in silicon have become popular as temperature sensors due to their very low cost. Figure 5.9 shows the forward bias characteristic of a silicon diode. It is well known that a voltage V_f has to be applied across the junction before a current will flow. For silicon V_f (which is often termed the *diode voltage drop*) is of the order of 600–700 mV. V_f is temperature dependent, and is very nearly linear over the temperature range from −50 to +150 °C. This covers the range of temperatures usually required in automotive engineering. The voltage V_f has a temperature characteristic which is essentially the same for all silicon devices of about −2 mV °C.$^{-1}$

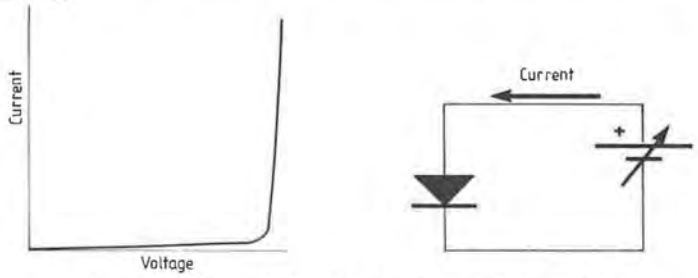

Figure 5.9 Forward bias characteristic of a silicon diode.

A typical circuit is shown in figure 5.10. Transistors are often used instead of diodes, with the base and collector terminals connected together. For the best performance a constant current should flow through the diodes, but in practice the error incurred by driving the circuit from a constant voltage is small.

There is one major disadvantage to using diodes as temperature sensors in control applications, which is that they are not fail-safe. If a diode temperature sensor is used to control a heater, any breakage of the diode wires will be interpreted by the controller as low temperature. More power will then be applied to the heater, resulting in an uncontrolled runaway.

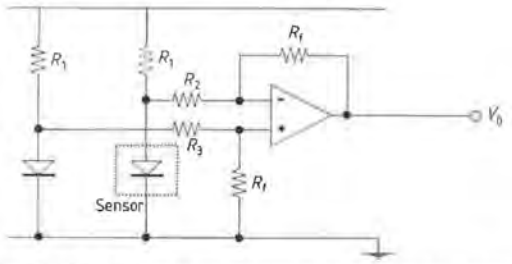

Figure 5.10 PN junction temperature transducing circuit.

5.5 LIQUID CRYSTAL TEMPERATURE SENSORS

A number of liquids (mainly organic) can be made to exhibit an orderly structure in which most or all of the molecules are aligned in a common direction. The structure can be altered by electric or magnetic fields. Most people are familiar with the liquid crystal displays used in watches and calculators which use compounds sensitive to electric fields. Less well known, however, is the fact that some liquid crystal materials are temperature sensitive [5]. One application which may be familiar is the use of cholestoric compounds as medical thermometers, in the form of a flexible plastic strip which is pressed against the skin to measure its surface temperature. The compounds involved have molecular structures similar to cholesterol, and for this reason are called 'cholestoric'. Cholestoric liquids form a helical structure, and this gives them their optical properties. The plane of polarisation of light passing through the compound is rotated by as much as 1000° per millimetre of path length.

The structure can be enhanced by confining the cholestoric liquid between two parallel sheets of a suitable plastic. The choice of polymer for the plastic is based on two requirements. These are first that it must be optically transparent, and second that it is sufficiently chemically active to bond to the liquid crystal molecules adjacent to the polymer and maintain their axes in the correct orientation.

When used for temperature measurement the liquid crystal is confined between two polymer sheets as described above, a few tens of micrometres apart. The surface of one plastic sheet is given a reflective coating as shown in figure 5.11. In figure 5.11(*a*) a light ray enters the sandwich and is reflected back. Since the liquid crystal is in its ordered state the reflected ray interferes destructively with the incident light, and the sensor appears opaque.

In figure 5.11(*b*) the liquid crystal has been raised to a temperature at which its ordered structure has broken down due to thermal agitation. The

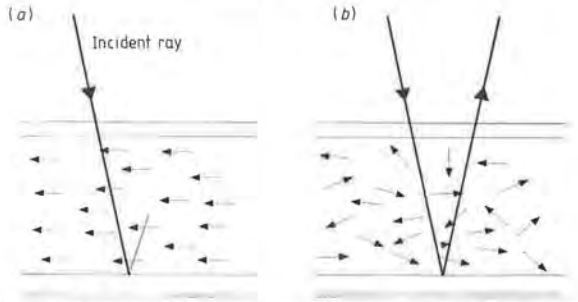

Figure 5.11 Liquid crystal temperature sensing: (*a*) destructive interference; (*b*) no interference.

temperature at which this occurs depends on the exact molecular structure. It is reasonably easy to create compounds tailored to specific temperature changes. If a ray of light enters the sandwich in this condition it is reflected back in the usual way, and the sensor appears to be transparent (the colour of the backing material is usually visible).

If polarised light is used this technique can detect temperature changes as small as 10^{-3} °C. If ordinary white light is used the resolution is of the order of 0.1 °C. The advantages of liquid crystal temperature sensors are that they are relatively cheap and immune to electromagnetic interference. This last characteristic may make them suitable for automotive applications where RFI problems are too severe for a conventional electronic sensor to be used. It is possible to interrogate a liquid crystal temperature sensor remotely, by means of a fibre optic link.

5.6 INFRARED EMISSION AND PYROMETRY

Most people are familiar with the fact that the amount and the wavelength of radiation emitted by a body are functions of its temperature. This dependence on temperature of the characteristics of radiation is used as the basis of a non-contact temperature measurement technique in which the sensors used are known as *radiation thermometers*. Automotive applications of these devices have been reported, for example in tyre condition monitoring (see [6]).

The total power of radiant flux of all the wavelengths R emitted by a black body of area A is proportional to the fourth power of the temperature of the body in kelvin

$$R = \sigma A T^4 \tag{5.7}$$

where σ is the Stefan–Boltzmann constant, which has the value $5.670\,32 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$. Most radiation thermometers are based on this law, since if a sensing element of area A at a temperature T_1 receives radiation from an object at temperature T_2, it will receive heat at a rate $\sigma A(T_2)^4$ and will emit heat at a rate $\sigma A(T_1)^4$. The net rate of heat gain is therefore $\sigma(T_2^4 - T_1^4)$. If the temperature of the sensor is small compared with that of the source T_1^4 may be neglected in comparison with T_2^4.

The above discussion applies to perfectly black bodies with an emissivity ε of unity. 'Real' objects have non-unity emissivities, and a correction must be made for this. The total radiant flux emitted by an object of area A and emissivity ε is

$$R = \sigma \varepsilon A T^4. \tag{5.8}$$

The flux R is equal to that emitted by a perfect black body at a temperature T_a, the apparent temperature of the body

$$R = \sigma A T_a^4. \tag{5.9}$$

Equating (5.8) and (5.9) gives

$$\sigma \varepsilon A T^4 = \sigma A T_a^4$$

therefore

$$T^4 = T_a^4/\varepsilon$$

and therefore

$$T = T_a/\sqrt[4]{\varepsilon}.$$

Radiation thermometers generally consist of a cylindrical body made from aluminium alloy or plastic. One end of the body carries a lens, which focuses energy from the target onto a detector within the tube. The lens may be made from glass, germanium, zinc sulphide, sapphire or quartz, depending on the wavelength. The heat detectors within the instrument are generally thermocouples, thermistors or PN junctions.

5.7 SOLID AND LIQUID EXPANSION TEMPERATURE SENSORS FOR AUTOMOTIVE USE

All water-cooled engines control the coolant circulation to avoid over-cooling. This is often done quite crudely, using a temperature-sensitive

valve called a *thermostat*. It seems likely that in the near future automotive engine cooling will be placed under more sophisticated control, perhaps by means of a variable-speed electric drive to the water pump and one or more electronic temperature sensors placed around the engine [7].

Although thermostats are not strictly temperature sensors they are worth including here, since they can be considered to be temperature transducers (transducing temperature to displacement). There are two kinds of thermostat in common use, the *bellows* and the *wax element* types.

5.7.1 Bellows Thermostats

The operating element is a sealed flexible metal bellows, partly filled with a liquid which has a lower boiling point than that of the engine coolant fluid. Alcohol, ether or acetone are commonly used. Air is excluded from the bellows, which contains only the liquid and its vapour. The pressure inside the bellows is therefore the vapour pressure of the liquid. This varies with temperature, being equal to atmospheric pressure at the boiling point of the liquid. The pressure is less below boiling point, and more at higher temperatures. The extension of the bellows is thus temperature dependent, and it is this which is used to control the coolant flow. A poppet-type valve is attached by a stem to the top of the bellows as shown in figure 5.12(*a*), and varies a circular opening in the flange to control the flow of water.

Complete sealing is not necessary and indeed is undesirable, since it would lead to air being trapped beneath the valve when the system is filled. A small hole is normally drilled in the valve to act as a vent. A loosely fitting *jiggle-pin* prevents this hole becoming clogged.

The *free length* of the bellows is such that when internal and external pressures are equal, the valve will remain open. Thus if the bellows develops a leak the valve will remain open, i.e. the thermostat is fail-safe.

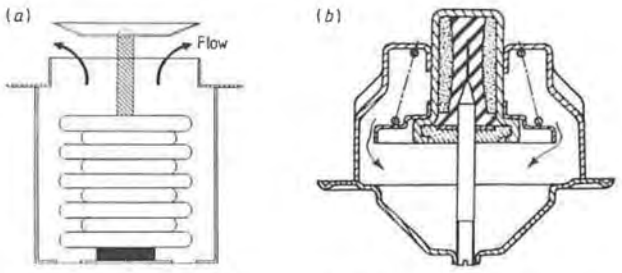

Figure 5.12 (*a*) Bellows thermostat. (*b*) Wax element thermostat.

5.7.2 Wax Element Thermostats

This type depends for its operation upon the considerable change in volume which occurs when certain types of wax melt. The operating element is a metal cylinder filled with such a wax, into which is inserted a thrust pin as shown in figure 5.12(*b*). A flexible rubber sleeve surrounds the pin and is sealed to the top of the capsule to prevent the wax escaping. A poppet valve is again used, and is held in the shut position by a spring when cold. When the thermostat is heated the wax melts and expands, pushing the thrust pin out of the capsule, and opening the valve.

The useful life of this type of thermostat is claimed to be in excess of 100 000 km. The opening temperature tends to increase, however, due to deterioration of the rubber sleeve.

Failure can take place in two ways. The most common mode is for a leak to develop, allowing the wax to escape and the valve to remain closed. The vehicle then overheats, so this device is not as fail-safe as the bellows type. Failure can also occur due to a leak developing in the rubber sleeve below the thrust pin. In this case the valve will stick open.

A vent hole and jiggle-pin are also provided for the reasons given in section 5.7.1.

5.8 HEAT FLUX GAUGES

It is sometimes necessary to make local measurements of convective, radiative or total heat transfer rates. In automotive engineering this is most likely to occur in engine and cooling system research. This requirement has led to the development of a class of sensors known as heat flux gauges.

One common type of heat flux gauge has the general form shown in figure 5.13. Two temperature measuring elements (usually thermocouples) are physically separated by a thermal insulator with known characteristics. When heat energy begins to pass through surface A, the thermocouple J1 generates a small voltage. Since the heat has to pass through the thickness of insulating material I to reach the second thermocouple on surface B, a different voltage is generated by J2. The differential voltage developed across J1 and J2 is proportional to their temperature difference. If the characteristics of the insulator I are known, the heat transfer rate may be obtained as a function of voltage.

Heat flux gauges of this type are fabricated as a string of thermocouples on a flexible backing. Up to 30 thermocouple junctions are placed on each

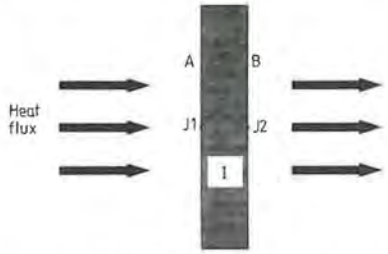

Figure 5.13 Heat flux gauge.

side of the insulator to increase the sensitivity of the device as discussed in section 5.2.2. Heat flux gauges of the type described in figure 5.13 somewhat resemble strain gauges, and are bonded to the surface to be measured in a similar fashion. However, strain gauges are usually considered to be disposable since they are relatively cheap. Heat flux gauges currently cost from £150–£300, although the price is falling, and it is usual to try and remove them for reuse. A number of successful removal techniques have been reported [8].

5.9 SUMMARY AND CONCLUSIONS

The types of temperature sensor required in the automotive industry may be divided into two categories: those used for research and development, and those fitted to production vehicles. The former category encompasses almost all of the temperature measurement methods available, and a complete treatment is beyond the scope of this book, although descriptions of the most commonly used devices and techniques have been given. The second class is of more interest to the automotive engineer since temperature measurement is essential for a number of control functions, notably that of electronic engine management, where inlet air temperature is required.

Up to about a decade ago most of the temperature sensors used on road vehicles were of the bimetallic variety. This has gradually changed, and nowadays thermistors are much more common. These are usually of the bead type, although it seems likely that thick-film thermistors will become standard because of their low cost and the fact that they can be integrated with other circuitry in the form of a thick-film hybrid.

A number of emerging applications such as tyre condition and propshaft

joint monitoring may require non-invasive temperature measurement techniques, and infrared emission sensors may be the most suitable for this purpose.

Silicon diode junctions exhibit useful thermal characteristics as discussed in section 5.4. These are likely to become common in applications where temperature measurement is required as part of an integrated silicon device. This might be the case in a micromachined silicon pressure sensor, where pressure and temperature measurements are frequently required together. If the expense of designing a custom silicon device is considered to be worthwhile, temperature sensing in the form of a PN junction can be included at very little extra cost.

As the amount of electronics on a vehicle increases, the requirement for temperature sensing is also likely to increase. This may be because temperature data are required to control a mechanical function. It seems more likely, however, that temperature measurements will be increasingly needed to provide thermal compensation of the electronic systems themselves.

REFERENCES

[1] Rogers G F C and Mayhew Y R 1983 *Engineering Thermodynamics, Work and Heat Transfer* 3rd edn (London: Longman) ISBN 0582 30500 4

[2] Dargie P N 1991 Investigation of the properties of thick-film thermistor pastes *USITT Report* U104-2 (available from USITT, University of Southampton, Southampton SO5 9NH, UK)

[3] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 44551 1

[4] Doebelin E O 1990 *Measurement Systems Application and Design* 4th edn (New York: McGraw Hill) ISBN 0 07 017338 9

[5] Moffat R J 1990 Experimental heat transfer *Proc. 9th Int. Heat Transfer Conf., 1990* (London: Hemisphere Publishing) pp 187–206 ISBN 0 89116 909 1

[6] Hutchinson M 1989 In infra-red tyre condition monitor *7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 271–90 ISBN 0 85298 697 1

[7] McBride J W and Reed M J 1989 An electrically-driven automotive coolant pump *Proc. 7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 113–20 ISBN 0 85298 697 1

[8] Burton M J and Bowen R J 1988 The use of thin-foil heat flux gauges to determine plug closure in cryogenic pipe freezing *Proc. 12th Int. Cryogenic Engineering Conf. (ICEC 12)* (London: Butterworth) pp 369–75 ISBN 0 40 80125 95

6

Combustion Sensors

Engine control systems have always required some method of measuring what is happening in the engine to enable control to be applied and the engine to be constrained to operate as required by the driver, and more recently by the regulatory authorities.

In early engines control of air, fuelling and ignition was largely manual, with ignition timing being set by hand from the driving position and the fuelling range being defined by the choice of jet size in the carburettor. Automatic mechanical control of ignition advance appeared in the 1920s with the distributor being mechanically rotated by centrifugal weights to increase ignition advance with increasing speed, and by a lever system connected to a pressure sensing diaphragm which sensed inlet manifold vacuum pressure (a good measure of engine load).

The arrival of electronic engine control in the 1960s, and the requirement, initially in the USA, to meet legally regulated emission levels, led to much more sophisticated control of fuelling and ignition timing based on the electronic measurement of inlet manifold mass air flow, engine speed and timing position. This required the use of look-up tables to relate these measured values at the input to the engine to the required fuelling and timing for 'best performance' as established on an engine test bed. This arrangement, however, relied entirely on the accuracy of the look-up tables and took no account of deterioration of the engine with age, changes of fuel quality, the variation between the test-bed engine and production engines or the difference between individual cylinders in the same engine. It was an open-loop system with no capacity to adapt.

The advent of more severe emission controls which required the use of exhaust catalysts made it essential to use feedback control capable of maintaining the air–fuel ratio at stoichiometry, the sensor for this being located in the exhaust system to measure exhaust gas oxygen and hence air–fuel ratio (see chapter 10). The exhaust gas oxygen (EGO) sensor is, however, slow in response (>35 ms) and only capable of measuring average exhaust gas oxygen content; and in its position in the exhaust manifold it can only measure the average output from all cylinders

combined. There is therefore an urgent need for a method of sensing much more rapidly the effectiveness and efficiency of the combustion process, and this can best be done by a sensor within the combustion chamber capable of making direct measurements on the combustion process. The sensor must, however, be of low cost, since for optimal engine control a sensor is required in each cylinder of the engine.

This requirement for low cost has to some extent determined the sensing methods which have been investigated for this application.

6.1 PRESSURE SENSORS

The first sensor to be used to investigate in-cylinder combustion was a pressure sensor using the piezoelectric effect. Originally, during the 1940s, expensive precision instrumentation sensors using quartz sensing elements were used; these became widely used for engine development after the invention of the charge amplifier in 1950 [1]. These sensors are capable of accurately measuring cylinder pressure over the whole range of engine operation, and provide detailed information on the variation of cylinder pressure and its timing during the engine cycle. They can also provide information on the knock, misfiring and abnormal combustion conditions which can seriously affect engine efficiency and smoothness.

For engine control the critical parameters are pressure amplitude and how the pressure variations relate in time to the engine timing cycle. The feedback of this information to the engine control system makes it possible to accurately control the engine to optimise power output while keeping exhaust gas emissions at a level which can be handled effectively by catalyst-equipped engines.

The response time of an in-cylinder pressure sensor should also be fast enough to provide adequate feedback control within one cycle for each cylinder. This is much faster than is possible using currently available exhaust gas oxygen sensors for the feedback control of air–fuel ratio.

A suitably developed powertrain management system using cylinder pressure feedback from each cylinder could be used to control spark timing, exhaust gas recirculation (EGR) rate, fuel injector pulse width and timing, throttle angle and transmission load matching. Such a system would provide excellent diagnostic information which could be used both for modification of engine operating conditions to permit continued operation under fault conditions, and for driver warning and servicing diagnostics.

It is not at present certain whether, when in-cylinder pressure sensing is

used, it will still be necessary to have an EGO sensor to maintain the accurate average stoichiometric air–fuel ratio essential to permit optimal exhaust catalyst operation. Certainly there will be no requirement for a knock sensor to detect over-advanced ignition timing as this is effectively detected by a cylinder pressure sensor, even at high engine speeds where the conventional knock sensor has major problems due to the high general vibration level.

A low-cost design developed by Anastasia and Pestana [2] is shown in figure 6.1. In this a hermetically welded, high-temperature-alloy diaphragm specially designed to minimise stress and deflection during cycling makes a flush fit at the end of the sensor which screws into the cylinder. This flush fit with the cylinder wall reduces the effect of the sensor on the combustion gas mixing to a minimum and seals the sensor from the combustion chamber. A ceramic pin is used to transfer the proportional deflection of the diaphragm under combustion pressure to the piezoelectric transducer element; in this way the piezoelectric device is isolated from the high combustion temperatures and the diaphragm is also rigidly supported against over-deflection and stress. The charge from the piezoelectric transducer element is amplified by a charge amplifier and electronic signal conditioner mounted within the sensor casing, and the signal processed to produce a voltage output proportional to the instantaneous cylinder pressure. The impedance of the signal path is also reduced to permit the signal to be sent to the electronic control unit without corruption by electromagnetic interference.

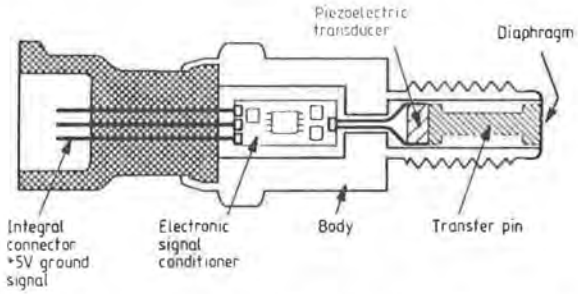

Figure 6.1 In-cylinder pressure sensor (Texas Instruments).

An alternative design has been described by Mock and Meixner [3] which uses a radially polarised piezoceramic tube (figure 6.2) instead of a conventional piezoelectric transducer. In this case the cylinder pressure is transmitted axially to the tube by a flush diaphragm at the cylinder end of the sensor which seals the sensor from the combustion chamber. The inner and outer walls of the tube are lined with electrodes to pick up the charge

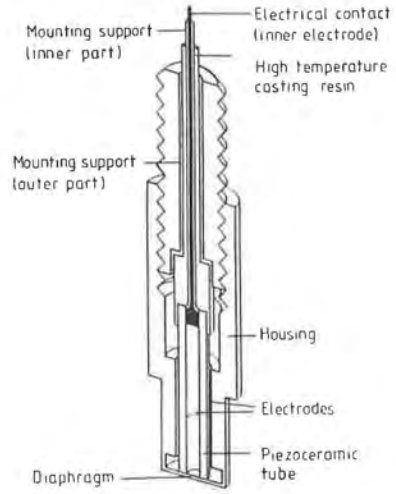

Figure 6.2 In-cylinder pressure sensor using radially polarised piezoceramic tube (from Mock and Meixner [3]).

signal generated by the transverse piezoelectric effect. In this design, with diaphragm temperatures up to 400 °C, there is some deterioration of the piezoelectric properties of the portion of the tube close to the diaphragm as the temperature goes above the approximately 200 °C Curie point. However, since a maximum of only half the tube is affected by this deterioration, the still-active part of the tube continues to provide a charge signal of about 200–300 $pC\,bar^{-1}$. Because of this high sensitivity compared with the conventional piezoelectric transducer, it is claimed that no electronics is required within the sensor and that signal processing may be done under low-temperature conditions away from the engine.

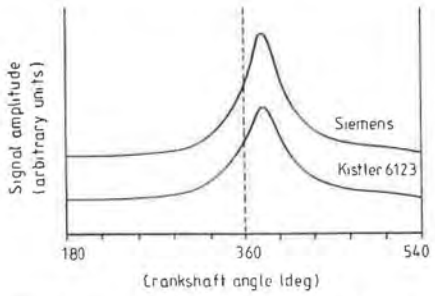

Figure 6.3 Signal output from piezoceramic tube pressure sensor compared with instrumentation pressure sensor output (Siemens).

The typical signal output from this low-cost piezoelectric flush diaphragm sensor is shown in figure 6.3 compared with that from a Kistler quartz instrumentation pressure sensor, with a similar comparison under knocking combustion in figure 6.4.

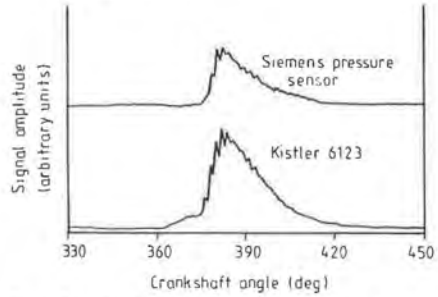

Figure 6.4 Signal output from piezoceramic tube pressure sensor compared with instrumentation pressure sensor output under engine knock conditions (Siemens).

A further alternative sensor has been described by Sasayama *et al* [4]. This makes use of a 0.3 mm × 4 mm flush silicon diaphragm in the combustion chamber and uses optical rather than piezoelectric methods to measure its deflection under pressure (figure 6.5). In this arrangement light from an LED is injected down half of a bundle of 400 multimode optical fibres; this produces small cones of light on the polished silicon diaphragm as shown in figure 6.5. The other half of the fibre bundle is used to transmit the light reflected to a photosensor. As the diaphragm is

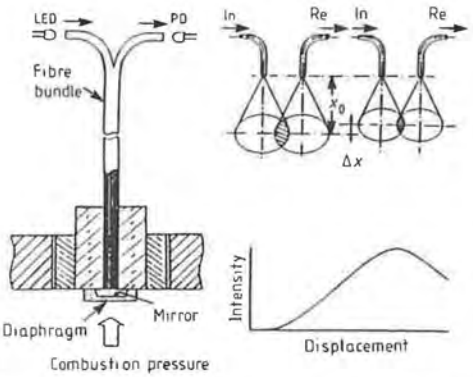

Figure 6.5 Optically sensed diaphragm pressure sensor (from Sasayama *et al* [4]): PD, photodiode; In, input; Re, reflected, x_0, zero position of diaphragm; Δx, deflection of diaphragm under pressure.

deflected towards the optical fibre under increased pressure, the overlap between the transmitting and receiving optical cones changes and the light reflected and reaching the photosensor also changes, giving the curve of intensity versus displacement shown in figure 6.5.

With the distance between the end of the optical fibres and the diaphragm optimised, a high sensitivity is possible. The characteristic resonance frequency of the diaphragm is more than 50 kHz and temperatures up to 500 °C can be handled, but compensation for temperature changes is required if pressure is to be accurately measured.

Finally, a sensor can be used which consists of a piezoelectric ceramic ring which can be clamped under the spark plug in place of the normally used washer (as shown in figure 6.6). Such a sensor was described by Kondo *et al* [5] and also by Randall [6]. The small movements of the spark plug under cylinder combustion change the compression of the ring and produce a voltage signal which can be processed as required.

In his work on this type of sensor Randall used the piezoelectric ceramic PZT-5a which is a composition of lead, zirconia and titanium. This ceramic has a Curie point of 365 °C and in ring transducer form a mechanical resonant frequency of over 2 MHz. The factor determining the frequency response of the sensor is therefore that of the spark plug. Vibrating as a free body its mechanical resonance has been calculated to be about 90 kHz; however, when installed in the engine significant damping is to be expected which would cause a reduction in sensor output at higher frequencies. The sensor should have adequate frequency response to

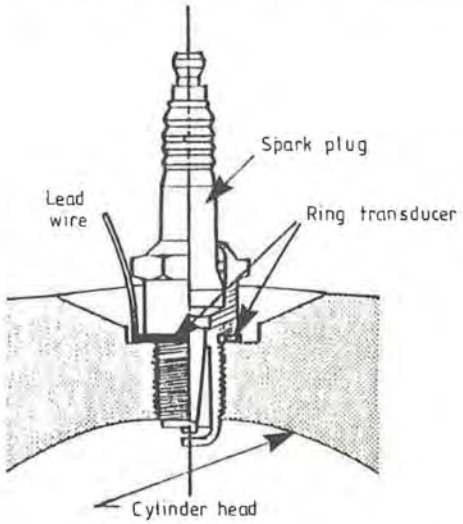

Figure 6.6 Spark plug washer ring pressure sensor.

detect detonation frequencies up to 15 kHz. There is a loss of sensitivity when the sensor is heated to the 200–300 °C temperature range found at the spark plug thread under high-speed engine operation and this has to be allowed for in the associated amplifier circuitry.

If possible, sensors with flush diaphragms at cylinder wall level are preferred to spark plug ring sensors because of the risk of unexpected resonances in the transmission of the pressure pulse to the spark plug and the vulnerability of the spark plug ring sensor to damage by incorrect spark plug tightening. However, the convenience of its application makes the spark plug ring sensor popular for engine development use.

6.2 OPTICAL SENSORS

A cylinder pressure sensor which uses optical methods to detect the deflection of a diaphragm was described in the previous section, but it is also possible to use optical methods to look directly at the combustion process as it happens. The method has a number of attractions: several parameters can be measured directly and could be easily multiplexed on multicylinder engines and the sensor system is immune from the very high level of electromagnetic interference present in the engine environment. It also has a number of disadvantages, chief among which is the need to provide optical access for a narrow-bore probe. This can either be provided by specially drilled holes, where this is possible, or by the use of a silicon rod down the centre of a spark plug.

Most experimental work has been done with specially drilled holes [7] which were designed to take both optical sensors and interchangeable piezoelectric pressure sensors to permit comparative measurement of optical radiation and cylinder pressure.

A silicon or silicon-clad glass rod is used as the initial transmission medium from the cylinder, and this is interfaced with an optical glass fibre bundle immediately outside the engine block which is then led to a large-area silicon photodiode in a cooler part of the engine compartment. Figure 6.7 shows the optical intensity of combustion measured in a gasoline engine compared with a cylinder pressure signal from a piezoelectric sensor mounted in an adjacent access hole. The optical signal is seen to be present only when fuel burning is actually in progress, while the pressure signal is also affected by the pressure changes caused by motoring the engine during non-combustion periods, but otherwise the characteristics of optical radiation and pressure during combustion are closely comparable. Figure 6.7 also shows that the rapid in-cylinder pressure variations (ringing) caused by 'knock', which is seen to be present as an oscillatory

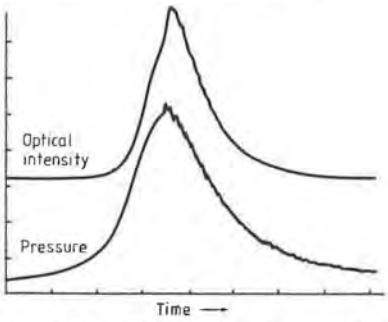

Figure 6.7 In-cylinder optical intensity variation during the combustion cycle compared with cylinder pressure variation (from Pinnock *et al* [7]).

trace on the pressure signal from maximum pressure onwards, is also reproduced as a ringing signal on the optical intensity trace, indicating that this pressure variation has a direct effect on the optical radiation measured. High-pass filters can be used to extract this knock information, and in production vehicles this could avoid the need for special block-mounted knock sensors (accelerometers). In fact the signal obtained continues to be easy to separate out even at maximum engine speed, which is not the case with either accelerometers or pressure sensors due to the high vibration levels in the engine.

Ignition timing and cylinder pressure can be directly derived from the optical signal and can be used for engine feedback control, as can the information from a pressure sensor. In the case of the optical sensor, however, flame temperature information can also be obtained from flame colour by the use of optical filters, and this technique may be useful in future for engine measurement and control when operating conditions are critical and a sophisticated analysis of the gas content is required.

It is interesting to note that none of these short-period measurements require knowledge of the degree of contamination of the front window face, provided that it is not sufficient to totally obscure the optical signal. In practice it has been found that contamination increases and decreases depending on engine operating conditions but that the optical face remains sufficiently clear for adequate optical transmission under all normal engine conditions.

The possible use of a silicon rod down the centre of the spark plug has already been suggested and could also make it possible to view the ignition process as well as combustion. This may give valuable additional information which can be used either for control or analysis of the critical early period of flame growth which is so crucial to an engine's performance and emissions.

It has been proposed that such a sensor could be used to fine-tune the position of the spark plug electrode and gap, and to study the effect that electrode erosion and air–fuel turbulence has on performance and emissions.

Recently Witze and Hall [8] have experimented with a standard spark plug instrumented with eight equispaced peripheral optical fibres inserted as shown in figure 6.8. These serve as 'eyes' looking at the developing flame kernel after ignition. The fibres are 200 mm in diameter and each fibre transmits flame emissions which originate within a 23° acceptance cone. The spatial resolution of the probe is complex and varies from about 10 to 50% of the detection circle radius, depending on where the flame kernel enters the acceptance cone of the optical fibre.

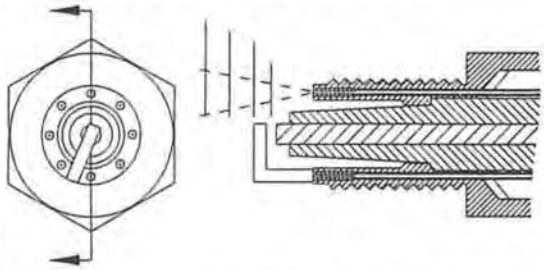

Figure 6.8 Spark plug instrumented with eight peripheral optical fibres (from Witze and Hall [8]).

In experimental use to determine the shape and path of the flame kernel after ignition, the light collected by the fibres is transmitted to photodetectors in a specially built signal processor which produces a digital signal if the light level is greater than a predetermined level. From these signals it is believed that definitive trends of the shape and path of the flame kernel can be obtained by statistical analysis.

Among other conclusions, Witze and Hall consider that the probe may be useful as a gas velocity diagnostic device in unmodified production engines. However, it seems unlikely that there is any useful role for the device beyond its use for the research and development of in-cylinder combustion technology.

6.3 IONISATION SENSORS

Another method of measuring combustion parameters makes use of an ionisation sensor in the combustion chamber. Ionisation sensors have been

used for many years to detect the presence or absence of the flame in automatic central heating boilers. They rely on the fact that large numbers of ions are produced in the flame when hydrocarbon fuels are burning, and if a simple insulated electrode connected to a low-voltage supply (usually between 10 and 300 V) is placed in the combustion chamber then a low resistance to the surrounding metal will be seen in the presence of the flame, and a high resistance when the flame is absent.

Bray and Collings [9] quote the experimental use of three different types of ionisation measurement in internal combustion engines. These are:

(i) in-cylinder measurement of flame arrival and pre-ignition;
(ii) in-cylinder measurement of post-flame ionisation levels;
(iii) exhaust manifold measurement of ionisation levels.

6.3.1 In-cylinder Measurement of Flame Arrival and Pre-ignition

In the case of this first type of measurement, the object is to detect the arrival of the flame front of the combusting air–fuel mixture on the far side of the combustion chamber from the point of ignition. This is done by the insertion into the combustion chamber of a metallic electrode which is insulated from the surrounding metal as shown in figure 6.9, and supplied with a low voltage through a resistor across which any change in ionisation gap impedance can be measured as a voltage change. The arrival of the flame front at this electrode produces a rapid increase in ionisation which in turn attracts electrons to the normally biased ionisation electrode and ions to the surrounding metal. This produces a large step current in the external circuit.

In 1984 May [10] proposed the use of an ionisation sensor of this type to measure the time of flame arrival as the basis for feedback engine control. He was looking for two items of information which could be used in the feedback control of ignition timing and fuelling in engines. Firstly, the time it takes for the flame to arrive at the far side of the chamber after the firing of the spark plug, a measurement which provides information on the suitability of the ignition timing for the engine operating conditions and which can therefore be used as a feedback signal to correct that timing. Secondly, the scatter of arrival times between successive firings, a measurement which provides information on the air–fuel ratio because a weak mixture produces greater variation in the arrival times of the flame front than a richer mixture. When the arrival times are integrated over a sufficient number of engine cycles they can be used as a feedback signal to control that parameter.

Experimental engines have been built using these control inputs and have been shown to be capable of optimising ignition timing and fuelling

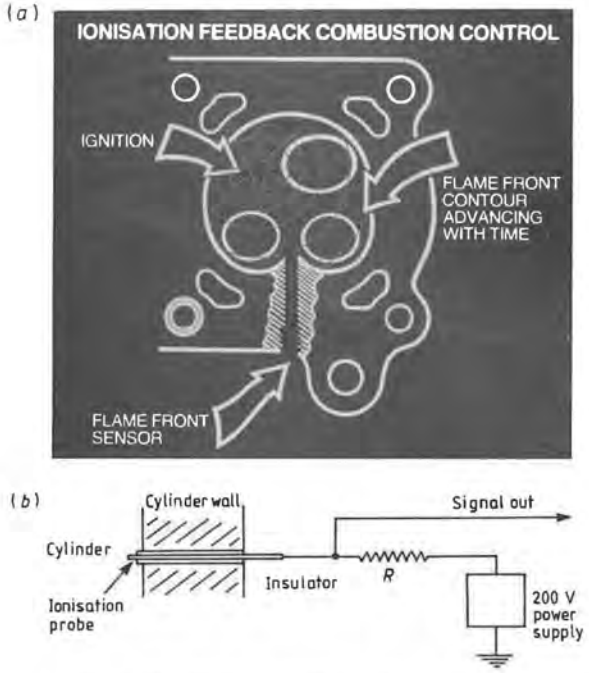

Figure 6.9 In-cylinder ionisation probe and its current.

under steady and slowly varying conditions. However, the long integration time required to obtain air–fuel ratio feedback information makes the use of the system impractical under real-life dynamic operating conditions.

6.3.2 In-cylinder Measurement of Post-flame Ionisation Levels

The use of the spark plug itself as an ionisation sensor after the ignition spark has been fired was first suggested by Blauhaut *et al* [11] in 1983, the circuit arrangement being as shown in figure 6.10.

In the post-ignition spark period the ion level in the spark plug gap decays slowly, its concentration being directly affected by the interrelated levels of cylinder pressure and temperature. In particular the rapid pressure rise in the cylinder caused by knock can be easily detected as it causes a rapid concentration of the ionised gas and a consequent reduction in spark plug gap resistance, and this information can be used to retard ignition timing to remove knock. However, slower changes in pressure are extremely difficult to separate from the variations caused by the many other disturbing effects.

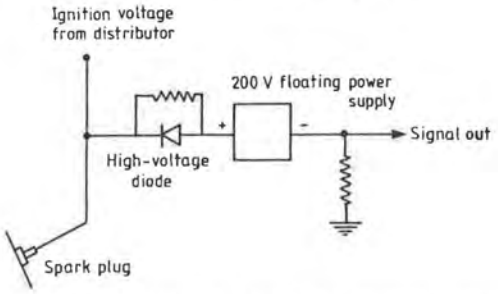

Figure 6.10 Circuit for spark plug ionisation sensor.

Not unexpectedly, electrical interference from other parts of the ignition system can cause severe interference and the method is not practical with distributorless ignition systems as sparks of both polarities are used.

If the voltage across the spark plug after the main spark event is maintained at a high enough level (a few hundred volts) and the current suitability limited, it is possible to maintain a controlled current arc between the plug electrodes in the remaining ionised gas. Under these conditions any gas movement across the plug electrodes will blow the arc into a curve instead of it taking the shortest path between the electrodes. Because of the consequent increase in arc length of the curved path and the accompanying increase in its resistance, it is possible to make measurements of gas velocity across the spark plug electrodes by measuring arc resistance, a potentially useful measurement in engine research.

6.3.3 Exhaust Manifold Measurement of Ionisation Levels

Ionisation sensors can be used in the exhaust manifold to detect the presence of a burning air–fuel mixture in the manifold, this being an indicator of late burning in the cylinder under excessively lean engine operation. The sensor can also warn of a misfire if no ionisation is detected immediately after what should have been a firing stroke. A circuit arrangement similar to that for the in-cylinder measurement described in section 6.3.1 would be suitable for this application of the ionisation sensor.

6.4 KNOCK SENSORS

Accelerometers will be covered comprehensively in chapter 9, but there is one particular type of accelerometer which is relevant to combustion sensing, and this is the knock sensor.

When an ignition system with electronic advance control is optimised for best performance and economy, it can, under some conditions, be set sufficiently far advanced to cause a condition known as 'knocking'. Under these conditions premature high-rate combustion takes place which, because of the rapid pressure increase, can quickly cause physical damage to vulnerable structures within the combustion chamber, such as the piston crown. In an engine with conventional ignition timing the advance angle is retarded sufficiently to avoid this condition; however, loss of efficiency results, so that it is highly desirable to operate the electronically controlled advance ignition as close to the knock limit as possible, but with the ability to retard the ignition within one or two engine cycles to a safe level. Operation in this mode requires a method of rapidly sensing the knock condition, and at present this is conventionally achieved by detecting the presence of knock by the use of an accelerometer, known, not surprisingly, as a knock sensor (see figure 6.11). This device is usually mechanically tuned to be sensitive to the characteristic knock ringing frequency, which in normal-size engines is in the region of 8 kHz. The knock sensor is placed on the engine block in a position chosen in an extensive vibration survey as the one at which approximately equal knock signals are received from all the engine cylinders.

Developments in in-cylinder measurements have made it possible to detect knock by any of the methods described earlier in this chapter, since knock is initiated by a rapid rise in pressure followed by acoustic ringing. Figure 6.4 shows this ringing superimposed on a cylinder pressure trace and figure 6.7 shows ringing superimposed on an optical trace. However, in the case of ionisation of the combustion flame front, the sensor on the far side of the combustion chamber from the spark plug does not detect the phenomenon, since it is lost within the variations which normally exist in the flame front arrival time. However, as described in section 6.3.2, it is

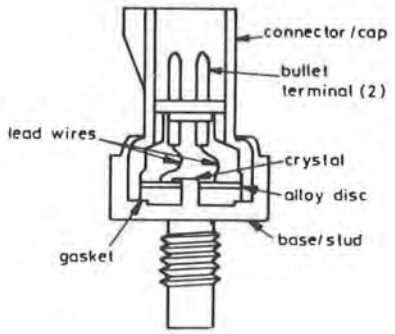

Figure 6.11 Knock sensor.

possible to use ionisation sensing methods to detect knock if measurements are made of the ionisation current across the spark plug after normal firing.

6.5 SUMMARY

Of all the methods developed for the measurement of combustion conditions in the cylinder it seems likely that the oldest — pressure measurement — has the most promise for production use in engine control feedback systems, provided that low-cost reliable devices can be made available.

Whether the benefits to be gained from the use of combustion sensors in sophisticated forms of feedback control can be justified for production vehicles depends on the improvements in economy and emissions that can be obtained compared with those already obtainable by the use of exhaust gas oxygen sensors in catalyst equipped engines.

There is no doubt, however, that many of the devices described in this chapter will continue to have wide application in engine research and development.

REFERENCES

[1] Kistler W P 1950 *Swiss Patent* no 267431

[2] Anastasia C M and Pestana G W 1987 A cylinder pressure sensor for closed loop engine control *Society of Automotive Engineers Paper* 870288.

[3] Mock R and Meixner H 1991 A miniaturized high-temperature pressure sensor for the combustion chamber of a spark-ignition engine *Sensors Actuators* A **25–27** 103–6

[4] Sasayama T, Suzuki S, Amano M, Kurihara N and Sakamoto S 1988 An advanced engine control system using combustion pressure sensors *Proc. 5th Int. Automotive Electronics Conf.* (*IMechE Conf. Publication 1985*) (London: IMechE) p 12.

[5] Kondo M, Niimi M A and Nakimira T 1975 Indiscope — a new combustion pressure indicator with washer transducers *Society of Automotive Engineers Paper* 750883

[6] Randall K W and Powell J D 1979 A cylinder pressure sensor for spark advance control and knock detection *Society of Automotive Engineers Paper* 790139

[7] Pinnock R A, Extance P and Cockshott C P 1987 Combustion sensing using optical fibres *Proc. 6th Int. Conf. on Automotive Electronics (IEE Conf. Publication 280)* (London: IEE) pp 280–3

[8] Witze P O and Hall M J 1990 Cycle-resolved measurements of the flame kernel growth and motion correlated with combustion duration *Society of Automotive Engineers Paper* 900023

[9] Bray K N C and Collings N 1991 Ionization sensors for internal combustion engine diagnostics *Endeavour* **16** (1) 10–12

[10] May M G 1984 Flame arrival sensing fast response double closed loop engine management *Society of Automotive Engineers Paper* 840441

[11] Blauhaut R B, Horton M J and Wilkinson A C N 1983 A knock detection system using spark plug ionization current *4th Int. Conf. on Automotive Electronics* (London: IEE) (late paper not included in the proceedings)

7

Torque Sensors

Torque seems to be the subject of some confusion, even amongst experienced engineers. Statements such as 'the torque is the amount of twist in a shaft' or 'torque is a measure of the wind-up' are often heard. It is sensible therefore to begin with a formal definition of torque, before moving on to consider methods by which it can be measured.

Torque can be defined as a measure of the tendency of a force to rotate the body on which it acts about an axis. Everyday experience tells us that the 'rotating effectiveness' of a force increases with its perpendicular distance from the pivot. For example, when opening a door it is normal to push or pull as far as possible from the hinges, and we attempt to keep the direction of the pull or push perpendicular to the door.

The magnitude of the torque acting in a plane perpendicular to an axis is obtained by multiplying the force (or the component of the force in a plane perpendicular to the axis of rotation), by the perpendicular distance from the axis to the line of action of the force, as shown in figure 7.1. The SI unit of torque is newton metre (N m).

The discussion above applies to both static problems (where, for example, a structural member is subjected to a moment), and to the more important case (for engineers) of rotating shafts in torsion. For the automotive engineer, torque in a rotating shaft is important since it is the

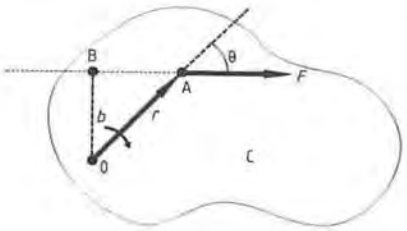

Figure 7.1 Definition of torque. Force F acting on body C tends to rotate C about 0. Torque $\tau = Fb = Fr \sin \theta$. In vector notation torque $\tau = \boldsymbol{r} \times \boldsymbol{F}$.

limiting factor in determining how much power the shaft can transmit. A shaft rotating with angular velocity ω and carrying power P will undergo a torque T, where

$$T = \frac{P}{\omega}. \tag{7.1}$$

A shaft of length l, polar moment of inertia J and modulus of rigidity G, subjected to a torque T, will experience an angle of twist θ given by

$$\theta = \frac{Tl}{JG}. \tag{7.2}$$

The maximum shear stress τ occurs at the surface of the shaft and is

$$\tau = \frac{Tr}{J} \tag{7.3}$$

where r is the shaft radius as shown in figure 7.2. For a solid circular shaft the polar moment of inertia $J = \pi r^4/2$, so by substitution we have

$$\theta = \frac{2Tl}{\pi r^4 G} \quad \text{and} \quad \tau = \frac{2T}{\pi r^3}. \tag{7.4}$$

Equations for the strains produced by torsion in shafts of other than circular cross section are readily derived [1].

The most important application of torque measurement in automotive engineering is in the assessment of engine power. Equation (7.1) shows how power can be readily calculated from measurements of rotational speed and torque. Engine speed measurement is relatively easy — often a white paint mark on the flywheel and a simple optical reflectance sensor will suffice. Torque measurement is more difficult to arrange, and for this reason it is still in the main the preserve of research and development

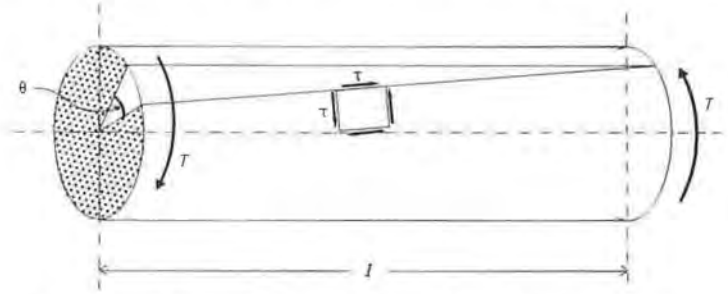

Figure 7.2 Circular shaft (radius r) under torsion. τ = shear stress.

laboratories. However, there are signs that this situation may be about to change, and torque sensors may soon become commonplace on production vehicles.

The designers of powertrain management systems would find a direct measurement of torque very useful. Strain-gauged torque sensors (see section 7.1) are available for use in development work, but are far too expensive (and unreliable in the long term) for use in production vehicles.

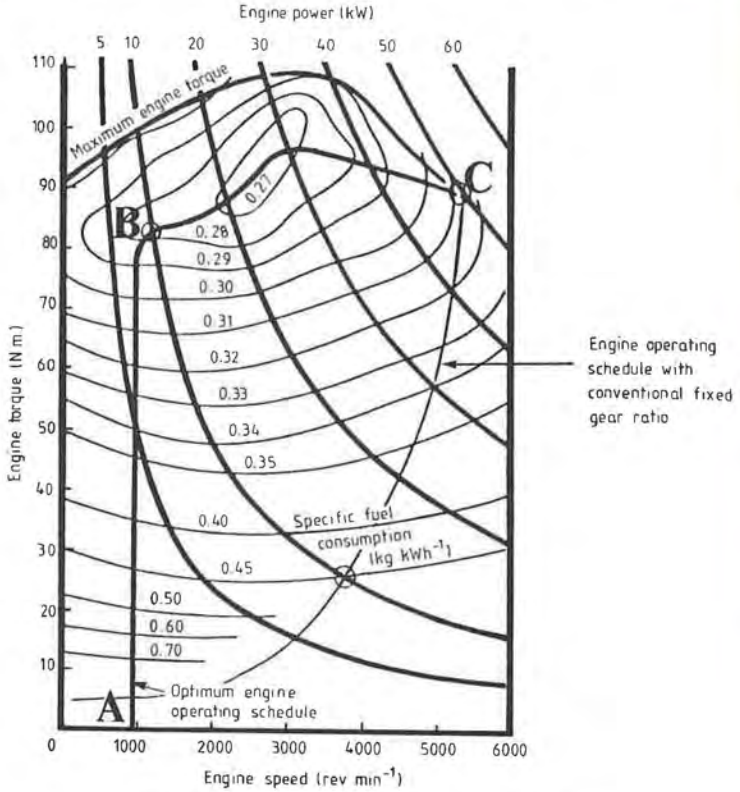

Figure 7.3 Typical engine torque–speed curves.

The utility of a reliable torque measurement device is apparent from consideration of the torque–speed characteristics of an engine and gearbox in combination. The gearbox is fitted between the engine and the road wheels of a vehicle, and serves purely as an impedance matching device. With a conventional manual gearbox the driver acts as the feedback loop, sensing speed and load and adjusting the transmission ratio

to what appears to be the optimum setting. The main sensory input used to achieve optimisation is engine speed, perceived in the form of noise. Unfortunately the pitch and noise level of an engine are poorly correlated with its power output or efficiency, so although gear changing controlled by acoustic stimuli may optimise the subjective acceleration and driveability, it gives poor economy and performance [2]. Figure 7.3 shows the torque–speed curves for a typical engine, and is taken from [2]. From these curves it is apparent that the optimum economy is obtained by keeping the engine at as low a speed as possible during acceleration (line A–B on figure 7.3), while using the gears to increase vehicle speed. The engine speed is only raised to increase the vehicle speed when the final gear ratio is reached at point B. Subsequently the powertrain is controlled for optimum fuel economy, and operates along line B–C. Good acceleration performance of course requires a somewhat different strategy.

7.1 MECHANICAL METHODS OF TORQUE MEASUREMENT

One of the earliest (and still very useful) methods of measuring the torque produced by an engine uses a device known as an *absorption dynamometer*, in which all the power produced by the engine is absorbed by friction in a brake. This is the origin of the phrase *brake horsepower*, although shaft power is a less misleading term. The engine is fitted with a rope or belt brake wrapped around the flywheel, which is often water-cooled. The rope passes once around the flywheel and is attached to a mass M at the bottom as shown in figure 7.4. The other end of the rope is connected to a spring balance which measures the tension in the rope, T. The force in the lower end of the rope arises from the weight, and is Mg. If the spring balance reading is T, the difference in tension between the ends of the rope is $Mg - T$. If the radius of the flywheel is R, the torque will be

$$\text{torque} = (Mg - T)R$$

and the power

$$\text{power} = 2\pi N(Mg - T)R \text{ (watts (W))}$$

(where N is the number of flywheel revolutions per second).

The danger inherent in this arrangement is that the brake may jam, throwing the weight over the top of the flywheel. To avoid this alarming possibility a strong safety rope or chain is always used as shown in figure 7.4, which prevents the weight being lifted more than a few centimetres.

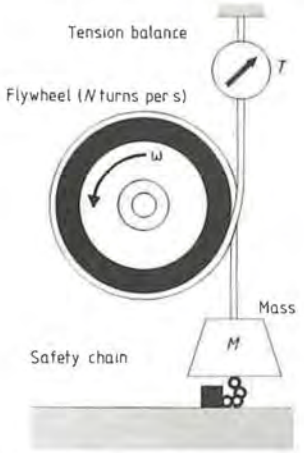

Figure 7.4 Brake dynamometer.

A more sophisticated brake dynamometer frequently used for torque measurement in automotive powertrain research is the hydraulic type, originally invented by W Froude. In a Froude dynamometer the energy from a rotating shaft is transferred to water, which is then brought to rest. The torque required to restrain the device is measured by (usually) a spring balance. The advantage of a Froude dynamometer is that unlike a rope brake device there is no possibility of it 'snatching'. However, Froude dynamometers are more expensive than rope brakes.

It is interesting to note in passing that the automotive fluid flywheel was almost certainly developed from the Froude dynamometer.

An entirely mechanical method of torque measurement is based on a measurement of the force required to restrain a gearbox. As far as the authors are aware it has not been applied in automotive engineering, although there seems to be no reason why it should not be successful. Any gearbox which changes the rotational speed of a shaft will change the torque in inverse proportion (assuming friction can be neglected). The *ratio* of the input torque T_{in} to output torque T_{out} is equal to the reciprocal of the speed ratio, and the *difference* between the input and output torques is the torque needed to restrain the gearbox. Thus

$$\frac{T_{in}}{T_{out}} = \frac{\omega_{out}}{\omega_{in}}$$

and

$$T_{in} - T_{out} = T_{restraining}.$$

By measuring the input and output speeds, and the restraining torque, T_{in} and T_{out} can be calculated. This approach has been found to be very useful

in university laboratories, where costs are of overriding importance, since it allows a torque measuring system to be improvised cheaply using a scrap back axle from a rear-wheel-drive vehicle. The engine is connected to one wheel shaft and the load to the other. The propeller shaft coupling is locked to the differential housing. The torque required to restrain the housing is measured with a spring balance or electronic force transducer, and is twice the torque being transmitted through the system since the input and output shafts revolve in opposite directions.

7.2 STRAIN GAUGE TORQUE TRANSDUCERS

Strain gauge torque transducers are created by applying strain gauges to a shaft to measure the shear strain caused by torsion, as shown in figure 7.2 and discussed in the introduction to this chapter. They are very widely used in research laboratories, and are probably the most common form of torque sensor. Their major disadvantage is that they require additional equipment to transmit power to the rotating shaft, and to retrieve data from it. This can take the form of a set of slip rings, rotary transformers or battery-powered radio telemetry equipment. Regardless of which of these is chosen, the need for some form of power and/or data transmission system and the consequential costs incurred probably rules out strain gauge based torque sensors for use on volume-produced vehicles. In addition slip rings (and to some extent rotary transformers) can be unreliable when operated in a dirty environment and may be prone to radio-frequency interference (RFI).

The shear stresses illustrated in figure 7.2 cause strains to appear at 45° to the longitudinal axis of the shaft. The conventional arrangement of strain gauges for torque measurement is shown in figure 7.5. The gauges must be placed precisely at 45° to the shaft axis, otherwise the arrangement is sensitive to bending and axial stresses in addition to those

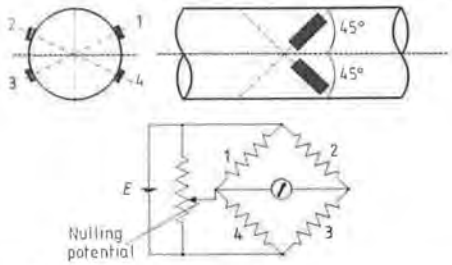

Figure 7.5 Strain gauges for shaft torque measurement.

caused by torsion. Accurate gauge placement is facilitated by the availability of special 'rosettes' in which two gauges are precisely positioned on a common backing. The use of four active strain gauges in a bridge arrangement gives complete thermal compensation [3]. Figure 7.5 shows an arrangement of strain gauges on a solid circular shaft. The same gauge positioning can be used on a hollow circular shaft, such as the propeller shaft on a rear-wheel-drive vehicle. When torque is applied to a thin-walled cylinder, such as a rear-wheel-driveshaft, the shear stress is assumed to be constant throughout the wall [4]. In such cases it is often convenient to place the strain gauges on the inner surface of the driveshaft where they are afforded a degree of mechanical protection.

Shafts of other than circular cross section are sometimes used for torque measurement, as shown in figure 7.6. For measuring low levels of torque the cruciform or hollow cruciform configuration is sometimes used. A solid square shaft is suitable for larger torque values, and has a number of advantages over the circular shaft of figure 7.5. The strain gauges are more easily aligned and attached to a flat surface, and since the corners of a square section in torsion are stress free [5] they provide a good location for the solder joints between leads and strain gauges. These joints are often a source of unreliability due to fatigue failure if they are located in a high-stress region. Finally, a square shaft is much stiffer in bending than a circular one of equivalent torsional stiffness, so the effects of bending (which will appear if the gauges are misaligned) are reduced.

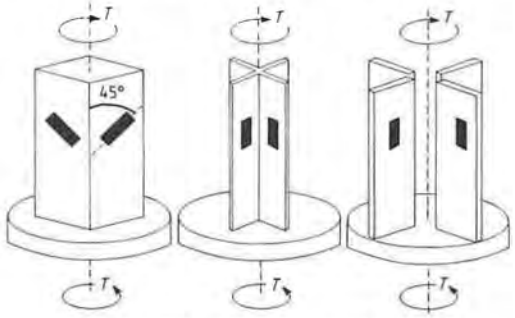

Figure 7.6 Torque sensor designs.

7.3 TORSION BARS

Torque in a shaft leads to elastic deflection. The resulting strain can be measured at a point as described in the preceding section, or alternatively

the gross relative motion between the ends of the shaft may be used to indicate the torque. Just as in the case of strain gauge systems, a major difficulty is the necessity of being able to measure the deflection while the shaft is rotating. However, there are advantages in using shaft deflection. Firstly, the need for precise location and orientation of the strain sensors is avoided. Secondly, since the effect of an applied torque is integrated along the length of the shaft, the influence of any local variation in material properties or shaft geometry is reduced. Thirdly, the (relatively) larger displacements available when movements of the two ends of a shaft are compared make it possible to design a variety of non-contact torque measurement systems which avoid the need for slip rings.

Figure 7.7 shows a typical torsion-bar torquemeter using an optical method for deflection measurement. The relative angular displacement between the ends of the torque-transmitting member is read from the position of the pointer on disc 2 relative to the calibrated scale fixed to disc 1. The 'persistence' of human vision and the stroboscopic effect of intermittent viewing make it possible to operate this system from about 600 rev min^{-1} (10 Hz) upwards.

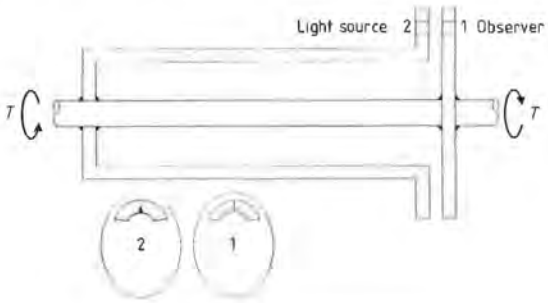

Figure 7.7 Torsion-bar torquemeter.

A torsion-bar system using capacitive torque sensing has been demonstrated for automotive use [6]. An automotive driveshaft was fitted with a concentric sleeve of dielectric material as shown in figure 7.8. The sleeve is fixed to the shaft at one end, and rests on a rubbing bearing at the other end. When torque is applied to the shaft it causes relative motion between the surface of the shaft and the free end of the concentric tube. This motion is used to vary the capacitance between two opposing patterns of conducting strips, one of which was applied to the shaft and one to the tube.

The capacitive torque sensor was connected to an inductor coil wound around the shaft. The resulting passive circuit has a resonance frequency which depends on the applied torque.

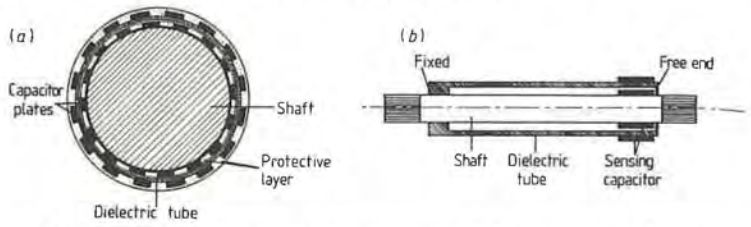

Figure 7.8 Construction of torque sensor: (*a*) cross section through shaft and sensor; (*b*) longitudinal section.

The passive resonant circuit rotates with the driveshaft, and is excited from an adjacent stationary location by inductive coupling using a second inductor coil driven by an oscillator as shown in figure 7.9. The problem of torque measurement then becomes one of measuring the frequency at which resonance occurs. When the oscillator frequency is the same as that at which resonance occurs in the passive circuit an increased current is drawn. If the frequency at which this occurs is measured it can be used to indicate the torque. The advantage of this arrangement for automotive applications is that no physical connection between the rotating shaft and the vehicle body is required.

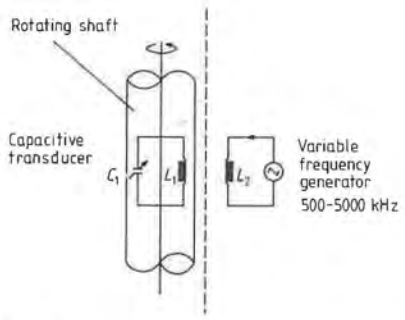

Figure 7.9 Rotating resonant circuit excited by inductive coupling.

Figure 7.10 shows a disassembled prototype device. The sensors were manufactured by thick-film printing and were found to be reasonably robust. In the prototype signal conditioning system the torque sensor was inductively coupled into a capacitive bridge as shown in figure 7.11. With this arrangement the output was found to vary as shown in figure 7.12. It can be seen that it is reasonably linear, and is probably adequate for automotive applications such as engine management or automatic gearbox control.

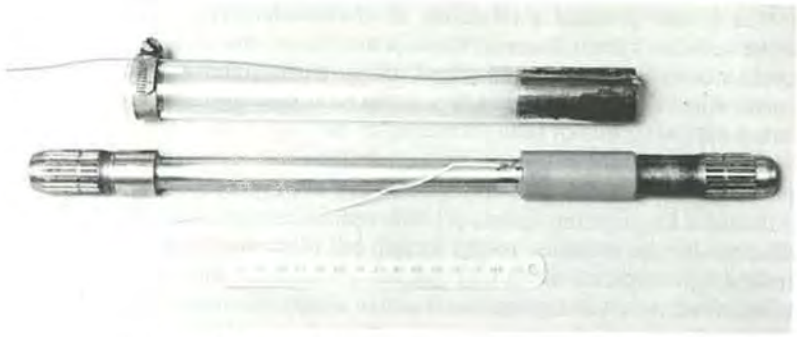

Figure 7.10 Disassembled capacitive non-contact torque sensor.

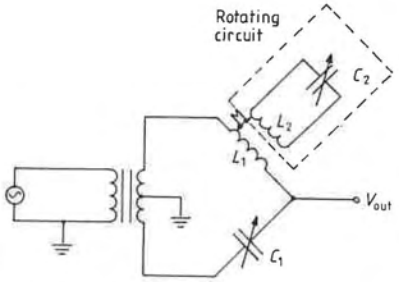

Figure 7.11 Bridge circuit for telemetry.

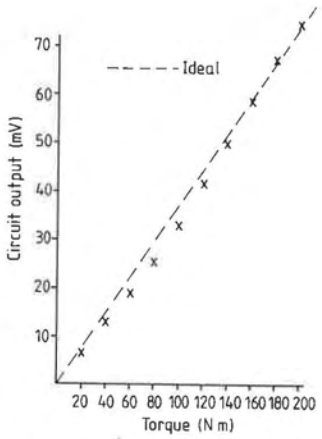

Figure 7.12 Output from bridge telemetry.

The prototype used a relatively short driveshaft from a front-wheel-drive vehicle (a Ford Escort). There is no reason why the same technique could not be applied to rear-wheel-drive vehicles, and in fact it may be easier since the driveshafts used tend to be longer and will have a larger 'wind-up'.

An optical torsion-bar sensor intended for use as part of an electric power-assisted steering (EPAS) system has been proposed by the Lucas Advanced Engineering Centre [7]. The optical sensor uses a pair of slotted discs positioned at the ends of a torsion bar as shown in figure 7.13. Light from a light-emitting diode (LED) passes through the slots in the discs and is received by a photodetector. Torque variations cause the amount of overlap between the discs to vary, and hence the output from the photodetector. However, the Lucas system exhibits a number of refinements which are intended to make it more suitable for automotive use. The most important of these is the use of a ratiometric technique to cancel out the effect of any variation in the source illumination intensity. The slotted discs are illuminated by a common LED as shown in figure 7.13. The amount of light emitted by the diode will vary if the supply voltage changes. Even if a well regulated supply is available, the light output from an LED reduces by up to 40% as the device ages. The ratiometric effect is achieved by arranging for each slotted disc to carry two tracks of slots, positioned so that as torque is applied in one direction the light intensity transmitted through the outer track (A) increases, while that passing through the inner track (B) decreases. The light passing through each track is measured by a pair of photodiodes as shown in figure 7.13. The torque is calculated by measuring the outputs from photodiodes A and B and then evaluating the expression

$$\text{torque} = \frac{A - B}{A + B}.$$

The magnitude of the result gives the torque, and the sign of the direction

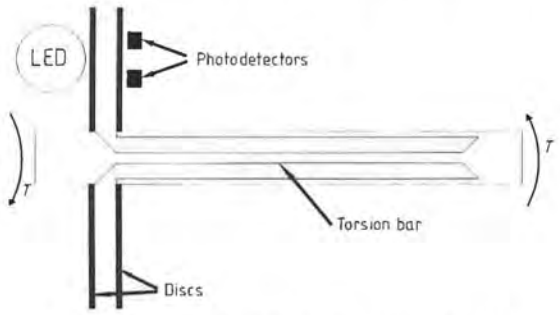

Figure 7.13 Lucas EPAS ratiometric torque sensor.

(i.e. clockwise or anticlockwise) in which it is applied. Provided both channels are affected equally, this technique ensures that the torque measurement is independent of the source intensity. Furthermore, for a given source intensity the quantity $A + B$ should be a constant which is independent of torque, and this value can be used to check that the sensor is operating correctly. If $A + B$ moves outside preset limits an appropriate warning may be given. A self-test facility of this kind is obviously essential in a safety-critical system such as vehicle steering.

The main problem with the Lucas system appears to be that the geometry of the photodetectors, and their location with respect to the slots on the discs, is critical if ripple in the sensor output is to be prevented as the discs rotate. Variations in the output can only be avoided if the sensitive area of the light detectors corresponds exactly to an even multiple of the slot area. Reference [7] proposes the use of masks to give the correct detection area and to collimate the light source. The ripple amplitude after these improvements is reported to be better than 10% of the full-scale measurement range. Although this level of accuracy would not be acceptable for a laboratory torque sensor, it is probably adequate for power-steering applications.

Work on measuring the twist or 'wind-up' along the crankshaft of an engine using slotted discs at each end has also been reported [8]. However, the very high levels of torque variation which result from multicylinder operation are alleged to make it difficult to obtain accurate results.

7.4 NON-CONTACT MAGNETIC METHODS

A number of torque sensors utilising the magnetostrictive effect have been reported. A good example of this approach is a device described by Spectrol Electronics [8] and shown in figure 7.14. Magnetostriction is an effect which occurs in ferromagnetic materials such as steel, where the magnetic permeability is affected by stress. Equation (7.3) shows that the stress in a shaft is proportional to the applied torque, and it follows that torque must change the permeability of the shaft if it is made of a magnetic material such as steel. The effect is small but can be measured by an arrangement such as that shown in figure 7.14. The torque sensor consists of five coils arranged as shown, wound onto a common five-armed core. The centre coil can be thought of as the primary winding of a transformer, and the four circumferentially positioned coils act as secondaries. Magnetic coupling between the primary and the secondaries is provided by the steel shaft, which is positioned close to the sensor as

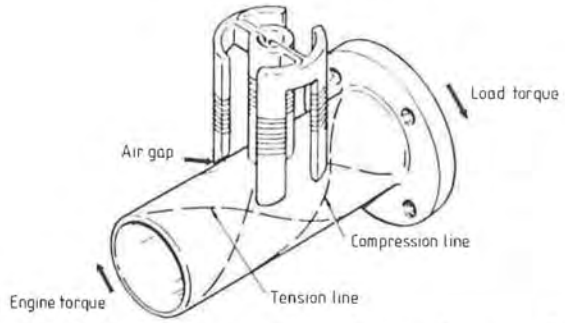

Figure 7.14 Magnetostrictive torque sensor (courtesy Spectrol Electronics).

shown. The primary coil is excited by an AC current, and produces an oscillating magnetic field within the shaft. The four secondary coils are connected together in a Wheatstone bridge arrangement, and are positioned so that they lie over the lines of principal stress, which follow a helical path at 45° for a cylinder in torsion. When the shaft is not under torsion equal currents are induced in the four secondaries and the bridge out-of-balance voltage is zero. When torque is applied to the shaft the permeability in the tension and compression directions will change by equal but opposite amounts, and the amplitude of the resulting bridge output voltage is proportional to the applied torque.

There are four main problems with this type of torque sensor. They are:

(i) inhomogeneity of the shaft material;
(ii) sensitivity to changes in the sensor/shaft gap;
(iii) thermal effects; and
(iv) variations in the sensor output due to changes in the shaft rotation speed.

The first of these effects is the most serious. The permeability of the material from which the shaft is made can vary by up to 50% around the circumference of the shaft. For a constant torque the output signal from the sensor can 'ripple' as a result at a frequency equivalent to the rotation rate. This characteristic makes it very difficult to measure instantaneous torque levels around a rotating shaft. However, the use of smoothing circuits allows the device to be used for measuring the average torque in the shaft by integrating over several revolutions.

Variable-permeability torque sensors of this type are probably not suitable for use as part of a high-speed real-time engine management

system, where instantaneous torque measurement is required, but they may be useful for less demanding tasks such as automatic transmission control.

7.5 SUMMARY

In conclusion, it appears that although automotive powertrain engineers would find a low-cost, reliable torque sensor very useful for engine and transmission control, one has yet to be produced which meets enough of the requirements to gain widespread acceptance. While strain gauge systems using sliding contacts or telemetry are adequate for development work, they are prohibitively expensive for use on production vehicles. The torsion-bar systems described in section 7.3 have a suitable performance, can be low-cost, and avoid the need for telemetry. However, the optical versions are liable to dirt problems, which may prohibit their use for powertrain applications. It is probably feasible, however, to use optical systems in 'clean' applications such as steering torque measurement [7]. The capacitive torque sensor described in [6] does not need to be kept as clean as an optical system, but is at present rather labour-intensive to construct and requires careful setting up. It is hoped that further development will reduce these disadvantages.

The magnetostrictive system described in section 7.4 also avoids the need for telemetry or an electrical connection to the rotating shaft, but suffers from the drawbacks discussed earlier. In addition an experimental version constructed for evaluation purposes was found to be very sensitive to temperature.

To sum up, it appears that torque sensing is a real need in automotive engineering, but to date no practical low-cost systems have been developed which are suitable for widespread use.

REFERENCES

[1 Roark R J and Young W C 1984 *Formulas for Stress and Strain* (New York: McGraw Hill) ISBN 0 07 085983 3

[2] Westbrook M H 1988 Automotive transducers: an overview *Proc. IEE* D **135** 339–47

[3] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 44551 1
[4] Benham P P and Crawford R J 1987 *Mechanics of Engineering Materials* (London: Longman) ISBN 0 582 28640 9
[5] Timoshenko S P and Goodier J N 1982 *Theory of Elasticity* 3rd edn (New York: McGraw-Hill) ISBN 0 07 085805 5
[6] Turner J D 1989 The development of a thick-film non-contact shaft torque sensor for automotive applications *J. Phys. E: Sci. Instrum.* **22** 82–8
[7] Hazelden R J 1992 Application of an optical torque sensor to a vehicle power steering system *Proc. IEE Coll.* C12: 1992/107 May
[8] Westbrook M H 1985 Sensors for automotive applications *J. Phys. E: Sci. Instrum.* **18** 751–8
[9] Wells R F 1988 Non-contacting sensors for automotive applications *Society of Automotive Engineers Paper* 880407

8

Position, Displacement and Velocity Sensors

Displacement sensors are probably the most widely used type in automotive engineering. Resistive position detectors seem to have been the earliest form of electronic sensor to be incorporated into a production vehicle. By the late 1920s remotely operated petrol gauges were becoming standard, in which a float on a lever arm in the petrol tank moved the wiper of a wire-wound potentiometer. This in turn controlled a remote (dashboard-mounted) voltmeter, which was provided with a pair of windings known as the *deflection* and *control* windings. The control coil replaced the hairspring found in an ordinary voltmeter. The two windings were interconnected in such a way that changes in the battery voltage did not affect the reading. For example, a decrease in battery voltage decreased the deflecting force, but also decreased the controlling force. The meter reading which resulted was a function of the ratio of the two forces, and was independent of battery voltage.

Modern fuel gauges are not usually voltmeters, since these have too fast a response and give a reading which fluctuates due to the fuel sloshing on hills and when cornering. Instead a meter is used which contains a bimetallic component and a heating coil. The deflection of the pointer depends on the current in the heating coil, which is in turn controlled by the resistance of the float sensor in the fuel tank. A voltage regulator is usually fitted to remove the effect of supply voltage variations. The thermal inertia of the system is made sufficiently large to smooth out most of the effects of fuel slosh.

Other automotive potentiometer applications followed later. In the late 1950s Bendix developed and patented their 'electro-injector' fuel injection system, in which the throttle position was sensed by a potentiometer. This was subsequently refined by Bosch to form the basis of the well known 'D-Jetronic' system, used extensively by Volkswagen and others in the 1960s.

Potentiometer sensors are now frequently used for sensing throttle and brake pedal position, steering wheel motion, suspension displacement, for automatic gearbox control, and are often provided in the form of controls

to allow the level of dashboard illumination or the frequency of intermittent windscreen wiper operation to be adjusted. Both rotary and linear versions are available, as discussed in section 8.1. In-car entertainment systems are probably the most frequent users of potentiometers for control. However, these systems lie outside the scope of this book.

Potentiometers are cheap and reasonably reliable for many applications, but suffer from the major disadvantage common to all devices which rely on a sliding contact, namely wear. While they may be adequate for, say, throttle position transduction, they tend to give rise to problems if used for applications such as shock absorber motion sensing. This is because a car body tends to remain close to one position relative to the wheels throughout a journey, but undergoes large numbers of small excursions around the 'mean' position. This phenomenon is known as *dither*, and unless special precautions are taken it can cause parts of the potentiometer track to become badly worn or even destroyed locally. For this reason many users are beginning to consider alternative, non-contact forms of displacement sensor, such as the inductive, capacitive or optical types discussed in detail in sections 8.2 to 8.4.

An eddy current probe consists essentially of a coil which is placed close to a conducting target. The coil is excited by a high-frequency (typically 1 MHz) signal, which induces eddy currents in the conducting target. If the distance between the target and probe changes the eddy currents vary, which alters the impedance of the coil. A bridge circuit is often used together with a second 'balance' coil, which provides temperature compensation. The bridge imbalance voltage is demodulated, low-pass filtered and linearised to produce a DC output proportional to target distance. Only a few applications of this technique have been reported in production road vehicles, but it is quite common in laboratory and development work. Section 8.2 contains details of this type of device.

Section 8.5 contains a description of ultrasonic displacement transducers. These are becoming quite common, and in automotive engineering have been used for applications such as ride height sensing in adaptive-suspension control, as well as for short-range collision warning in the form of parking and reversing aids. Ultrasound is also used in some aircraft for fuel-tank level sensing, and this practice may spread to include automotive applications.

The Hall effect occurs in semiconductors, and leads to an output voltage which depends on the strength of a transverse magnetic field [1]. Proximity or distance measuring systems based on this principle may use a permanent magnet fixed to the displaced object, or else require a ferrous target, the approach of which changes the reluctance of an internal magnetic circuit whose flux is measured by the Hall sensor. In automotive applications Hall sensors are often used to sense rotation rate, for example

in anti-lock braking systems (ABS), where a Hall probe is placed close to a toothed wheel. Section 8.6 gives a detailed discussion of the Hall effect and shows how it may be applied in a variety of sensor applications.

This chapter is brought to a conclusion by a discussion of brake-pad and clutch-plate wear sensors. These often consist of an embedded wire link, which is broken when the component wears down. They can therefore be considered as a crude form of displacement sensor.

8.1 POTENTIOMETERS

A potentiometer consists essentially of a resistive element which is provided with a movable contact. In the earliest forms of potentiometer the resistive element or 'track' was made from high-resistance wire such as Nichrome, wound onto an insulating former of suitable shape. The contact consists of a springy conducting arm, which is arranged so that it can be moved along the potentiometer track. The contact motion can be linear, rotary or a combination of the two such as helical movement.

Translational (also called linear†) potentiometers are available with strokes from about 5 to 1000 mm. Rotary versions range from about 10° to as much as 60 turns.

8.1.1 Linearity

If the resistance of a potentiometer is linear with respect to its travel, the output voltage e_o (see figure 8.1) is a linear function of displacement x_i when the terminals are open-circuit and no current is drawn. However, all circuit inputs draw some current, so any signal conditioning arrangement connected to the potentiometer will degrade its linearity to some extent. Figure 8.2 shows the usual arrangement, and from simple circuit analysis it

†The term 'linear'; is used somewhat confusingly with respect to potentiometers, and can have two meanings. It is often used to denote a device in which the contact motion is translational. However, it is also frequently used to describe a device in which the resistance is proportional to displacement (rather than being, say, logarithmic). Such a 'linear' potentiometer can of course have a contact which undergoes circular, helical or translational motion, and hence the confusion arises.

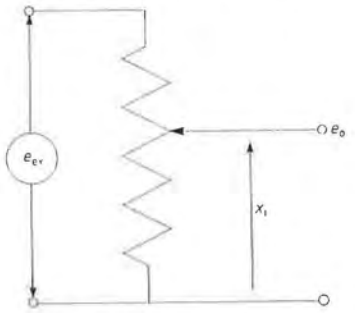

Figure 8.1 Potentiometric displacement sensor.

can be shown that

$$\frac{e_o}{e_{ex}} = \frac{1}{(x_t/x_i) + (R_p/R_m)\,(1 - x_i/x_t)}. \tag{8.1}$$

Under ideal conditions $R_p/R_m = 0$ for an open circuit, and equation (8.1) becomes

$$\frac{e_o}{e_{ex}} = \frac{x_i}{x_t}. \tag{8.2}$$

Thus when no current is drawn the input–output relationship is a straight line. In practice $R_m \neq \infty$, and as shown in figure 8.3 there is a nonlinear relationship between e_o and x_i. If $R_p = R_m$, the maximum deviation from linearity is about 12%. If $R_p = 10\%$ of R_m, the error drops to about 1.5%. For values of $R_p/R_m < 0.1$ the position of maximum error is in the region where $x_i/x_t \approx 0.67$, and the maximum error is approximately 15 R_p/R_m% of full scale.

To achieve good linearity, therefore, the input impedance R_m of any circuit connected to a potentiometer should be high compared with the

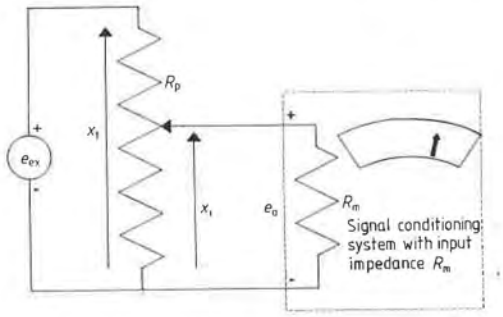

Figure 8.2 Potentiometer connected to input impedance R_m.

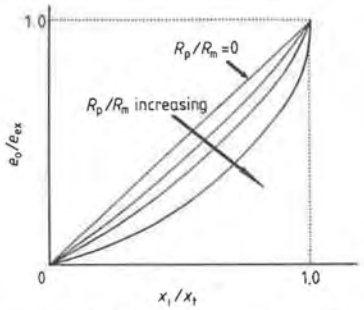

Figure 8.3 Potentiometer loading effects.

potentiometer impedance R_p, which should be kept as low as possible. Unfortunately this requirement conflicts with the almost invariable need for high sensitivity. Since the output e_o is directly proportional to e_{ex}, it appears to be possible to get any desired output simply by increasing e_{ex}. However, potentiometers have a fixed power rating, which is determined by their heat-dissipating capability. If the limiting heat dissipation is H W, the maximum allowable excitation voltage is

$$e_{ex}\,(\max) = \sqrt{HR_p} \tag{8.3}$$

Thus, a low value of R_p allows only a small e_{ex}, and therefore a reduced sensitivity. The choice of R_p must be a compromise between considerations of loading and sensitivity.

8.1.2 Resolution

The resolution of a potentiometer depends on its construction. In a wire-wound type the variation in resistance proceeds in small steps as the wiper moves from one turn of resistance wire to the next, as shown in figure 8.4.

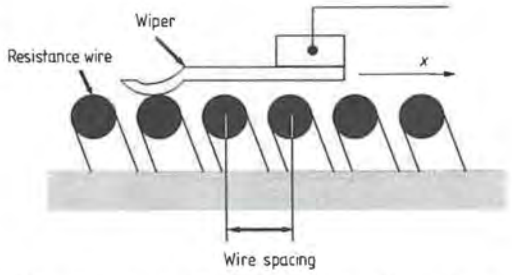

Figure 8.4 Wire-wound potentiometer resolution depends on wire spacing.

The finest wire spacing that can be achieved is around 25 turns/mm, and thus for wire-wound translational devices the best resolution that can be achieved is about ± 40 μm.

The resolution of *carbon film, cermet* (a mixture of ceramic and metallic materials) or *conductive plastic* (a mixture of plastic resin and a conducting powder) resistance elements is higher than that of wire-wound devices, but they are not so hard-wearing. This type of potentiometer is often described as having infinite resolution, since the resistance element presents a smooth surface to the wiper. The resolution cannot actually be measured and quantified, since the deviations of e_o from the ideal straight line are random (in contrast to the repeatable output 'steps' of a wire-wound device). The quantity usually cited in an attempt to specify the output smoothness is the ratio of the peak amplitude of the random variations to e_{ex}, with 0.1% being a typical value.

A further problem with non-wire-wound resistance elements is that they cannot tolerate high excitation currents. As we have seen in section 8.1.1, this leads to a reduction in sensitivity.

A compromise solution is offered by the so-called *hybrid potentiometer*, in which a layer of conductive plastic is applied on top of a wire-wound track. This approach combines the best features of both types of potentiometer, but at an increased cost.

8.1.3 Noise

Electrical noise in a potentiometer arises from output voltage fluctuations due to contact slider bounce, dirt and contact or track wear. In dynamic applications a potentiometer can cause significant mechanical loading due to the inertia and friction of its moving parts. This can affect the characteristics of the motion being measured.

In a wire-wound potentiometer the sliding contact may 'bounce' at certain speeds as it passes over the turns of wire, causing intermittent contacting. This effect can be particularly severe if the speed and wire spacing are such that an exciting force occurs at or close to the fundamental frequency of the spring-loaded contact arm. A solution to this problem which is sometimes adopted is to use a contact divided into two or more parts, each with a different resonance frequency, so that if one part is resonating the other still makes a good contact.

8.2.4 Examples of Automotive Potentiometric Displacement Sensors

Automotive sensors have to withstand severe environmental conditions, including temperatures from -40 to $+140$ °C. For applications such as throttle position sensing, where very high temperature operation is

unlikely, conductive polymer tracks are frequently used. The conductor material is usually screen-printed onto a substrate moulded from thermosetting plastic. The substrate material is selected to have a thermal coefficient of expansion (TCE) similar to that of the resistor track. If prolonged operation at the high end of the temperature range given above is likely, a hybrid (wire-wound/conductive plastic) design may be required.

The slider contacts are usually multifingered to ensure good performance under vibration, and are normally designed to withstand vibration levels up to 50*g*. The contacts are usually made from precious metals such as gold, silver or platinum, welded onto a base metal carrier.

Automotive potentiometers are sealed to a higher standard than similar devices intended for purely electronic use. An automotive potentiometer has to be able to withstand petrol, oil, hydraulic fluid, steam, boiling water and in some cases even battery acid without damage. Figure 8.5 shows a selection of automotive potentiometric sensors, and figure 8.6 a potentiometer throttle position sensor used on a Ford Formula 1 racing engine.

Figure 8.5 Typical potentiometric position sensors for automotive use (courtesy Colvern Autosensors).

8.2 INDUCTIVE DISPLACEMENT TRANSDUCERS

Inductive position transducers do not suffer from the problems associated with a sliding contact, since they are inherently non-contact devices. The

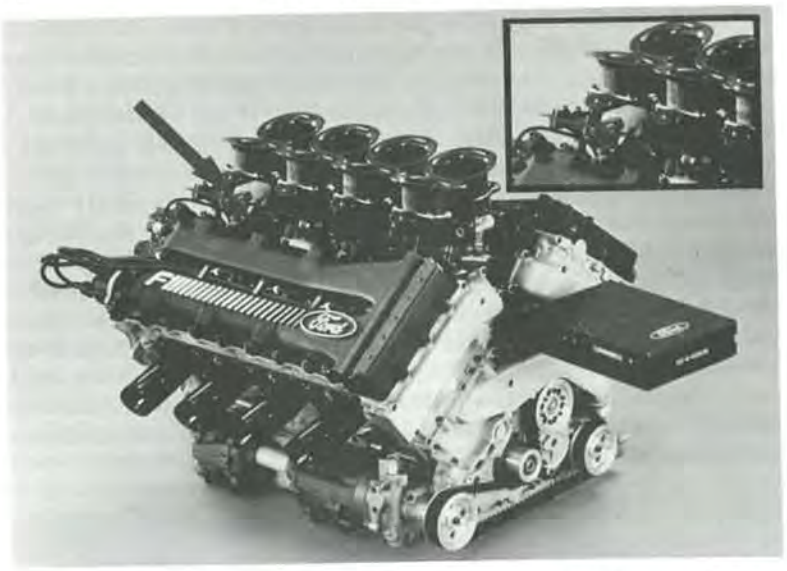

Figure 8.6 Ford racing engine fitted with potentiometer throttle sensor (courtesy Colvern Autosensors).

resolution available from a good-quality LVDT is equal to that obtained from a potentiometer. However, for automotive applications inductive sensors suffer from one inherent disadvantage: they are essentially AC devices, and since vehicle electrical systems are almost invariably DC at present some inconvenience and added cost is generally associated with their use.

8.2.1 Variable-reluctance Transducers

To understand how inductive displacement transducers work the concept of a *magnetic circuit* is required. In an electric circuit an electromotive force or voltage V drives current I through a resistance R. From Ohm's law the magnitude of the current is

$$V = IR. \tag{8.4}$$

In a magnetic circuit, such as the example shown in figure 8.7(a), a current i passes through a coil of n turns, which is wound onto a loop of ferromagnetic material. By analogy we can regard the coil as a source of *magnetomotive force* (MMF) which drives a flux ϕ through the magnetic

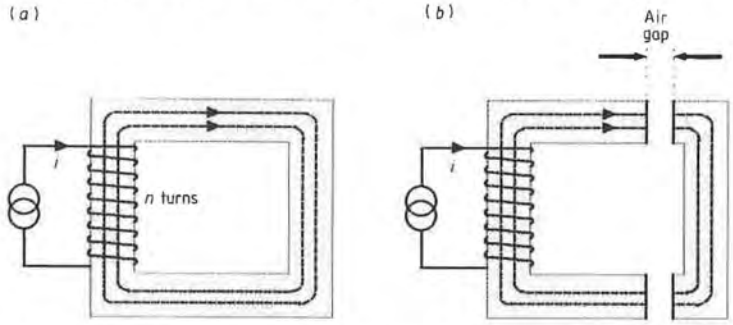

Figure 8.7 Variable-reluctance sensing: (*a*) simple magnetic circuit; (*b*) typical reluctance sensor.

circuit. The corresponding equation is

$$\text{MMF} = \text{flux} \times \text{reluctance} = \phi\,\mathfrak{R} \tag{8.5}$$

where $\mathfrak{R}$ is the *reluctance* which limits the flux in the magnetic circuit, just as resistance limits the current in an electric circuit. In the example shown in figure 8.7 the MMF is ni, so in this case the magnetic circuit flux ϕ linked by a single turn of the coil is

$$\phi = \frac{ni}{\mathfrak{R}} \text{ (Wb)}. \tag{8.6}$$

The total flux N linked by the entire coil of n turns is

$$N = n\phi = \frac{n^2 i}{\mathfrak{R}}.$$

By definition the self-inductance L of the coil is the total flux per unit current [2]. Therefore

$$L = \frac{N}{i} = \frac{n^2}{\mathfrak{R}}. \tag{8.7}$$

Equation (8.7) allows us to calculate the inductance of a sensing element given the reluctance of the magnetic circuit. The reluctance $\mathfrak{R}$ of a magnetic circuit is given by

$$\mathfrak{R} = \frac{\mathscr{L}}{\mu\mu_0 A} \tag{8.8}$$

where $\mathscr{L}$ is the total length of the flux path, μ is the relative permeability of the magnetic circuit material, μ_0 is the permeability of free space ($= 4\pi \times 10^{-7}\,\text{H m}^{-1}$), and A is the cross-sectional area of the flux path.

Figure 8.7(*b*) shows an arrangement similar to that of figure 8.7(*a*), except that the magnetic core has been separated into two parts by an air gap of variable width. The total reluctance of the circuit is now the sum of the reluctances of the two parts of the core, and of the air gap. The relative permeability of air (μ_{air}) is close to unity, while that of the core material can be several thousand times greater. Thus, the presence of the air gap causes a large increase in the circuit reluctance, and a corresponding decrease in flux and inductance. It is this effect which is used in sensor construction, since a small change in the width of the air gap causes an easily measurable variation in inductance. In automotive engineering variable-reluctance sensors are most frequently used for sensing rotation rate, where (as shown in figure 8.8) the sensor is placed close to a toothed wheel. Movement of the ferromagnetic teeth past the device causes a variation in the coupling between the coils. After suitable signal conditioning the output signal is produced with a frequency which is linearly related to rotation rate.

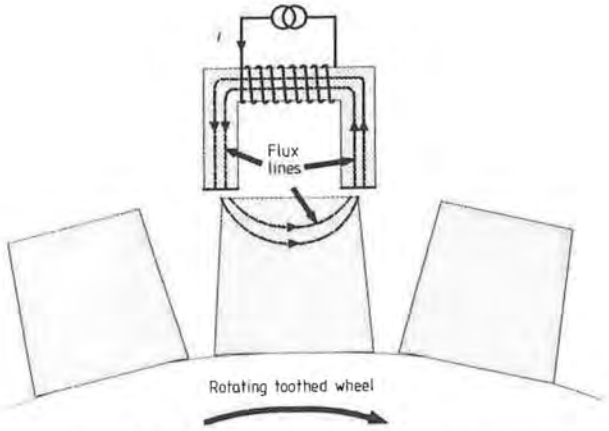

Figure 8.8 Variable-reluctance rotation rate sensor.

8.2.2 Variable-coupling Transformers: LDTs and LVDTs

A simple form of inductive displacement transducer known as a *linear displacement transducer* or LDT may be made by winding a pair of coils (or a single centre-tapped coil) onto a hollow cylindrical former, and allowing a ferromagnetic plunger to move along the axis as shown in figure 8.9(*a*). The plunger and coils have the same length *d*. As the plunger is moved the inductances vary. The two inductances L_1 and L_2 are normally connected in a bridge circuit with a pair of balancing resistors *R*, followed by an

amplifier as shown in figure 8.9(*b*). If the inductances are both L when the plunger is in the central position, and if the plunger and coil lengths are all d, a displacement of δd will produce opposing inductance changes $+\delta L$ and $-\delta L$. Ideally therefore $\delta L/L = \delta d/(d/2)$, and the corresponding bridge output will be $(V_{ex}/2)(\delta L/L)$.

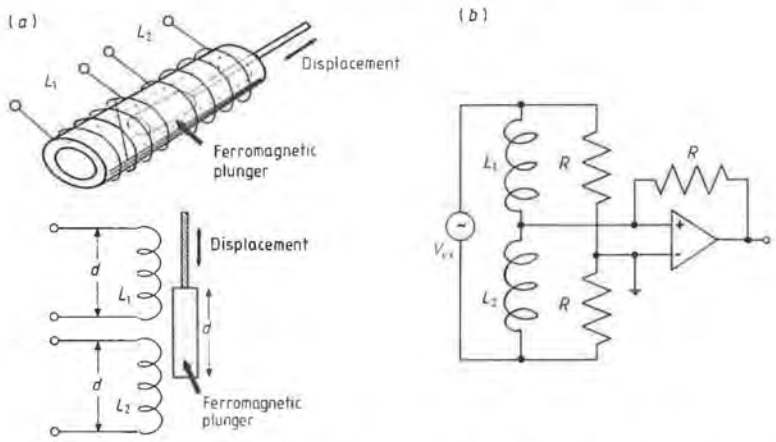

Figure 8.9 Linear differential transformer used as a displacement sensor: (*a*) linear displacement transducer; (*b*) linear differential transformer connected to bridge circuit.

A problem with the device described above is that the output is only linear over a limited region, when the plunger is close to the centre. An improved performance is obtained from the slightly more complex device known as a *linear variable differential transformer* or LVDT.

An LVDT is a transformer with a single primary winding and two secondaries, wound end-to-end on a tubular former as shown in figure 8.10. A ferromagnetic plunger (often a ferrite rod) moves inside the former, and varies the coupling between the primary and the secondaries. Both linear and rotational displacements can be sensed as shown in figure 8.10.

The centre (primary) coil is fed from an AC excitation supply, and induces voltage across the two outer (secondary) coils. The induced voltages have equal magnitudes when the plunger is positioned symmetrically. If the excitation voltage e_{ex} is $V_s\sin(2\pi f_s t)$, the output e_o is given by the difference $V_1 - V_2$ of the voltages induced in the two secondaries. The secondaries are normally connected in series opposition, as shown in figure 8.10, so that the output voltage is

$$e_o = V_1 - V_2 = V_{out}\sin(2\pi f_s t + \phi).$$

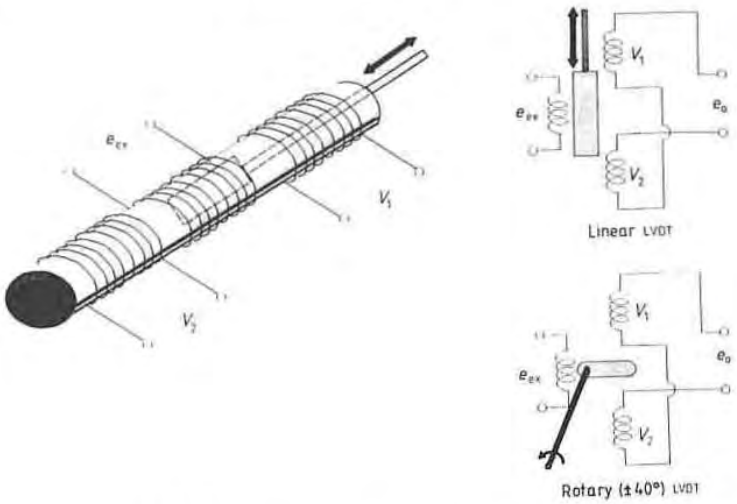

Figure 8.10 Linear variable differential transformer (LVDT).

An LVDT is usually excited by a sinusoid of up to 24 V in amplitude, with a frequency in the range 50 Hz to 25 kHz. With the series opposition arrangement a null position exists when the plunger is centrally placed for which the output $e_o \approx 0$. A displacement of the plunger away from the null position increases the coupling (mutual inductance) between the primary and one secondary, while decreasing the coupling for the other secondary. The amplitude of e_o is almost linear with respect to displacement for a considerable range either side of the central 'null' position as shown in figure 8.11.

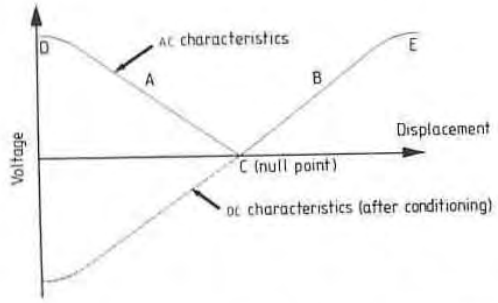

Figure 8.11 Amplitude of LVDT output voltage as a function of displacement.

e_o also undergoes a 180° phaseshift when the plunger is moved past the central position. An LVDT is usually used in conjunction with a signal conditioning circuit which converts the output signal into DC. The conversion is carried out in such a way that the regions A and B on figure 8.11 are distinguished, where the amplitude is the same but the phase difference is 180°. A *phase-sensitive demodulator* is used [3] to sense the phase difference, and gives a negative output voltage for displacements in the A region and a positive output for B. These signal conditioning circuits are sometimes incorporated within the casing of the LVDT, and the operation of the device is then 'transparent' to the user: a DC supply voltage is used to energise the device, and a DC position signal is returned.

Nonlinear effects occur at the extremes of travel as shown by regions D and E on figure 8.11. A typical nonlinearity figure for an LVDT is 1% over the whole range of the device.

LVDTs are available to cover ranges from ±0.25 mm to ±500 mm. The useful frequency range is limited by the inertia of the moving parts of the device. For automotive applications the main problem with an LVDT is the cost, which has so far precluded their use on production vehicles. However, LVDTs are very frequently used in laboratory and prototype development work.

8.2.3 Eddy Current Displacement Transducers

In appearance an eddy current transducer resembles a variable-reluctance sensor, but with one important difference: the 'target' used by a variable-reluctance device has to be ferromagnetic, while the object being sensed by an eddy current probe simply has to be electrically conducting. Thus, eddy current probes can be used in more situations than variable-reluctance devices.

An eddy current probe consists of a pair of coils, one active, which is influenced by the proximity of a conducting target, and a second balance coil which is included to complete a bridge circuit and provide temperature compensation.

The bridge is excited by a high-frequency (typically 1 MHz) AC signal, and magnetic flux lines from the active coil pass into the surface of the conducting target. Eddy currents are produced, largely at the surface, and diminish away from it. They are negligibly small a short distance below the surface.

As the target moves with respect to the probe the eddy currents vary, which changes the impedance of the active coil. A bridge imbalance results which is related to the distance between the active coil and the target. This imbalance signal is filtered and rectified to provide a DC voltage signal. The signal is inherently nonlinear, and linearisation circuits are sometimes

employed to remedy this defect.

Probes are commercially available to cover the ranges from ±0.25 mm to ±30 mm, with a maximum resolution of 10^{-4} mm. The recommended measuring range for a given probe begins at a 'standoff' distance which is normally about 25% of the range. Thus, a probe with a ±1 mm range is mounted with a static target/probe distance of between 0.5 and 1.5 mm.

Targets are not supplied with eddy current probes, since an existing machine part is normally used. The motion of a non-conducting target can be sensed if a piece of conducting foil is fixed to the surface. Adhesive-backed aluminium foil tape is commercially available for this purpose.

Flat targets should be the same diameter as the probe, or larger if possible. Curved targets (such as circular shafts) behave like flat surfaces if the shaft diameter exceeds five transducer diameters. Special four-probe systems are available for measuring the orbital motion of rotating shafts.

Eddy currents transducers have not been reported in use on a production vehicle, probably due to the rather complex and expensive signal conditioning arrangements required. However, systems of this type are in frequent use in laboratories concerned with rotating machine dynamics, and are commonly used for engine development work where their good high-frequency response is essential.

8.3 CAPACITIVE DISPLACEMENT TRANSDUCERS

A capacitor consists of two electrically conducting components separated by a dielectric (non-conducting) medium. When a voltage V is applied across the capacitor, equal and opposite charges $\pm Q$ appear on the two conducting components. The ratio of charge to voltage defines the capacitance of the device

$$C = \frac{Q}{V}.$$

If the capacitor consists of two parallel conducting plates of area A, separated by a distance d, then the capacitance C is

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \text{ (farad (F))} \tag{8.9}$$

where ε_0 is the permittivity of free space (a vacuum) which has a value of 8.85×10^{-12} F m^{-1}, and ε_r is the relative permittivity of the dielectric material separating the conducting plates. For dry air $\varepsilon_r = 1$.

Capacitance is therefore a function of geometrical parameters (A and d) as well as the material property ε_r. Any change in A, d or ε_r causes a corresponding change in the capacitance as shown by equation (8.9). This fact is used in the design of capacitive displacement sensors, several versions of which are shown in figure 8.12. Capacitive sensing is common where accurate measurement of small displacements is required, such as in pressure and acceleration sensing (see chapters 3 and 9).

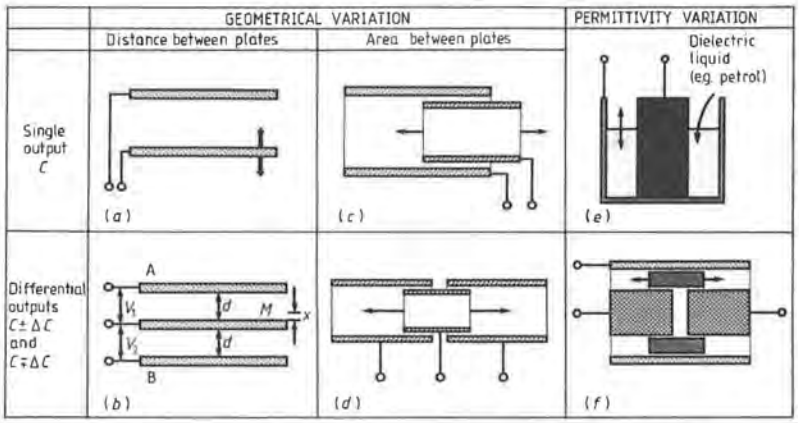

Figure 8.12 Capacitive displacement transducers: (*a*) single pair of plates; (*b*) three-plate arrangement; (*c*) concentric tubes; (*d*) differential concentric tubes; (*e*) liquid level gauge; (*f*) movable dielectric ring.

Single or differential outputs can be used, although in practice it is often inconvenient to employ only two plates since the output is nonlinear as discussed below. A three-plate differential arrangement is normally used. There are several common geometries for both translational and rotary displacement sensing, some of which are shown in figure 8.12. The nominal capacitance of this type of sensor is generally below 1000 pF (1 nF). To reduce the impedance levels and to allow the determination of high-speed movement, the frequency of the signal used for measurement is usually 100 kHz or greater. The transducer impedance at this frequency is still quite high, so the insulation resistance must also be high to avoid 'shunting' the transducer unduly and reducing its sensitivity. A charge amplifier is almost invariably used (see chapter 9). Capacitive displacement transducers are normally used to measure displacements of less than 1 mm.

Capacitive sensors are often unreliable when used in conditions where the humidity may vary. This is probably why their use in automotive

engineering is largely confined to pressure and acceleration sensors, where the capacitive element can be sealed in a vacuum or enclosed in an atmosphere of dry gas.

A sensor in which the plate area A or separation d is varied requires a physical connection between the motion being sensed and the moving part of the capacitor (see figure 8.12). Transducers based on permittivity variation ε_r on the other hand do not, and this has led to their use being preferred in some circumstances.

8.3.1 Linearity

In a simple two-plate arrangement such as that shown in figure 8.12(a), the variation of C as d alters is hyperbolic and can only be treated as linear for small excursions. In this case the sensitivity $\Delta C/\Delta d$ is proportional to $1/d^2$.

If a three-plate differential arrangement such as that shown in figure 8.12(b) is used, a linear output can be obtained. Referring to the diagram, it is apparent that the fixed plates A and B and the movable plate M form a pair of capacitors with values C_1 and C_2. When M is centrally positioned midway between A and B, a distance d from each, $C_1 = C_2$. If M is moved a distance x towards A the capacitances become

$$C_1 = \frac{\varepsilon_0 \varepsilon_r A}{d + x} \quad \text{and} \quad C_2 = \frac{\varepsilon_0 \varepsilon_r A}{d - x}. \tag{8.10}$$

If a voltage V_{ex} is applied across AB, the output V_{out} is given by

$$V_{out} = V_1 - V_2 = V_{ex}\left(\frac{C_2}{C_1 + C_2} - \frac{C_1}{C_1 + C_2}\right) = V_{ex}\frac{x}{d}.$$

The three-plate arrangement is therefore linear, with V_{out} proportional to x, while the sensitivity V_{out}/x is inversely proportional to d.

8.4 OPTICAL MOTION SENSORS

Angular and translational (linear) motion are often measured optically using pulse-counting methods. Optical position transducers show signs of becoming popular in automotive applications for the following reasons:

(i) pulse counting is inherently digital;
(ii) optical devices are immune from electrical interference;

(iii) no mechanical connection to the sensing element is made;
(iv) they can be of low cost, especially if plastic optical components are used.

Their main drawbacks are that they are relatively fragile, and that their performance suffers badly if dirt is allowed to get onto the optical components. For these reasons they have so far largely been restricted to applications such as steering-wheel position and velocity detection, where the sensor can be mounted in a relatively clean environment and does not experience the extremes of temperature found in the engine compartment. Examples of this kind of use are described in [4] and [5].

8.4.1 Angular Encoders

The simplest form of angular encoder consists of a slotted or striped disc, a light source and a photosensitive detector. The slots in the former 'chop' the light falling on the detector. If a reflective disc with matt segments is used, light is either reflected or not. With a reflective system the light source and detector can be placed on the same side of the disc, and if a slotted version is used they are placed on either side. The resulting pulse train indicates rotational speed. If the disc rotates at n revolutions per minute, the output pulse rate is $nS/60$ Hz, where S is the number of slots in the disc. Simple devices such as this cannot give any indication of the direction of rotation, nor in general can they measure angular position unless a starting reference point is provided and pulse-counting techniques used.

The difficulty over sensing the direction of rotation may be overcome if two tracks and two sensing heads are used. The tracks are positioned so that their electrical outputs have a phase displacement which lags or leads, depending on the direction of rotation.

The transducers described above are usually referred to as *incremental angular encoders*, since unless the start-up point is known it is very difficult to estimate the absolute position. This difficulty is overcome by *absolute angular encoders* such as the four-bit examples shown in figure 8.13. These devices consist of a number of concentric tracks containing a pattern of opaque and transparent (or reflective) sections, such that a unique binary code can be read from the patterns for any angle. The patterns are read by means of photocells as before.

If natural binary coding is used, as on the disc shown in figure 8.13(a), a potentially large error can occur if a readout is attempted when the photocells lie along a line midway between two sectors of the disc. It can be seen from the diagram that, for some positions, a number of tracks

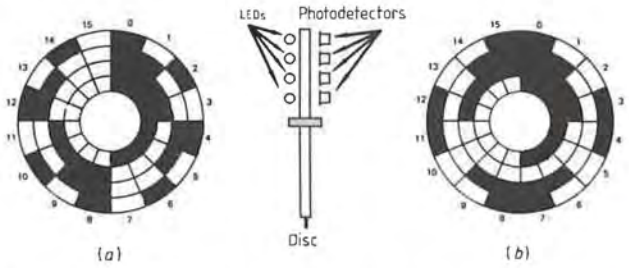

Figure 8.13 Binary (*a*) and Gray code (*b*) absolute encoders.

change state simultaneously (these are known as the *cardinal transitions*). This problem does not arise if Gray code is used, as shown on figure 8.13(*b*), since only one bit changes state at each sector. It is therefore impossible to generate an output which is in error by more than one bit.

Absolute encoders are available with 8, 10, 12 or 16 tracks, but they become increasingly expensive and inconveniently large as the number of tracks increases. The resolution obtained from an n-track absolute encoder is $360/(2^n)°$. Thus, an eight-track device for example is divided into 256 ($= 2^8$) sectors, and can resolve to within $360/256 = 1.41°$.

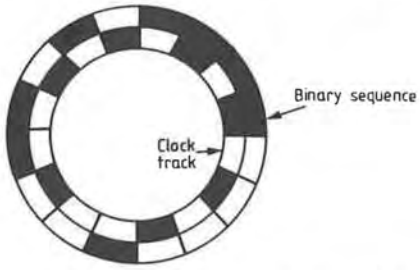

Figure 8.14 Absolute encoder with four-bit maximal-length binary sequence.

An alternative approach which dispenses with the need for n tracks to obtain an accuracy of $360/(2^n)°$ is to use a single track containing a maximal-length binary sequence of 2^n digits for n bits. Such a sequence contains all the possible combinations of n bits exactly once, and every angular readout (determined by n adjacent bits) is unique. A separate clock track is often used with this scheme. Figure 8.14 shows a simple four-bit version of this arrangement.

8.4.2 Translational Encoders

The arrangements described above for angular displacement sensing can also be applied to translational motion. A device commonly used in high-precision displacement sensing, which has yet to find an automotive application, consists of two diffraction gratings as shown in figure 8.15(*a*). One part (the scale grating) is fixed, while the other (the index grating) is attached to the moving object and slides over the fixed grating. As the index grating moves the appearance of the device alternates between light and dark, and this change is readily sensed by means of a photocell. The photocell does not have to be the same size as the lines on the grating, and indeed there are advantages in making it much larger, since any errors introduced by imperfections in individual lines are automatically averaged out.

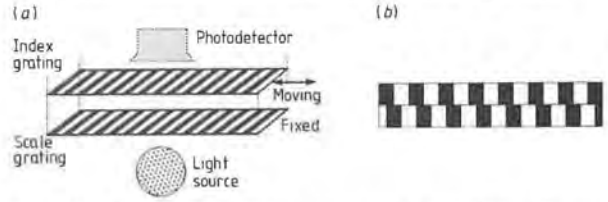

Figure 8.15 Translational optical sensors: (*a*) grating displacement sensor; (*b*) two-phase index grating.

Just as for angular position sensing, a single photocell feeding a counting or frequency measuring system gives no indication of the direction of movement. If this is required the index grating is made in two parts as shown in figure 8.15(*b*), where the two sets of lines are displaced by one-quarter of a line spacing. The resulting quadrature signals are often used to feed a direction detector and an up/down counter.

Absolute translational encoders are also available. These operate in much the same way as the angular versions, and consist of a number of parallel tracks containing a pattern of opaque and transparent (or reflective) sections, such that a unique binary code can be read from the patterns for any angle. Once again Gray code is used to reduce errors.

8.5 ULTRASONIC DISPLACEMENT TRANSDUCERS

The term *ultrasound* refers to sound waves generated at frequencies higher than the human ear can detect, i.e. at frequencies above about 18 kHz. Ultrasonic waves obey the same basic laws of wave motion as lower-frequency acoustic waves [6]. They have, however, the following advantages:

(i) Higher frequencies imply shorter wavelengths, which means that diffraction does not occur so readily. It is easier to produce a directed and focused beam of ultrasound than it is for 'ordinary' sound.

(ii) Ultrasound waves pass easily through the metal walls of a structure. This means that for applications such as fuel-tank level sensing the measurement system may be mounted externally.

Ultrasound has been used for distance measurement or rangefinding for many years. Examples with which the reader will be familiar include marine sonars, medical ultrasound imaging equipment and defect location systems for use in structural materials such as steel and concrete. In each case brief pulses of ultrasound are emitted, and the range of the object under investigation is deduced by timing the echo.

The possibility of using an ultrasonic echo method for obstacle detection was first suggested by L F Richardson at the time of the *Titanic* disaster in 1912 [6]. The idea was developed intensively during the First World War with the aim of detecting enemy submarines, and the first practical system was developed by P Langevin. Figure 8.16 shows a diagram of Langevin's quartz ultrasonic transducer, which was used both as a transmitter and receiver. A beam of ultrasound was propagated vertically downwards into the sea and reflects from the bottom, or from any intervening object such as a submarine. Langevin's work forms the basis of all modern sonar (*so*und *n*avigation *a*nd *r*anging) systems, both for military uses and for applications such as navigation and fish detection.

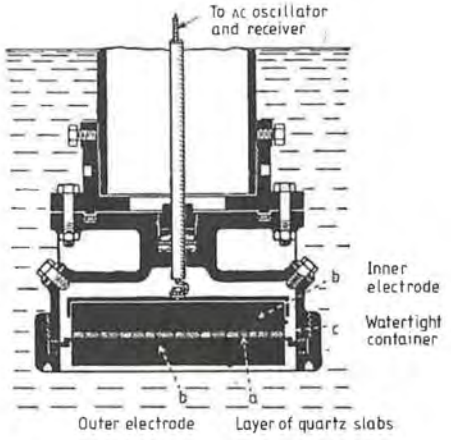

Figure 8.16 Langevin's ultrasound transducer.

Sonar was the first engineering application of ultrasound, and its development preceded that of radar by more than 30 years. The development of radar led to a number of improvements in electronic technology, many of which were in turn applied to ultrasound systems. One of the most important of these developments was the possibility of using a phased array [8] to control the direction of the ultrasound beam.

Radar cannot be used underwater because (since seawater is conducting) electromagnetic waves are very rapidly attenuated. The attenuation of compression waves in water, however, is very low, and submarine sonars can operate over distances of several kilometres. The attenuation of ultrasound in air is much greater than in water, but practical ultrasonic ranging systems can still be built which operate satisfactorily over ranges of tens of metres. The attenuation of ultrasound in air is greatest at high frequencies, so in airborne applications frequencies of the order of tens of kHz are used. For example, airborne ultrasound is used to provide automatic focusing in some cameras, and for this application tone bursts with a frequency between 20 and 25 kHz are used.

It automotive engineering ultrasound has been used to measure the *ride height* of a vehicle for adaptive-suspension control, to monitor the road profile of a vehicle in advance of the wheels, again for adaptive-suspension control, and for local collision warning and avoidance systems (often called parking or reversing aids).

A number of aircraft fuel gauging systems use ultrasound for tank level sensing. This development was introduced because it offers the possibility of non-invasively (and therefore safely) measuring the amount of liquid in a fuel tank. Sophisticated systems have been developed in which measurements from a number of transducers are integrated to give a reading which is independent of the motion of the fuel 'sloshing'. It seems likely that, as has happened before, a technology originally developed for aerospace will eventually be used in automotive engineering.

To date the only commercial application of ultrasound on a production vehicle has been for measuring the ride height as part of an adaptive-suspension system. Ultrasonic transducers for this purpose have been fitted to some Japanese and US vehicles, but have not so far made a widespread appearance on those manufactured in Europe.

8.6 HALL EFFECT SENSORS

If a conducting or semiconducting material carrying a current is placed in a magnetic field B, which is normal to the direction of current flow I as shown in figure 8.17, a voltage V is produced across the width of the conductor. The effect is greatest in semiconductors, and these are

normally used in commercial Hall effect devices. Briefly, the Hall voltage appears because the magnetic field causes the current carriers to follow a curved path as shown in the diagram. An excess of current carriers appears along one edge of the device, while a shortage occurs at the opposite side. This charge imbalance gives rise to the Hall voltage, which remains as long as the magnetic field is present and as long as the current is maintained. For a rectangular (semi)conductor of thickness t the Hall voltage V is given by

$$V = \frac{K_H BI}{t} \tag{8.11}$$

where K_H is the Hall coefficient for the material, which depends on the charge mobility and resistivity of the conductor. Iridium antimonide is widely used in Hall effect devices, for which $K_H = 20\ V\,T^{-1}$.

Hall effect transducers may be used for non-invasive current measurement, for magnetic field measurement†, or for inferred distance, velocity or proximity sensing. This last group of applications are the most common in automotive engineering, where Hall probes are often used to sense rotation rate by means of a toothed wheel.

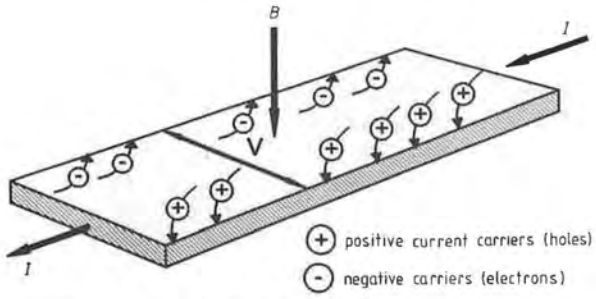

Figure 8.17 The Hall effect in a (semi)conductor.

8.6.1 Hall Probe Rotation Rate Sensors

Hall pick-ups for sensing rotation rate sometimes use a permanent magnet fixed to the moving object, such as in the tyre pressure sensing system mentioned in chapter 3 and discussed in detail in [8]. More frequently a

†Hall effect solid state compasses are available. These may have applications in the area of in-car navigation systems. However, just as for other magnetic forms of compass, care has to be taken to eliminate the effects of non-terrestrial magnetic fields arising from, for example, overhead power cables.

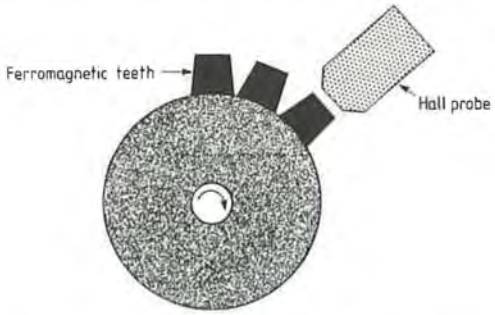

Figure 8.18 Hall probe used to measure the rate of rotation of a toothed wheel.

ferrous target (such as a toothed wheel) is used, whose approach changes the reluctance of an internal magnetic circuit which controls the Hall probe output. This kind of system is frequently used to sense wheel speed for ABS and traction control systems [9]. A typical arrangement is shown in figure 8.18.

8.6.2 Hall Probe Distance Measurement

Hall effect transducers give an output voltage which is linearly and repeatably related to the magnetic field intensity. If a suitably shaped permanent magnet is used the Hall probe output can be made inversely proportional to the distance from the permanent magnet. This effect can be exploited to give a non-contact displacement sensor which is reasonably accurate (typically $\pm 1\%$). A permanent magnet is fixed to the moving object as shown in figure 8.19, while the Hall probe is fixed. Distance measurements can be carried out using this approach within the range 0–20 mm. However, care should be taken to provide adequate magnetic shielding, since errors will be introduced if the sensor is exposed to transient magnetic fields. This is probably the reason why only a few

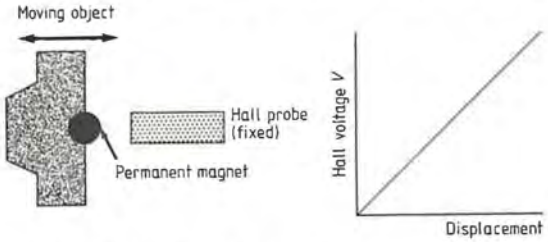

Figure 8.19 Hall transducer used for measurement of displacement.

automotive uses of this technique have been reported. One example is discussed in [10], in which a Hall effect accelerometer is described. A permanent magnet acts as the seismic mass in this device, and the Hall probe senses its motion.

8.7 BRAKE-PAD AND CLUTCH WEAR SENSORS

A number of automotive components rely upon dry friction for their operation. The most well known examples are in clutch and braking systems, where pads or shoes of a composite material (formerly asbestos) are forced together or against a steel surface by hydraulic pressure. Since a large amount of friction results the component wears away, and the replacement of these parts was formerly a regular feature of routine vehicle maintenance.

Improvements in materials and design have extended the life of, for example, brake pads, to the point at which they can now last for several years. One consequence of this improved life is that failure of the component due to wear is often unexpected, since it happens very infrequently. A means of warning the operator when, say, 90% of the material has been abraded is therefore necessary, to ensure that remedial action is taken before it becomes critical or damage has occurred.

The problem is essentially one of sensing a displacement, since as the brake shoe or clutch plate wears away its metal backing moves steadily nearer the abrasion surface. The approach normally adopted for brake pads and shoes is to embed a network of fine wires within the abradable material. These wires carry a small current, which is used to test for continuity. When the component is worn away to the point where the wires become broken the current is interrupted, and an appropriate dashboard warning lamp can be illuminated.

Clutch plates present a slightly more difficult problem, since they rotate, which makes it difficult to use the broken-wire technique. A non-contact technique is probably the most appropriate in this case. A method which has been successfully demonstrated [11] uses a stationary Hall probe and a permanent magnet fixed to the rim of the clutch plate. As the clutch rotates the Hall probe output appears as a series of pulses, with the engine speed determining the frequency. The pulse amplitude, however, is inversely proportional to the distance between the Hall probe and the magnet, and this fact can be used to trigger an alarm when the appropriate threshold amplitude is exceeded. A subsidiary benefit of this system is that the pulse rate can be used to give an electronic indication of engine speed.

REFERENCES

[1] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 44551 1
[2] Grant I S and Phillips W R 1975 *Electromagnetism* (New York: Wiley) ISBN 0 471 32246 6
[3] Noltingk B (ed) 1990 *Instrumentation Reference Book* (London: Butterworths) ISBN 0 408 01562 4
[4] Goodyer E N 1989 An overview of a range of novel automotive sensors *Proc. 7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 79–86 ISBN 0 85295 697 1
[5] Smith D S and Jackman P R 1992 Optical sensors for automotive applications *IEE Computing & Control Division Coll. 1992/107 on Automotive Sensors* (London: IEE) pp 2/1–2/3
[6] Turner J D and Pretlove A J 1991 *Acoustics for Engineers* (London: Macmillan) ISBN 0 333 52143 9
[7] Cracknell A P 1980 *Ultrasonics* (London: Wykeham) ISBN 0 85109 770 7
[8] Hill M, Sadek H and Turner J 1992 The development and testing of a car tyre pressure sensor *Proc. 25th ISATA Conf. on Mechatronics* (Automotive Automation) pp 57–64 ISBN 0 947719 49 0
[9] Podeswa R and Lachmann U 1989 A differential Hall IC for geartooth sensing *Proc. 7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 93–8 ISBN 0 85298 697 1
[10] Bicking R E 1989 A Hall effect accelerometer *Proc. 7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 105–8 ISBN 0 85298 697 1
[11] Sheldrake C D 1991 Clutch wear sensor *BEng Project Report* (Southampton University Mechanical Engineering Department)

9

Accelerometers

An accelerometer is an electromechanical transducer which generates an electrical output when subjected to mechanical shock or vibration. Accelerometers are inertial sensors which make measurements by virtue of Newton's second law (force = mass × acceleration). Unlike displacement and velocity, which are usually determined with respect to an arbitrary reference point, acceleration can be measured on an absolute basis.

Shock and vibration measurements are vital to the development, testing and operation of structures and machines in all fields of engineering. Accelerometers are widely used for this purpose because of their accuracy, robustness, wide frequency response and sensitivity compared to other forms of vibration sensor. In general accelerometers are smaller, lighter and easier to install than other types of vibration sensor, such as seismometers (velocity transducers) or displacement sensors. In automotive applications accelerometers are mainly used for suspension control, for crash detection (to trigger the inflation of airbags etc), and for anti-knock systems (where engine vibration is monitored).

In this chapter the design characteristics, operating principles and limitations of the most common forms of accelerometer are described. Clearly, few if any users will wish to design their own accelerometer. However, understanding how a sensor operates is an essential prerequisite to its intelligent selection and use. Section 9.1 gives an introduction to the fundamental theory of accelerometer operation.

Most of the accelerometers used today are of either the piezoelectric (PE) or piezoresistive (PR) type, although capacitive types are likely to become widespread (see section 9.2.3). The piezoelectric effect is a phenomenon exhibited by certain crystalline and ceramic materials, in which a potential difference appears across opposite sides of a block of the material as a result of mechanical deformation. Piezoelectric accelerometers consist of a small inertial mass mounted on a piece of PE material. When the assembly is accelerated, Newton's second law dictates that the mass applies a force to the PE mount. The resulting deformation causes a voltage to appear, which can be amplified and used

to measure the magnitude of the acceleration. There are a number of different designs of PE accelerometer, and these are described in section 9.2.1.

The piezoresistive effect occurs when a material changes its electrical resistance in response to an applied stress. Piezoresistance changes should not be confused with the resistance changes due to dimensional alteration, such as those that occur in conventional wire or foil strain gauges. In a foil strain gauge the resistance change is mainly due to elongation of the gauge in its active direction. In a PR gauge the dimensional changes are small, and the measured resistance change is essentially a function of applied stress rather than strain.

Piezoresistive accelerometers are constructed by bonding PR strain gauges to the root of a small cantilever. A seismic mass is often fixed to the free end of the cantilever to enhance the sensitivity. The strain gauges are connected in a Wheatstone bridge circuit, and the resulting out-of-balance voltage is proportional to acceleration. The advantage of PR accelerometers is that their frequency response extends down to DC, making them suitable for low-frequency work. However, they are not usually as rugged as PE devices, and they can be more expensive. PR accelerometers may be micromachined from silicon, and can be made very small. Further details are given in section 9.2.2.

An alternative, cheaper approach to PR accelerometer production is to use thick-film printing. While sensors manufactured using thick-film technology are generally larger than those made from silicon, the reductions in unit and set-up costs can be considerable. If necessary signal conditioning circuits can also be incorporated within the sensor housing by means of thick-film hybrid construction. Further details of thick-film PR accelerometers are given in section 9.2.2.

A third phenomenon which is sometimes used as the basis of an accelerometer design is capacitance variation. A cantilever, or some other beam configuration, carries a seismic mass and forms one-half of a parallel-plate capacitor. The beam moves in response to acceleration, and this alters the plate separation, which changes the capacitance. Variable-capacitance accelerometers can be micromachined from a block of single-crystal silicon. The resulting sensors are very small and rugged and can also incorporate signal conditioning circuits. The frequency response extends down to DC. However, apart from switching applications such as airbag triggering, the use of capacitive accelerometers is usually restricted to measurements of low acceleration levels, since large changes in plate separation can cause nonlinearities in the output.

Confusion can arise over the best form of accelerometer to use in a given situation. Section 9.2 is therefore concluded by a table to assist with sensor selection, in which the advantages and disadvantages of each type of sensor are highlighted.

When using an accelerometer (or any other form of sensor) the effect of environmental changes on the output signal must be considered. Changes in the thermal, magnetic or acoustic environment in which the transducer is situated may cause changes to the output signal. Accelerometers can also give erroneous outputs because of base strain if incorrectly mounted, from humidity changes, or from electromagnetic interference (EMI). These topics are discussed in sections 9.3 and 9.4.

Piezoelectric and capacitive devices give a charge rather than a voltage output. Piezoresistive accelerometers are electrically equivalent to the Wheatstone bridge, and therefore require a different form of signal conditioning. Section 9.5 contains a discussion of the signal conditioning circuits appropriate to different types of accelerometer.

9.1 THEORY OF OPERATION

Most vibration sensors are based on the damped mass-spring system shown in figure 9.1. Depending on whether the frequency to be measured, ω, is higher or lower than the natural resonance frequency of the sensor, ω_n, displacement, velocity or acceleration may be measured. Transducers of this type operate by sensing the motion of the suspended mass M relative to the case.

The equation of motion for the system of figure 9.1 is

$$m\ddot{x} = -c(\dot{x} - \dot{y}) - k(x - y) \tag{9.1}$$

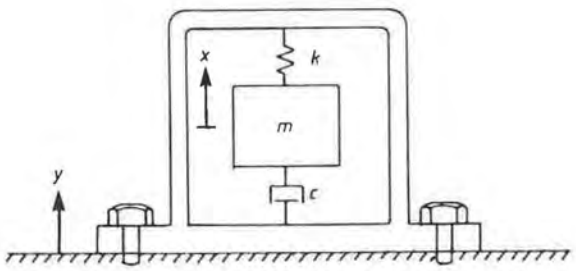

Figure 9.1 A damped mass-spring system. See text for notation.

where x and y are the displacement of the seismic mass and the vibrating body respectively. If the relative displacement between the seismic mass and the case is

$$z = x - y \tag{9.2}$$

and if the vibrating body to which the case is attached undergoes sinusoidal motion of the form $y = A \sin \omega t$, then equation (9.1) becomes

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 A \sin\omega t. \tag{9.3}$$

This is the well known equation for the response of a system with a single degree of freedom to forced vibration. The steady-state solution can be assumed to be of the form $z = Z \sin(\omega t - \phi)$ where z is the amplitude of the mass-case displacement and ϕ is the phase of the displacement with respect to the exciting force. By comparison with the standard solutions we can write down equations (9.4) and (9.5)

$$z = \frac{m\omega^2 A}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} = \frac{A\,(\omega/\omega_n)^2}{\sqrt{[1 - (\omega/\omega_n)^2]^2 + [2\zeta\,(\omega/\omega_n)]^2}}. \tag{9.4}$$

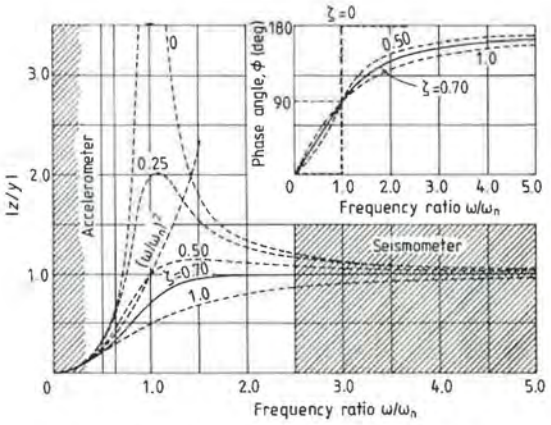

Figure 9.2 Plot of equations (9.4) and (9.5).

$$\tan\phi = \frac{\omega c}{k - m\omega^2} = \frac{2\zeta\,(\omega/\omega_n)}{1 - (\omega/\omega_n)^2} \quad (9.5)$$

It is obvious from these expressions that the important parameters are the frequency ratio ω/ω_n and the damping factor ζ. Figure 9.2 shows a plot of these equations. The type of sensor is determined by the relationship between the frequency to be measured and the resonance frequency of the transducer.

9.1.1 Vibration Velocity Sensors (Seismometers)

When the natural frequency of the sensor ω_n is low compared with the vibration frequency being measured ω, the ratio ω/ω_n is large. The amplitude of the relative displacement Z approaches that of the vibration A, regardless of the value of the damping ratio ζ. In these circumstances the mass m remains stationary while the surrounding case moves with the vibrating body. Sensors of this type are called seismometers. The relative motion z is usually converted to a voltage by making the seismic mass a permanent magnet which moves relative to coils fixed to the case. Since the voltage generated is proportional to the rate at which the coils cut the magnetic field lines, the output of the sensor is proportional to the vibration velocity. A typical seismometer will have a resonance frequency between 1 and 5 Hz, and a useful bandwidth from 10 Hz to around 1 kHz.

The main disadvantage of seismometers as vibration sensors is their large size. Since $Z = A$, the relative motion of the seismic mass has to be of the same order as that of the vibration being measured. The sensor housing must be large enough to accommodate this motion.

9.1.2 Accelerometers

When ω_n is high compared to ω the sensor output is proportional to acceleration. Examination of equation (9.4) shows that the factor

$$\sqrt{[1 - (\omega/\omega_n)^2]^2 + [2\zeta\,(\omega/\omega_n)]^2}$$

approaches unity for $\omega/\omega_n \to 0$, so that

$$Z = \frac{\omega^2 A}{\omega_n^2} = \frac{\text{acceleration}}{\omega_n^2}. \quad (9.6)$$

Thus Z, the displacement amplitude of the seismic mass with respect to

the case, is proportional to the acceleration of the motion being measured with a factor $1/\omega_n^2$. The useful range of the accelerometer can be seen from figure 9.3, which is a plot of

$$\frac{1}{\sqrt{[1-(\omega/\omega_n)^2]^2+[2\zeta(\omega/\omega_n)]^2}}$$

for various values of damping ζ. The diagram shows that the useful range of frequencies for an undamped accelerometer is severely limited. However, when $\zeta = 0.7$ the useful frequency range extends to around 20% of the resonance frequency, and within this range the maximum error is less than 0.01%.

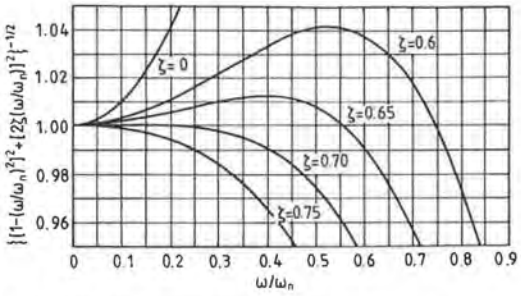

Figure 9.3 The useful range of an accelerometer. See text for notation.

Phase distortion. The time delay between applying a mechanical input to an accelerometer and the appearance of the resulting electrical output is known as the phase shift. If this delay is not the same for all frequencies contained in the mechanical input, the phase relationship between the frequency components of the vibration waveform will be altered, and the resulting electrical output will become distorted. This effect is known as phase distortion, and to avoid it we require either that the delay be zero, or else that all frequency components are delayed by the same amount. The first case, a zero delay or zero phase shift, corresponds to $\zeta = 0$ and $\omega/\omega_n < 1$ (see equation (9.5)). However, as we have just seen, zero damping is undesirable in an accelerometer.

The second case, an equal timewise phase shift applied to all frequency components, is almost satisfied when $\zeta = 0.7$ and $\omega/\omega_n < 1$. It can be seen from figure 9.2 that when $\zeta = 0.7$ and $\omega/\omega_n < 1$, the phase angle ϕ is given approximately by

$$\phi = \frac{\pi}{2}\,\frac{\omega}{\omega_n}.$$

Thus for $\zeta = 0.7$ (and $\zeta = 0$) phase distortion is almost eliminated.

Accelerometer resonance frequencies. The resonance frequency of an accelerometer is not constant, although a constant value will be specified on the sensor calibration chart. It depends not only on the seismic mass and the stiffness of the PE or other transducers to which it is attached, but also on the mass and stiffness of the vibration test object to which the device is attached, and to some extent on the stiffness of the mounting method used.

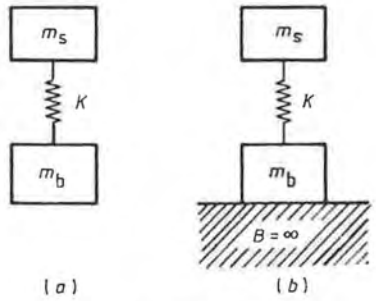

Figure 9.4

The situation is illustrated in figure 9.4. A seismic mass m_s rests on a transducer, such as a piezoelectric element, which is attached to the transducer base. K is the equivalent stiffness of the transducer(s) and their connection to the base. The mass of the base and housing is m_b. When the accelerometer is not coupled to any other object, as shown in figure 9.4(a), the resonance frequency f_r is

$$f_r = f_s \sqrt{1 + \frac{m_s}{m_b}} \tag{9.7}$$

where f_s is the resonance frequency of the seismic mass m_s on the stiffness K, which is given by

$$f_s = \frac{1}{2\pi} \sqrt{\frac{K}{m_s}}. \tag{9.8}$$

From equation (9.7) it can be seen that the 'free-hanging' resonance frequency depends upon the ratio of m_s to m_b. At first sight therefore equation (9.7) seems to imply that the base of the accelerometer should be made as light as possible (m_b small), so that f_r is as large as possible and the usable bandwidth extended. However, when the accelerometer is mounted on a test object of large mass and stiffness as shown in figure 9.4(b), m_b tends to infinity, and the accelerometer resonance approaches f_s.

There is therefore no point in trying to make the sensor base light. Other considerations, such as the susceptibility to base strain discussed in section 9.3, dictate the use of a stiff, and consequently heavy, base.

9.2 ACCELEROMETER DESIGNS

There are many different types of accelerometer. In each design the transducer relies on detecting the motion of a seismic mass to measure acceleration, as discussed in section 9.2. However, there are differences in the way in which this motion is sensed. These differences give rise to variations in parameters such as the bandwidth, sensitivity or susceptibility to interference of the sensor. Thus, some designs are well suited for some applications but will give poor performance in others. It is important therefore that the automotive engineer is aware of all the features, good and bad, of the many competing forms of accelerometer, so that problems due to an inappropriate choice are avoided. In this section the intention is to survey the types of accelerometer available, and to highlight the good and bad features of each type.

9.2.1 Piezoelectric accelerometers

A piezoelectric accelerometer contains two elements: the mass, on which the acceleration acts to produce a force, and a piezoelectric transducer, which converts the force into electric charge. In one of the most common designs (of which an example is shown in figure 9.5(*a*)), the mass rests on a number of piezoelectric discs. The mass and discs are pre-loaded by a spring and the whole assembly is sealed inside a housing. When the accelerometer is vibrated along its axis the mass exerts a force on the PE material, which develops a variable charge in proportion to the applied force. For frequencies below that of the resonance of the transducer the acceleration of the mass is equal to that of the whole accelerometer.

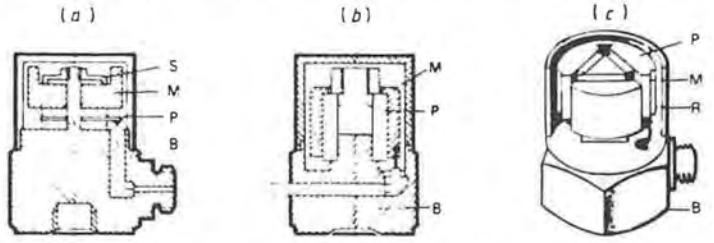

Figure 9.5 Designs for piezoelectric accelerometers: (*a*) centre-mounted compression (CM) type; (*b*) annular shear (AS) type; (*c*) delta shear (DS) type (courtesy of Bruel and Kjaer).

A PE material is one which develops an electrical charge when it is subjected to stress. Naturally occurring PE materials include single crystals of quartz and Rochelle salt. However, the PE effect is small in these substances. PE accelerometers normally use man-made PE compounds known as ferroelectric ceramics, which have a much higher responsivity. These consist of mixtures of barium titanate, lead zirconate and lead metaniobate. Ferroelectric ceramics have the further advantage that, unlike crystalline materials, they can be produced in any desired shape or size. The sensitivity of PE materials to stress is temperature dependent, and ceases completely above a critical temperature known as the *Curie point*. Curie point temperatures range from 120 °C (for barium titanate) to over 600 °C.

Three different mechanical constructions are used for PE accelerometers. The most common is the *centre-mounted compression* (CM) type, an example of which has already been shown in figure 9.5(*a*). This construction gives moderate sensitivity and can withstand high levels of continuous vibration or shock without damage. In a CM accelerometer the piezoelectric mass-spring system is mounted on a cylindrical centre post attached to the base of the accelerometer, which increases the resonance frequency and thus the usable bandwidth. However, despite the use of a thick base, CM accelerometers are more sensitive to base strain and thermal transients than other PE designs.

Since no bonding is used in the construction of a CM accelerometer, this type can be used at temperatures up to 250 °C without damage.

The *delta shear* (DS) type has high sensitivity and a high resonance frequency, and is more immune to base strain and thermal effects than CM sensors. A DS piezoelectric accelerometer contains three piezoelectric elements, arranged in an equilateral triangle as shown in figure 9.5(*b*). Each piezoelectric transducer carries its own seismic mass. Acceleration along the axis of the sensor creates shear stresses in each of the PE elements. The use of shear rather than compression increases the stiffness of the PE elements, which raises the resonance frequency and makes the accelerometer more immune to external environmental changes than the CM type. However, the more complex construction means that DS piezoelectric accelerometers are usually more expensive than the simpler CM type.

If the seismic masses in a DS accelerometer are bonded to the piezoelectric transducers, the upper operating temperature is limited by the need to avoid softening the bonding adhesive. To avoid this problem the seismic masses are not usually bonded, but instead are clamped into position by a pre-loading ring as shown in figure 9.5(*b*). This raises the upper operating temperature to around 250 °C.

The *annular shear* (AS) type represents an attempt to combine the advantages of the DS accelerometer with the simplicity of manufacture and

consequent low cost of CM devices. In an AS accelerometer an annular mass is bonded to a piezoelectric ring, which is mounted on a cylindrical centre post as shown in figure 9.5(*c*). This approach lends itself to miniaturisation and enables very small, light accelerometers with very high resonance frequencies to be constructed. However, the bonding technique used to attach the mass to the piezoelectric element normally limits the maximum upper operating temperature range to around 100 °C.

Frequency response of PE accelerometers. The electrical output of a PE accelerometer as a function of frequency is shown in figure 9.6. This diagram is obtained by considering an accelerometer subjected to constant-amplitude sinusoidal acceleration, and plotting peak electrical output as a function of frequency. From this figure it can be seen that the useful bandwidth extends from a few hertz to about one-third of the resonance frequency. Thus PE accelerometers are used mainly for sensing frequencies in excess of 5–10 Hz. In automotive engineering they are frequently used for sensing engine vibrations.

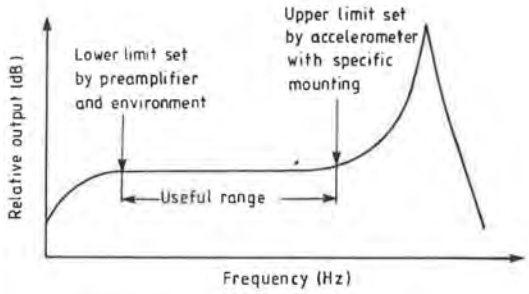

Figure 9.6 The electrical output of a PE accelerometer as a function of frequency.

The low-frequency limit obtained when using a PE accelerometer depends mainly on the electronic amplifier to which the sensor output is connected, rather than being a feature of the sensor itself. The low-frequency limit is determined by the *RC* time constant formed by the accelerometer output impedance and the preamplifier input impedance, and is usually less than 5 Hz. To obtain lower frequencies, preamplifiers with very high input impedances ($>10^9\ \Omega$) must be used.

Cross-axis (transverse) sensitivity of PE accelerometers. An ideal accelerometer should be sensitive to motion along a single axis, and should not produce an output when it is accelerated in any direction at 90° to this axis. Real accelerometers approach this ideal with varying degrees of

success. However, there is always some sensitivity to vibration perpendicular to the intended sensing axis. The size of the cross-axis sensitivity depends upon the exact orientation of the sensor with respect to the motion, but is usually within the range 0–5%. The dependence of transverse sensitivity on orientation is a consequence of the fact that the maximum charge and voltage sensitivity of the PE transducer elements is often not perfectly aligned with the axis of the accelerometer. This produces directions of maximum and minimum transverse sensitivity, which are at 90° to each other and perpendicular to the main axis. The calibration details supplied with an accelerometer always specify the maximum transverse sensitivity. The direction of minimum sensitivity is often marked with a spot on the accelerometer housing.

We have already seen that an accelerometer has a resonant response to vibration along its main axis. Similarly a resonance will occur if it is exposed to vibration in transverse directions. In the region of this resonance the transverse sensitivity may reach 100% of the main axis (off-resonance) sensitivity. For most PE accelerometers the transverse resonance occurs at about 30% of the main-axis resonance, and therefore lies within the bandwidth normally considered as the useful operating range of an accelerometer. Thus, spurious results may be generated if an accelerometer is subjected to transverse vibration at around one-third of the main resonance frequency, or if it is exposed to transverse shock loads which can contain broad-band energy extending to high frequencies.

PE accelerometers should not be dropped or knocked, since the transverse loads so produced may exceed the design limits and can cause irreparable damage.

9.2.2 Piezoresistive (PR) accelerometers

The piezoresistive effect occurs in silicon and other materials, and is used in the fabrication of miniature accelerometers. As the name implies, a change in electrical resistance occurs in response to changes in the applied stress. Piezoresistive vibration sensors are formed by placing stress-sensitive resistors on highly stressed parts of a suitable mechanical structure. The PR transducers are usually attached to cantilevers, or other beam configurations, and are connected in a Wheatstone bridge circuit. The beam may carry a seismic mass or may utilise its own self-weight. Under acceleration the beam deflects due to the inertial forces and undergoes stress changes. These stress variations are converted into an electrical output, which is proportional to acceleration, by the PR transducers.

Piezoresistive accelerometers are relatively easy to construct, provide a low-frequency response extending to DC, and work well over a relatively large temperature range (−50 to +150 °C). They are frequently used in

automotive applications such as active-suspension control, where the primary requirement is for good low-frequency performance. A further feature which makes them valuable in automotive engineering is their ability to include signal processing and communication functions within the sensor package at little extra cost.

The drawbacks of PR devices are that the output signal level is moderate (typically 100 mV full scale for a 10 V bridge excitation), the sensitivity can be temperature dependent, and the usable bandwidth is not as large as that which may be obtained from a PE sensor.

Before considering the design of PR accelerometers in detail, we need to understand the behaviour of piezoresistors.

Piezoresistors. If a rectilinear resistor has length l, width W, thickness t and a bulk resistivity ρ, its resistance R will be

$$R = \frac{\rho l}{wt}. \tag{9.9}$$

The gauge factor or strain sensitivity is defined as k, where

$$k = \frac{\mathrm{d}R/R}{\varepsilon}.$$

ε is the relative change in length of the resistor (the strain) due to a stress σ, applied to the substrate parallel to its length. Figure 9.7 shows the consequences of the applied stress. The length increases by an amount $\mathrm{d}l$, while the width and thickness decrease by $\mathrm{d}w$ and $\mathrm{d}t$ due to Poisson's ratio ν. It is clear that $\mathrm{d}w = -\nu w\varepsilon$ and $\mathrm{d}t = -\nu t\varepsilon$.

The original cross section was

$$A = wt.$$

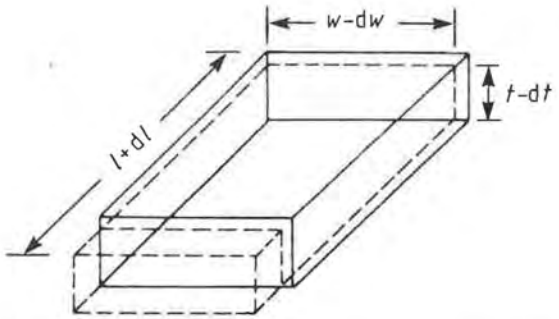

Figure 9.7 Consequences of applying stress to a piezoresistor. See text for details.

Owing to the strain ε, the new cross-sectional area is

$$\begin{aligned} A' &= (W - \mathrm{d}w)\,(t - \mathrm{d}t) \\ &= wt + 2\nu wt\varepsilon + \nu^2 wt\varepsilon. \end{aligned}$$

The term $\nu^2 wt\varepsilon$ is very small compared with the other two terms in the equation and can be neglected. We can therefore write the change in cross-sectional area as

$$A - A' = \mathrm{d}A = -2\nu\varepsilon A$$

giving

$$\frac{\mathrm{d}A}{A} = -2\nu\varepsilon.$$

Differentiating equation (9.9) gives

$$\frac{\mathrm{d}R}{R} = \frac{\mathrm{d}\rho}{\rho} + \frac{\mathrm{d}l}{l} - \frac{\mathrm{d}A}{A}$$

and hence the gauge factor k is

$$k = \frac{\mathrm{d}\rho/\rho}{\varepsilon} + (1 + 2\nu). \tag{9.10}$$

Typically ν will be between 0.2 and 0.3. Equation (9.10) therefore shows that the longitudinal gauge factor is a function of changes in both longitudinal resistivity and geometry. In conventional foil or wire strain gauges the piezoresistive effects are negligible, and the variations in resistance are mainly a function of dimensional changes. For a foil gauge k is approximately 2. For piezoresistive strain gauges the first term in equation (9.10) is significant, and higher gauge factors (typically around 10) can be achieved, giving enhanced sensitivity. It should be noted, however, that the resistivity of most PR materials is strongly temperature dependent, and that as a result PR strain gauges generally have a higher thermal sensitivity than other types.

Silicon piezoresistive accelerometers. The first silicon accelerometer was demonstrated in 1976 [1]. It consisted of a single cantilever carrying PR strain gauges near its root. The device was fragile and required the inclusion of a liquid-filled cell for damping. Improved designs have since appeared.

Three types of silicon PR accelerometer have since evolved, as shown in figure 9.8. These are the *single cantilever*, the *doubly supported* structure, and the '*top-hat*' design. It will be noted that, despite its name, the single cantilever design can have one, two or more supporting beams carrying the seismic mass. The distinguishing feature of this type is that the support beams are all placed along one edge. The doubly supported cantilever of figure 9.8(*b*) uses four supporting beams, and for this reason is sometimes

referred to as a quad cantilever. The top-hat approach is implemented when a device capable of undergoing large displacements is required. If the supports are folded as shown in figure 9.8(*c*) the effective length can be increased, while the package size is kept constant.

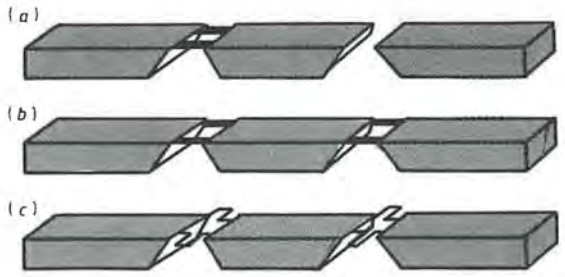

Figure 9.8 Three types of silicon PR accelerometer: (*a*) single cantilever; (*b*) doubly supported structure (double cantilever); (*c*) double cantilever with top-hat springs ('top-hat' design).

In all three types viscous damping is provided by the inclusion of a small volume of air.

The single cantilever type has the highest gain, but can also have a large transverse sensitivity. The doubly supported and top-hat designs provide good off-axis cancellation and are reasonably robust. They are consequently the most common configuration for PR silicon accelerometers.

(i) *Resonance frequency*. The resonance frequency of a silicon PR accelerometer is determined by the stiffness of the support structure and the seismic mass as discussed in section 9.1. Typical resonance frequencies for silicon PR accelerometers are in the range 500–5000 Hz. Thus the bandwidth is considerably lower than that of a typical PE device. For comparison, thick-film devices usually have resonance frequencies between 300 and 2000 Hz.

(ii) *Sensitivity*. The sensitivity increases with seismic mass m_s, decreases with support stiffness K, and is modulated by a transduction efficiency term b

$$\text{sensitivity} = b\,\frac{m_s}{K}.$$

The main parameters which determine b are the position of the PR transducers, the number of PR transducers and the transduction efficiency

of the transducers, which is a function of their geometry and chemical constitution.

The sensitivity is also inversely proportional to the square of the resonance frequency f_r

$$\text{sensitivity} = \frac{(2\pi)^2 b}{f_r^2}. \tag{9.11}$$

Equation (9.11) shows that if a sensitive accelerometer is required the resonance frequency should be as low as possible. Thus the sensitivity requirement is in conflict with the need for a large bandwidth. A typical doubly supported design with an output of 5 $mV\,g^{-1}$ per volt of bridge excitation will have a resonance around 500 Hz, implying that the useful bandwidth extends from DC to around 150 Hz. If the design is modified so that the resonance moves to 37 kHz, which is approaching the practical limit for this type of accelerometer, the sensitivity decreases to 5 $\mu V\,g^{-1}$ per volt of bridge excitation.

(iii) *Off-axis modes and transverse sensitivity.* A major advantage of the doubly supported type is a substantial reduction in transverse sensitivity and unwanted resonant modes compared with the single cantilever. The three principal modes of vibration for a doubly supported cantilever are shown in figure 9.9. The intended axis of sensitivity is vertical, and therefore only the mode shown in figure 9.9(*a*) is desired. However, for the modes shown in figures 9.9(*b*) and (*c*), which are excited by off-axis acceleration, opposite sides of the structure undergo opposite forms of bending. Careful positioning of the PR transducers and the use of a bridge circuit can therefore be used to give a high degree of immunity to outputs originating from these modes. No such symmetry exists in the case of a single cantilever. The maximum transverse sensitivity of a doubly supported PR silicon accelerometer is comparable with that of a PE device, and is typically 5% of the main axis sensitivity.

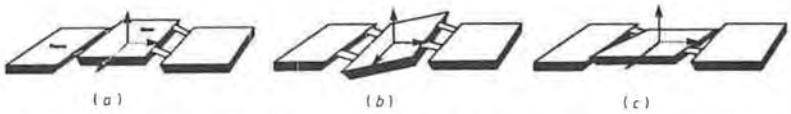

Figure 9.9 Principal modes of vibration for a doubly supported cantilever.

Thick-film piezoresistive accelerometers. Thick-film circuits are formed by the deposition of layers of special pastes (usually referred to as inks, although there is little resemblance to conventional ink) onto an insulating

substrate. The printed pattern is fired in a manner akin to the production of pottery, to produce electrical pathways of a controlled resistance. Parts of a thick-film circuit can be made sensitive to strain or temperature. The thick-film pattern can include mounting positions for the insertion of conventional silicon devices, in which case the assembly is known as a *thick-film hybrid*. The process is relatively cheap, especially if large numbers of devices are produced, and the use of hybrid construction allows the sensor housing to include sophisticated signal conditioning circuits. These factors indicate that thick-film technology is likely to play an increasingly important part in automotive sensor design.

Three main categories of thick-film inks exist: conductors, dielectrics (insulators) and resistors. Conductors are used for interconnections, such as the wiring of a bridge circuit. Dielectrics are used for coating conducting surfaces (such as steel) prior to laying down thick-film patterns, for constructing thick-film capacitors and for insulating cross-over points, where one conducting path traverses another. Resistor inks are the most interesting from the point of view of sensor design, since many thick-film materials are markedly piezoresistive.

The main constituents of a thick-film ink are the *binder* (a glass frit), the *vehicle* (an organic solvent), and the *active elements* (metallic alloys or oxides). After printing, each layer of a thick-film pattern is dried to remove the organic solvents (the vehicle) which give the ink its viscosity. Drying also improves the adhesion properties, bonding the ink to its substrate and rendering the pattern immune to smudging. This stage is usually performed in a conventional oven at 100 to 150 °C.

A final high-temperature firing is required to remove any remaining solvent and to sinter the binder and the active elements. During the firing cycle a thick-film pattern is raised to a temperature between 500 and 1000 °C. The glass frit melts, wets the substrate and forms a continuous matrix which holds the functional elements. The heating and cooling gradients, the peak temperature and the dwell time determine the *firing profile*. This has a critical effect on the production of a thick-film circuit, since it allows the electrical characteristics of the inks to be modified. Resistor materials are especially sensitive to the firing profile, and the resistor layer is usually therefore the last to be fired. However, the need for passivation of a circuit often necessitates covering it with a dielectric layer. To avoid changing the resistor values, a low-melting-point dielectric is often used for the final layer.

The need for high-temperature firing can cause problems if thick-film piezoresistors are to be applied to previously heat-treated components. The temperatures used can adversely affect, for example, the properties of toughened or hardened steels.

Thick-film circuits and sensors are created by screen printing. This is essentially a stencil process, in which the printing ink is forced through the

open areas of a mesh-reinforced screen onto the surface of a substrate. The screen stencils are formed by photolithography. In this process a photosensitive mesh-filling material is exposed to ultraviolet light through a mask depicting the required pattern. The image is photographically developed, and those parts of the pattern which have not been fixed are subsequently washed away.

(i) *Thick-film materials as primary sensors.* The use of thick-film technology was introduced as a means of miniaturising circuits without incurring the expense associated with fabrication in silicon. It was soon noted that thick-film materials had temperature- and stress-dependent properties. Although this was awkward from the point of view of circuit fabrication, it has since been turned to good account in sensor design. The linear temperature coefficient of resistance (TCR) possessed by certain platinum-containing conductive inks has allowed resistance thermometers to be constructed wholly in thick-film form [2]. More importantly from the point of view of accelerometer design, the PR properties of thick-film resistor (TFR) inks can be used to form strain sensors. This approach has been used to make a number of pressure sensors (see for example [3]), and by 1993 was being exploited to produce accelerometers.

9.2.3 Capacitive Accelerometers

The use of a capacitance change is well established in sensor design as a measuring technique. Pressure and displacement have frequently been evaluated by this means, and a number of manufacturers now supply capacitive accelerometers. These are claimed to give better performance than conventional types for measuring low-frequency, low-level acceleration.

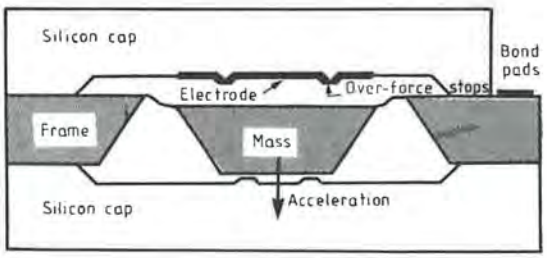

Figure 9.10 Typical design for a capacitive accelerometer.

Capacitive accelerometers are micromachined from single-crystal silicon. A typical design is shown in figure 9.10. A conducting layer is deposited onto one surface of a silicon block. A second conducting layer is laid down on one side of a second block, which acts as the seismic mass. The mass is supported by beams, and is separated from the base by an air gap as shown in figure 9.10. The two halves of the sensor are electrostatically bonded together. Signal processing electronics can be incorporated within the sensor package.

Capacitive accelerometers are fairly expensive to manufacture, and usually cost about the same as the PE equivalent. They are claimed to provide a frequency response down to DC, stable damping characteristics and a useful bandwidth which is larger than PR but smaller than that provided by PE devices. In addition their output is extremely stable over a wide temperature range. As shown by table 9.1, this is their primary advantage over the other, more common accelerometer types.

The output impedance of any capacitive sensor is intrinsically very high. However, if a capacitive accelerometer includes signal conditioning circuits within the package a low output impedance is usually provided, so that vibration signals may be transmitted over lengthy cables without loss or distortion.

To sum up therefore, the inherent advantages of capacitive sensing are its low power, high output, wide dynamic range and relative immunity to

Table 9.1 Comparison of different accelerometer types.

Parameter	Piezo-electric	Piezoresistive		Capacitive
		Silicon	Thick film	
DC response	No	Yes	Yes	Yes
Bandwidth	Wide	Moderate	Low	Wide
Self-generating	Yes	No	No	No
Impedance	High	Low	Low	Very high
Signal level	High	Low	Low	Moderate
Temperature range (°C)	−55–100[a]	−55–150	−50–120	−200–200
Linearity	Good	Moderate	Moderate	Excellent
Static calibration (turnover)	No	Yes	Yes	Yes
Cost	High	Low	Low	High
Ruggedness	Good	Moderate	Moderate	Good
Suitable for shock	Yes	No	No	No

[a]Unless special designs are used.

thermal effects. Unfortunately the signal processing requirements for this type of sensor are not straightforward, and it can be difficult to obtain a linear output.

9.2.4 Inertia Switches

Many motor manufacturers have considered the adoption of one or more passenger protection systems for use in the event of a crash. A number of such systems have been demonstrated and have appeared on production vehicles. The devices considered include airbags, collapsible steering columns, and ignition and fuel cut-off switches. All of these have one feature in common: they rely upon sensing the rapid deceleration associated with a crash for correct operation. The sensors used must discriminate between a crash and 'legitimate' manoeuvres, such as driving over kerbs, hitting potholes or heavy braking.

Crash sensing simply requires a determination of when a preset acceleration threshold is exceeded, rather than an accurate measurement. The deceleration associated with a crash is at least an order of magnitude greater than anything likely to be experienced in normal driving. It is not usually considered economic to use accelerometers for crash sensing, since a fairly crude estimation of acceleration is sufficient. For this reason inertia switches are often used instead. These consist of a weight, often in the form of a steel ball or cylinder, which is fixed to one end of a spring. The weight and spring are contained within a tube. When the assembly is accelerated or decelerated along its axis, inertia forces cause the weight to compress or extend the spring. The deflection of the weight is proportional to acceleration, and is sensed by placing an appropriate sensor in the side of the tube. An alternative design consists of a ferromagnetic sphere, retained at the bottom of a 'saucer' by a permanent magnet. Rapid deceleration causes the ball to break free of the magnet, and it rolls to the rim of the saucer where it closes a switch.

Reluctance probes and Hall effect sensors have also been used experimentally for designing acceleration switches. The sensor placing determines the acceleration level at which an output is given. Signal conditioning circuits are often included in an inertia switch to provide a digital output.

9.3 ENVIRONMENTAL EFFECTS

Accelerometers are frequently used to make measurements under severe environmental conditions. In automotive applications these can include both high and low temperatures (typically −40 to +150 °C), severe shock

loadings up to several hundred g, a wide range of humidities (up to 100%), exposure to electromagnetic radiation, and exposure to potentially damaging chemicals such as petrol, oil, hydraulic fluid and water. It is therefore important that all accelerometers used in automotive applications are sealed, are as rugged as possible and have a sensitivity to environmental changes which is as low as possible.

9.3.1 Thermal Sensitivity

Table 9.1 shows the temperature ranges over which the various types of accelerometer may be used with confidence. With the exception of piezoelectric sensors, all the accelerometers discussed in this chapter will behave satisfactorily within the typical automotive temperature range (−40 to +150 °C). Some forms of piezoelectric accelerometer (especially the bonded DS type) begin to give erroneous results above 150 °C, and should not be used at high temperatures unless the manufacturer states that this is permissible. However, in general they may be used in applications where they are not placed close to the engine or exhaust system.

In addition to the temperature limits shown in table 9.1, accelerometers can also exhibit a slowly varying output when subjected to temperature transients. This output arises from two causes: a pyroelectric effect and non-uniform thermal expansion of the accelerometer structure, which subjects the sensing element to a stress variation. The effect of a temperature transient is usually seen as a low-frequency electrical output, which can be removed by appropriate filtering. It is unlikely to interfere with measurements unless very low-frequency, low-amplitude vibrations are being investigated. This is not usually the case in automotive engineering.

9.3.2 Humidity

Most commercial accelerometers are of sealed construction, with either welded or epoxy bonded housings. They have a high resistance to the majority of corrosive agents encountered in industry. If it is necessary to immerse an accelerometer, or to use it in an environment where heavy condensation is likely, it may be necessary to take special precautions with cabling. Impervious Teflon cables can be obtained for this purpose. These should be sealed at the accelerometer/cable connection using a silicone rubber sealant such as Dow Corning Silastic 738RTV. This material is suitable for use from −70 to +260 °C.

9.3.3 Acoustic Sensitivity

All accelerometers are to some extent sensitive to acoustic excitation. However, with careful design unwanted acoustically generated outputs can be avoided. Most accelerometers are of rigid, mechanically isolated construction, and pressure variations in air will have little effect on the force transmitted to the sensing element. In general, vibrations of the surface to which the sensor is attached will give rise to much higher signals than will direct acoustical excitation of the accelerometer. Care must be taken when trying to measure very low-level accelerations in the presence of an intense sound field. However, this situation is unusual in automotive engineering, since most of the acceleration levels of interest are fairly large.

9.3.4 Base Strain Sensitivity

When an accelerometer is mounted on the surface of a structure undergoing strain variations, a spurious output will be generated by the accelerometer if any of the strain is transmitted to the sensing element. The base strain sensitivity is usually specified by the manufacturer for individual accelerometers, in units of acceleration (g or $m\,s^{-2}$) per microstrain (μ).

Commercial accelerometers usually have thick, stiff bases to reduce strain effects. A further reduction in base strain sensitivity is possible for PE accelerometers by using DS designs (see section 9.2.1).

9.3.5 Electromagnetic Interference (EMI)

Most commercial accelerometers are housed within welded steel or stainless steel enclosures. These normally provide adequate screening from the effects of electric or magnetic fields, which arise from spark ignition systems and other electrical wiring. However, EMI problems can arise in the cabling used to connect accelerometers to a measurement system. The problem is particularly common when PE sensors are used, since these devices have a very high impedance. If PE transducers are used in an automotive application it is desirable to employ the type which incorporates a charge amplifier (see section 9.5) within the sensor housing. The sensor output is then of low impedance, and is consequently much less likely to be affected by EMI.

9.4 MOUNTING TECHNIQUES

It is very important to mount an accelerometer properly. If it is not rigidly fixed to a vibrating surface the high-frequency performance will be poor. The mounting technique commonly recommended for commercial sensors uses a steel stud, and is shown in figure 9.11. If this technique is used it is essential that a good surface contact is obtained between the accelerometer and the test specimen. The mounting surface should therefore be smooth and flat, with the hole for the mounting stud perpendicular to the surface. A thin film of silicone grease can be applied to the base of the accelerometer before screwing it down, to improve the mounting stiffness and hence the high-frequency performance.

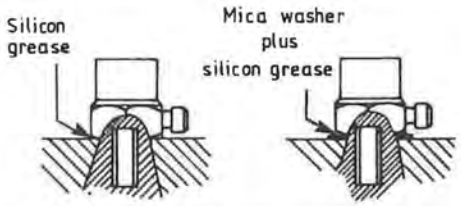

Figure 9.11 Recommended mounting technique for commercial sensors (courtesy of Bruel and Kjaer).

9.4.1 Connecting Cables

Piezoelectric and capacitive sensors are very high impedance devices. They are therefore susceptible to triboelectric noise, which can be generated by connecting cables when these are subjected to mechanical motion. Dynamic bending, compression or tension of cables momentarily separates the cable screen from the dielectric, and this results in local capacitance changes and the release of triboelectric charge as shown in figure 9.12. The problem is especially severe at low frequencies, and when using conventional coaxial cable rather than specially treated accelerometer leads. To minimise this source of electrical noise accelerometer cables having special noise reduction treatment should be used. These should be clamped to the vibrating specimen using epoxy glue, wax or adhesive tape so that relative movement is avoided as far as possible.

In some circumstances it may not be possible to clamp the cable to the vibrating surface. For example, it may be too hot. In this case an accelerometer with a top-mounted output connector should be used, and the cable should be led away from the accelerometer and clamped as soon as possible.

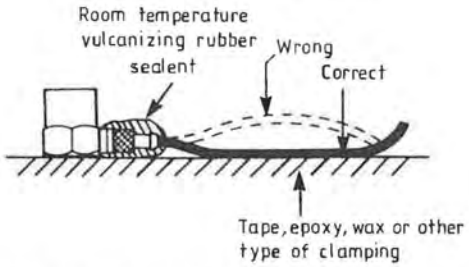

Figure 9.12 Correct clamping of connecting cables (courtesy of Bruel and Kjaer).

9.5 PREAMPLIFIERS

In general it is difficult to separate the characteristics of a sensor from those of its associated preamplifier. Thus, it is fitting to conclude this chapter with a brief discussion of the preamplifiers used in conjunction with various accelerometers.

The output signal generated by a PE accelerometer appears across an extremely high impedance, and is of very low power. Piezoelectric sensors act as charge sources, and the preamplifiers used with this type of accelerometer are known as *charge amplifiers*. The usual arrangement is shown in figure 9.13, where it is shown that a charge amplifier consists of an operational amplifier (or op-amp) with capacitive feedback. The circuit effectively converts a charge to a voltage and provides amplification. The input capacitance $C_p = C_F(G - 1)$, where G is the open-loop DC gain of the op-amp. (For more details of op-amps see [4].) The amplifier output voltage is given by

$$V_{out} = \frac{QG}{C_S + C_C - C_F(G-1)}.$$

The cable capacitance C_C, source capacitance C_S and op-amp input capacitance C_P are usually much smaller than C_F. Since the open-loop gain of an op-amp is very high, the output becomes

$$V_{out} = -Q/C_F.$$

With C_F constant the output voltage V_{out} is proportional to the input charge Q. Only when the cable becomes so long that the value of C_C approaches that of C_F will the sensitivity of the circuit be affected.

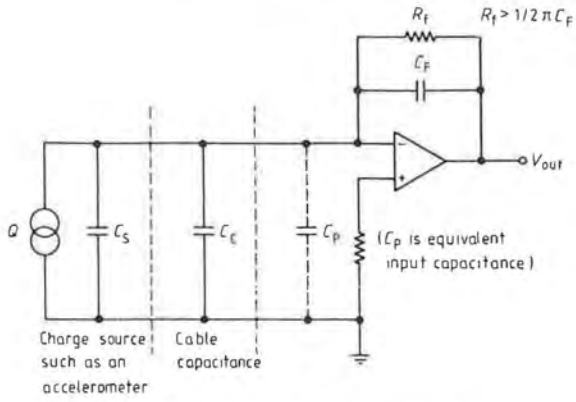

Figure 9.13 A charge amplifier circuit.

Piezoresistive sensors are usually connected as part of Wheatstone bridge circuit. The bridge may contain one, two or four active sensors, and can provide first-order temperature compensation [4]. A bridge circuit is essentially a pair of parallel potential dividers, as shown in figure 9.14, and an excitation voltage must be applied before any output signal is obtained. Changes in the sensor resistance produce changes in the out-of-balance voltage. This can be amplified by a conventional high-impedance voltage amplifier, usually an op-amp. Further details of bridge circuits and suitable amplifiers are given in [4].

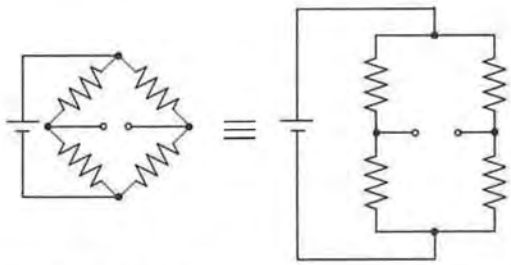

Figure 9.14 Similarity of Wheatstone bridge and a pair of potential dividers.

A signal may be produced from a capacitive sensor in three ways. The most common method is to include one or more capacitive sensors in a bridge circuit such as the Blumlein bridge [5]. An alternative approach is

to include a capacitive sensor in an oscillator circuit as discussed in Chapter 3, in such a way that it determines the oscillator frequency. Finally, if a DC voltage is connected in parallel with a capacitive sensor a charge amplifier may be used. As the measurand varies the capacitance will change, and thus a charge variation occurs, which can be converted to a voltage and amplified by a charge amplifier such as that shown in figure 9.13.

9.6 SUMMARY

In conclusion, it can be seen that care must be taken to select the correct accelerometer for each measurement task. If a large bandwidth is required, piezoelectric sensors are probably best. If a piezoelectric accelerometer is required to work at high temperatures it will probably have to be of the centre-mounted compression type. If low-frequency measurements are required, piezoresistive or capacitive devices should be used. If temperature extremes are likely, capacitive accelerometers may be best, and so on. Table 9.1 gives a comparison of the accelerometer types discussed in this chapter, and should be used to aid sensor selection.

REFERENCES

[1] Roylance L M and Angell J B 1979 A batch-fabricated silicon accelerometer *IEEE Trans. Electron Devices* **26** 1911

[2] Reynolds Q M and Norton M G 1985 Thick-film platinum temperature sensors *Proc. Test and Transducer Conf. (London, 1985)* Vol 2 (Trident International Exhibitions) pp 31–44

[3] Morten B, Prudenziati M and Taroni A 1982 *Proc. Int. Kongress Mikroelektronik* (International Society for Hybrid Microelectronics) pp 345–54

[4] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 44551 1

[5] Jones B E 1978 *Instrumentation, Measurement and Feedback* (New York: McGraw Hill) ISBN 0 07 099383 1

10

Gas and Fuel Composition Sensors

The problem of urban air pollution by motor vehicles has existed for many years. Before the days of vehicle emission controls, the use of diesel-powered vehicles was significantly inhibited by their reputation for smoky, evil-smelling exhaust emissions, although in fact these diesel emissions were less toxic than the normally invisible emissions from gasoline-powered vehicles. Since the 1970s, however, the issue of air pollution from both gasoline- and diesel-powered vehicles in urban areas has become of major importance to the automotive industry as well as becoming a significant factor in worldwide atmospheric pollution which could result in eventual global warming — the 'greenhouse' effect.

This problem was first recognised in the 1960s in the Los Angeles area of California, where the particular temperature inversion conditions which prevail there, combined with the very large number of vehicles in use, had led to severe air pollution. This resulted in the application in 1966 of the first control regulations to restrict the exhaust emissions of gasoline-powered vehicles used in California. Initially only carbon monoxide and hydrocarbons were controlled, but in due course the contribution of the oxides of nitrogen to acid rain and ozone generation was recognised and the emission regulations were extended to cover these as well.

During the 1970s emission control standards were extended to the remainder of the USA, while their increasing severity made the accurate control of fuelling and ignition timing and the introduction of exhaust gas recirculation essential and forced the introduction of catalysts to further reduce the regulated exhaust gas pollutants. The catalysts require the use of unleaded fuel as their catalysing surfaces are rapidly 'poisoned' and made ineffective by lead deposits if leaded fuel is used. What was not considered important at the time was that one of the gases to which carbon monoxide and hydrocarbons were converted in the catalyst was carbon dioxide, considered at that time to be totally innocuous.

In Europe, emission regulations were introduced in the early 1970s, followed in the 1980s by the implementation of the '1973 US' non-catalyst standards in Switzerland and Sweden. Since then, later US regulations,

which cannot be met without the use of catalysts, were implemented in Europe as an alternative to the unique European emission regulations then existing, and it is these regulations which have been mandatory from the beginning of 1993 onwards.

In Europe, during the period between the late 1970s and late 1980s, it was seen to be possible to meet what were then expected to be the unique European regulations by a technique known as 'lean burn', where the engine is operated with more air and less fuel. This differs from the condition required to operate a catalyst effectively, since this requires the air–fuel ratio to be at a richer level known as 'stoichiometry' (14.7 parts of air to one of fuel for gasoline). In the case of lean-burn operation, fuel economy is improved by up to 10% and good control of the oxides of nitrogen can be achieved, but it is not possible by this method to meet the very low emission levels now required in the United States and throughout the European Community. Most of the work done on lean burn engines in the 1980s, which had resulted in the introduction of a range of production engines, has therefore been abandoned for the next generation of stoichiometric three-way catalyst engines.

All is not lost, however, as far as lean burn operation, with all its economy and emission advantages, is concerned, as Toyota have recently published [1] details of their third-generation lean burn system which extends the capability already shown by their second-generation system to operate under lean burn conditions at light loads, while changing the engine over rapidly to stoichiometric operation with a catalyst when high loads are applied.

Both stoichiometric and lean burn operation require the use of exhaust gas oxygen (EGO) sensors if the engine is to be controlled accurately enough to attain the low emission levels permitted by the regulations. Therefore the accuracy required of these sensors needs to be high; this must be combined with an acceptable cost to make their use on every vehicle practical. In section 10.1 we will describe their principles, construction and method of use in some detail. Section 10.3 will describe possible future sensors for individual exhaust gases.

Another method of improving engine emissions is by the use of alternative fuels, in particular ethanol or methanol. Both are alcohol-based fuels, ethanol being derived from vegetable matter and methanol from a mineral base. They are significantly more expensive than petrol or diesel fuel, they have cold start and mixing problems due to the higher latent heat of evaporation and also require twice the volume of fuel compared with gasoline for a similar range and performance. For these reasons they are generally used mixed with a small percentage of gasoline; for example M85 is a fuel with 85% methanol and 15% gasoline.

To effectively control engines using these mixtures it is essential to provide information to the control system of the ratio between the

alternative fuel concerned and the gasoline in the fuel tank as the percentage of each constituent fuel changes during use and refuelling. Sensors to do this have been developed and will be described in section 10.4.

10.1 EXHAUST GAS OXYGEN (EGO) SENSORS

The EGO sensor is essential to allow effective feedback control of the air–fuel ratio in a gasoline engine which is intended to operate at stoichiometry with a 'three-way' catalyst. Operation in this mode is necessary if the critical pollutants of carbon monoxide (CO), hydrocarbons (HC) and the oxides of nitrogen (NO_x) are to be effectively controlled, as can be seen from figure 10.1, which shows the effect of using a three-way catalyst on these emissions.

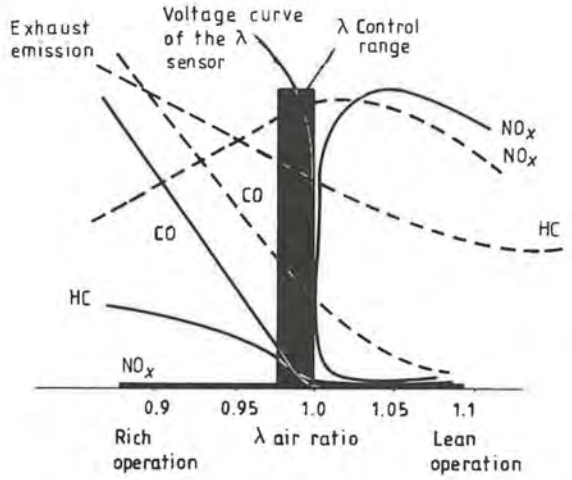

Figure 10.1 Effect of a three-way catalyst on emissions of HC, CO and NO_x above and below stoichiometry ($\lambda = 1$), showing emission levels without catalyst (broken curves) and with catalyst (full curves). Also shown is the voltage curve of the EGO (λ) sensor as the air–fuel ratio passes through stoichiometry.

As can be seen from the figure, oxidation of HC and CO, and maximum reduction of NO_x only occurs very close to the stoichiometric air–fuel ratio (14.7 parts of air to one of fuel for gasoline, shown on the diagram as $\lambda = 1$, where λ = actual air–fuel ratio/stoichiometric air–fuel ratio). The

EGO sensor, by measuring the relative partial pressure of oxygen in the exhaust gas compared with that in the ambient air, gives an accurate indication of the change of air–fuel ratio at stoichiometry.

Figure 10.1 shows the rapid change in the response curve of the EGO sensor as the air–fuel ratio passes through stoichiometry. This enables the sensor to respond to very small changes of the air–fuel ratio around stoichiometry and therefore to provide good feedback control. However, as can be seen, there is very little change in EGO sensor output in the lean burn region, so that it has been necessary to develop special lean burn sensors to provide feedback under those operating conditions; these sensors will be described later.

10.1.1 Zirconia Oxygen Sensor

Figure 10.2 shows the construction of the most widely used EGO sensor. The sensor is shaped rather like a spark plug and is mounted in a hold in the exhaust manifold on the engine side of the catalyst. The sensing end comprises a zirconia (ZnO_2) ceramic thimble with noble metal catalytic platinum surface electrodes which project into the exhaust gas stream, with the inner surface of the thimble exposed to ambient air. The ceramic is conductive for oxygen ions, and if there is a pressure difference between the partial pressure of oxygen in the exhaust gas and in the ambient air then oxygen ions will migrate and generate a voltage between the platinum electrodes on the inner and outer sides of the zirconia thimble; this voltage is a logarithmic function of the ratio of the oxygen partial pressures on the two sides.

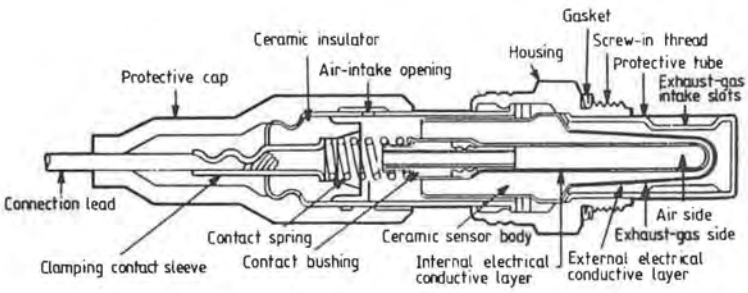

Figure 10.2 Construction of a zirconia ceramic EGO sensor.

The potential difference generated is described by the Nernst equation

$$E = \frac{KT}{2e} \ln \left(P(O_2) \text{ side } 1 / P(O_2) \text{ side } 2\right)$$

where K is the Boltzmann constant, T is the temperature in kelvin, e is the electronic charge and $P(O_2)$ is the partial pressure of oxygen. For air–fuel ratios below stoichiometry the exhaust gas is low in oxygen and the partial pressure is very low, whereas above stoichiometry the oxygen partial pressure increases rapidly by at least an order of magnitude. This, combined with the catalytic action of the platinum electrodes, causes the voltage generated to drop rapidly, so that for an air–fuel ratio change of 0.2 to 0.3 the output voltage changes from a maximum of about 900 mV to a minimum low level of about 50 mV as the air–fuel ratio passes through the 14.7 : 1 stoichiometric level from rich to lean.

Using the sensor as a sensitive switch to indicate the transition from rich to lean makes it possible to control air–fuel ratio to within 0.1 of stoichiometry.

The response rate of the EGO sensor is, however, significantly slower than would be ideal, although recent improvements have reduced it to about 50 ms for a step change in air–fuel ratio from the ⩾200 ms of earlier devices.

The EGO sensor works effectively over the normal range of exhaust gas temperatures (350–800 °C). However, below 350 °C, the sensor impedance becomes very high which results in difficulty in maintaining the output voltage. The response time also becomes much slower than in a hot sensor operating at normal exhaust gas temperatures. This problem, which leads to inadequate control signals being generated in a cold engine and under some slow-running and idle conditions, has resulted in the development of a heated version of the EGO sensor (a HEGO sensor).

In the HEGO a rod-shaped heater element is inserted into the interior of the zirconia thimble and a high and controlled current applied at engine switch-on to bring the sensor up to at least 350 °C as rapidly as possible. The heater can then be turned off when the exhaust gas temperature is consistently above 350 °C, but can be turned on again if necessary under conditions of low engine load. This facility makes it possible to match the sensor response more closely to heated catalyst warm-up times and control emissions more effectively earlier in the engine cycle.

It has been reported by Shulman and Hamburg [2] that the air–fuel set point of a zirconia oxygen sensor in a feedback system can shift lean as engine load increases and as cylinder-to-cylinder air–fuel distribution is worsening. The actual air–fuel shifts are, however, small and only affect high-precision systems.

Although the zirconia EGO sensor is by far the most widely used in production vehicles, other sensors have been investigated as an alternative.

10.1.2 Titania Oxygen Sensor

The titania EGO sensor relies on the fact that the electrical resistance of the semiconductor titania (TiO_2), changes in an oxidation–reduction reaction upon exposure to oxygen. This change in conductivity happens as a result of a deficiency of oxygen ions in the titania crystal lattice which is dependent on the temperature and partial pressure of oxygen in the gas surrounding the semiconductor. Under lean air–fuel ratios the deficiency of oxygen ions is small, while under rich air–fuel ratios the deficiency is large. Since the resultant lattice oxygen defects donate free electrons into the conduction band, the resistance of the semiconductor is high at lean air–fuel ratios and low at rich air–fuel ratios.

The resistance of the titania semiconductor (R_t) has been quoted by Pfeifer and Wertheimer [3] as being proportionally related to activation energy and temperature as

$$R_t \propto (P(O_2)\text{exhaust})^n \exp(E/KT)$$

where E is the activation energy of titania and n is approximately equal to 4 divided by the operating temperature. A change of three orders of magnitude in the resistance between rich and lean air–fuel ratios at the operating temperature of 800 °C has been reported.

It is possible to simulate the switching characteristics of the zirconia sensor by connecting the titania element in a voltage divider circuit with a

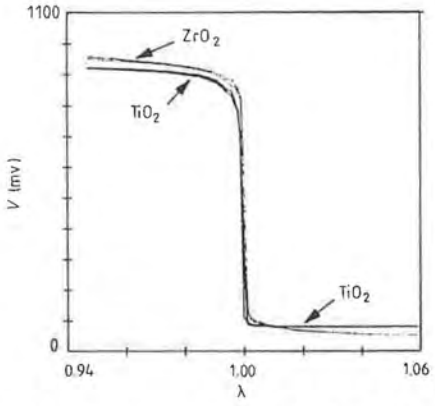

Figure 10.3 Voltage swing on a zirconia sensor as the air–fuel ratio passes through stoichiometry, compared with the output from a matched voltage divider containing a resistive titania EGO sensor (from Pfeifer and Wertheimer [3]).

compensating resistor selected to be midway on a logarithmic scale between the rich and lean resistance values of the titania sensor element. The voltage drop across the compensating resistor then goes from near zero under lean conditions to close to full value when the exhaust gas is rich, thus matching effectively the switching characteristics of the zirconia sensor. The voltage swing can be determined by the voltage applied to the divider and can be made similar to that of the zirconia sensor, so making the two devices interchangeable (figure 10.3).

Because of the temperature dependence of the titania sensor element it is necessary to compensate for this either by the use of a negative temperature coefficient resistor mounted in the exhaust gas stream as the compensating resistor in the voltage divider, or by the use of a heater to maintain the titania element near the highest temperature it will see in the exhaust environment. This is most easily achieved by depositing the titania sensor element as a thick film on one side of an alumina substrate and applying a thick-film heater element to the other side (figure 10.4). This arrangement permits rapid heating of the sensor

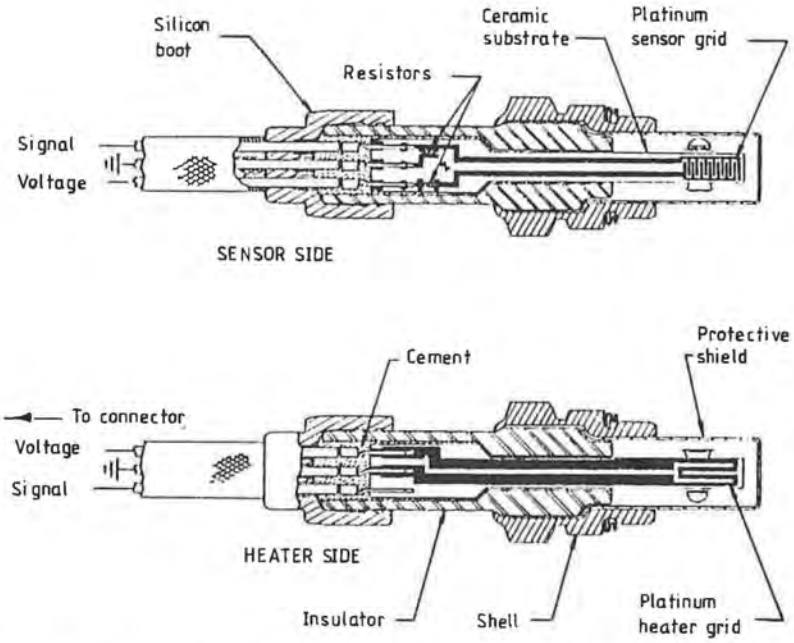

Figure 10.4 Construction of a heated titania EGO sensor (from Pfeifer and Wertheimer [3]).

element to near maximum operating temperature and its subsequent stabilisation to within 100 °C of this level. Under these conditions switching times comparable with, and in some cases rather faster than, the older type of zirconia sensor are achievable. However, recent improvements in zirconia sensor switching times have removed this advantage.

Although the titania sensor has advantages in simplicity of fabrication, compactness and low cost, it has not replaced the zirconia sensor because there is still a question over the level of reliability which can be obtained.

Other materials such as cobalt and niobium dioxides, strontium, lanthanum and tin have also been investigated as possible oxygen sensors. They all work by exploiting the change in electrical resistance through oxidation or reduction reactions with oxygen in a way analogous to titania, but no developments have so far been reported which challenge the dominance of the zirconia-based devices.

Both EGO and HEGO sensors are intended to provide feedback control of the engine operation at stoichiometry to permit effective operation of the three-way catalyst. However, as has already been indicated, the Japanese have shown that significant advantages in fuel consumption and in the emission of some of the undesirable exhaust gasses can be obtained by operation of the engine for at least a part of its operating time under lean burn conditions. Then feedback control of the engine using a lean burn EGO sensor is essential. The next section will describe sensors suitable for lean burn operation.

10.2 LEAN BURN OXYGEN SENSORS

If engines are to be effectively operated to give best emissions consistent with good economy and adequate performance under lean burn conditions, then full feedback control of ignition timing, fuelling and probably other functions such as exhaust gas recirculation (EGR), valve opening, swirl and induction manifold length is essential. This is even more critical for the type of mixed lean/stoichiometric cycle engine described by Katoh *et al* [1] since changeover from lean to stoichiometric operation is a critical function of the level of exhaust emissions under rapidly changing operating conditions. This critical feedback control makes an accurate lean burn exhaust gas oxygen (LEGO) sensor essential, and much work has been done to develop sensors which will have a good response from stoichiometry up to an air–fuel ratio of at least 25 : 1 and ideally 30 : 1.

The method developed to make it possible to operate EGO sensors over the lean burn operating range is known as 'oxygen pumping'. With this method a constant voltage is applied between the platinum electrodes of the EGO sensor thimble from a controlled source (this is instead of using the relative partial pressure of oxygen between the exhaust gases and the ambient air to control the migration of oxygen ions across the zirconia ceramic of the EGO sensor thimble, so generating a voltage between the platinum electrodes on each side). Oxygen is then pumped through the zirconia ceramic from one side to the other, with the current required being a direct measure of the number of oxygen ions pumped.

If a cell construction of the type shown in figure 10.5 is used, in which exhaust gas can diffuse through a small aperture into a cell with a small enclosed volume, then the zirconia oxygen pump just described can be arranged to remove oxygen from the cell quickly enough to reduce the overall concentration in the cell to the stoichiometric level. This condition can then be sensed by a zirconia stoichiometric EGO sensor which is arranged as a detection cell to compare the relative partial pressure of oxygen in the pumped cell with that in ambient air as described in section 10.1, and to give a rapid voltage change when this balance is reached. The current to drive the pumping cell and maintain a stoichiometric oxygen concentration in the comparator cell is then a measure of the lean air–fuel ratio in the exhaust gas.

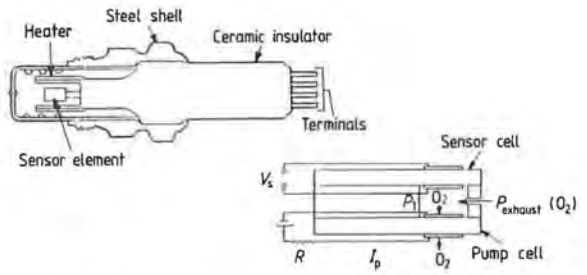

Figure 10.5 Zirconia heated lean burn EGO sensor showing diffusion aperture pump cell and sensor cell (right) and heater element (left).

Alternative lean burn cell configurations based on the same general principles have been described by various workers and are now briefly described.

Hetrick *et al* [4] describe an incremental titration design in which, instead of continuously removing oxygen, the oxygen pump is run in an oscillatory mode giving small deviations in oxygen concentration above and below stoichiometry in the comparator cell. The detection cell then

senses those deviations and the oscillation period required to produce a specified peak-to-peak output is a measure of exhaust oxygen level.

Haaland [5] describes another alternative in which the comparator is sealed and the oxygen pump is run until the detection cell senses a stoichiometric level of oxygen in the closed volume, then oxygen is added by the oxygen pump until the detection cell senses equilibrium between the oxygen in the closed volume and the exhaust gas. The total integrated pump current required to change the oxygen concentration in the enclosed volume from the stoichiometric oxygen level to the EGO level is then a direct measure of EGO.

The continuously pumped lean burn oxygen sensor has been most widely used, and a microelectronic version has been described by Velasco *et al* [6]. Figure 10.6, taken from their paper, shows the construction. Test results on continuously pumped lean burn oxygen sensors show that the pump current exhibits a linear relationship to air–fuel ratio from stoichiometry to at least 20 : 1, this lean limit being determined by the maximum oxygen diffusion rate of the pump cell.

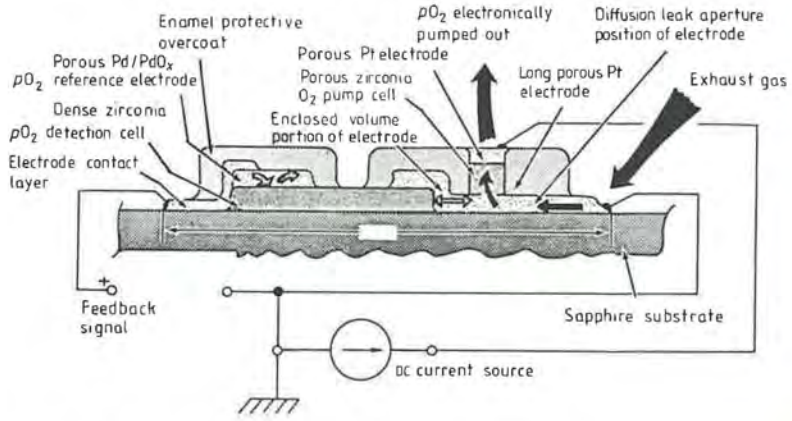

Figure 10.6 Zirconia lean burn EGO sensor constructed using microelectronic technology (from Velasco *et al* [6]).

In closed-loop feedback operation the sensor can be set with a constant pump current so that a switch signal is received from the detection cell at a specific air–fuel ratio. Variations in the pump current can then be used to select specific air–fuel ratios for engine operation. This is potentially of particular use for the control of engines operating on a mixed lean/stoichiometric cycle.

Another simplified lean burn sensor has been described by Kamo *et al* [7] and also by Katoh *et al* [1] for use with mixed lean/stoichiometric cycle engines. In this (see figure 10.7(*a*)) the oxygen supply available to the zirconia pumping cell is controlled by a diffusion layer which ensures that the number of oxygen ions available at the first electrode of the pumping cell is a direct function of the oxygen concentration in the exhaust gas. The current required to drive the pumping cell to its diffusion limit is then proportional to oxygen concentration.

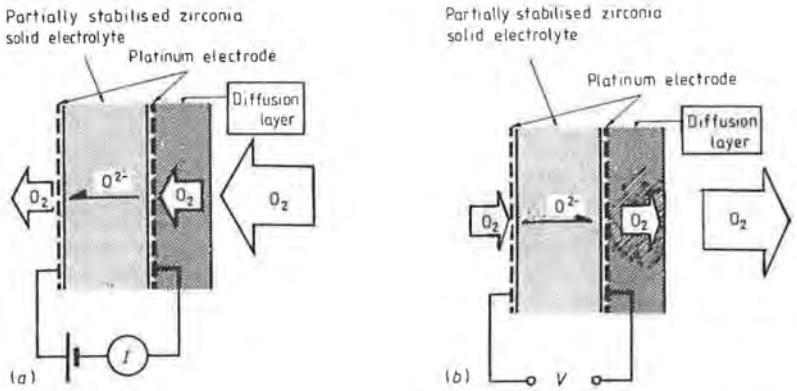

Figure 10.7 (*a*) Lean burn EGO sensor using a voltage-driven zirconia element as an oxygen pumping cell controlled by a diffusion layer (from Kamo *et al* [7]). (*b*) Stoichiometric EGO sensor using an open circuit zirconia element as an oxygen ion flow voltage generator (from Katoh *et al* [1]).

The sensor will also work as a stoichiometric device if it is operated under open circuit conditions (see figure 10.7(*b*)). Under these conditions a voltage is generated by oxygen ion flow in the opposite direction through the cell immediately the engine air–fuel ratio drops below the 14.7 : 1 stoichiometric level and the exhaust gas becomes low in oxygen. Figure 10.8 shows the output characteristics of this combined lean and stoichiometric sensor. Sensor heating is provided to bring the sensor up to its optimum operating point as rapidly as possible.

The construction of a practical version of this sensor is described by Kamo *et al* [7] and is shown in figure 10.9. The sensor element is a thimble of zirconia with platinum electrodes on the inner and outer surfaces. The diffusion layer over the outer electrode is made by the carefully controlled

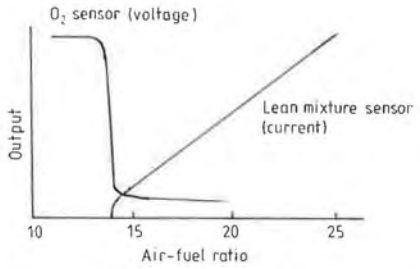

Figure 10.8 Output characteristics of the lean burn and stoichiometric sensors shown in figure 10.7 (from Katoh *et al* [1]).

plasma spray coating of a suitable ceramic. A cylindrical ceramic heater keeps the sensor element above 650 °C and the thimble has a double protection cover to reduce the effect of sudden temperature changes on the exhaust gas.

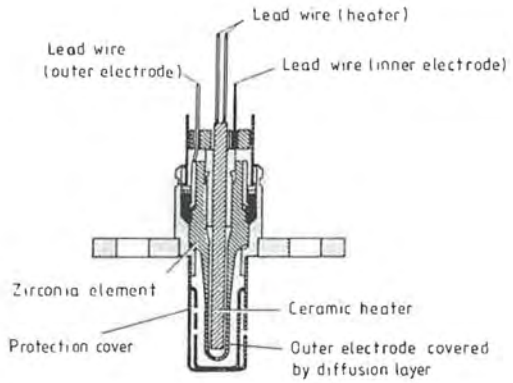

Figure 10.9 Construction of diffusion-controlled lean burn EGO sensor (from Kamo *et al* [7]).

In practical use with the Toyota lean burn engine the operation is confined to stoichiometric or richer or alternatively lean burn above a 20 : 1 air–fuel ratio. This avoids the operation of the engine between air–fuel ratios 15 : 1 and 18 : 1 where NO_x emissions are high. The sensor can be switched by the control system between stoichiometric and lean burn operation to accommodate these sharply different conditions.

One final method that can, under certain limited conditions, be used to monitor lean burn operations, is the use of a conventional heated zirconia exhaust gas oxygen (HEGO) sensor maintained at a temperature above

600 °C. Wiedenmann *et al* [8] describe experiments done with such a sensor which has had its heater power increased to 18 W and has been enclosed in a special protective tube to damp out rapid variations in exhaust gas pressure and temperature.

Steady-state tests for static output of a heated sensor for lean burn exhaust gas at 350 °C under steady load engine operation showed an exponential reduction in sensor output voltage from 60 mV at an air–fuel ratio of 15 : 1 to 20 mV at an air–fuel ratio of 19 : 1 (see figure 10.10).

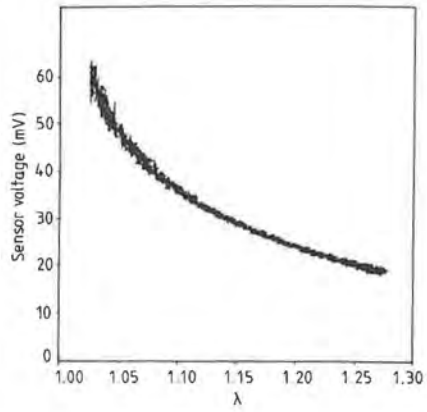

Figure 10.10 Output voltage of stoichiometric heated EGO sensor operated under lean burn conditions (from Wiedenmann *et al* [8]).

There is a small rise in sensor voltage with increase in exhaust gas temperature amounting to about 0.2 air–fuel ratios for an increase in temperature from 350 °C to 650 °C. This is about four times less than would be expected from the Nernst equation and was achieved by a special choice of materials and design of the sensor. Temperature correction can be applied if necessary in a practical application. Response times of between 50 and 150 ms are reported, depending on engine speed for step changes in air–fuel ratio.

The reducing rate of variation in output voltage with increase in air–fuel ratio makes it possible that the device would not be sufficiently accurate as a lean burn sensor for air–fuel ratios greater than 18 : 1.

In summary, therefore, the increasing availability of accurate lean burn sensors is likely to make possible more sophisticated engine control patterns in which optimisation of both emissions and driveability can be obtained at much better economy than is obtainable today with conventional stoichiometric catalyst-equipped engine operation. In the

longer term there is likely to be a need to measure the amount of each specific regulated gas in the exhaust rather than inferring air–fuel ratio from EGO measurements. To do this requires low-cost methods of measuring CO, HC and NO_x separately, as well as any other gas such as CO_2. Sensors to do this are being investigated, but are still a long way from the stage when they could be considered for use in engine development.

10.3 SELECTIVE GAS SENSORS

To obtain optimised control of an internal combustion engine it would be highly desirable to be able to measure the individual gases which contribute to exhaust emissions. In particular, measurement of CO, HC and NO_x, with the possible addition of CO_2, would be required if feedback techniques were to be used to optimise the emissions of a mixed-cycle (lean burn/stoichiometric) operating engine.

Many techniques for gas sensing exist; for a thorough review references [9] and [10] should be consulted. Moseley *et al* [10] list a number of techniques, including the galvanic oxygen sensor which has already been described in sections 10.1 and 10.2, the catalytic gas detector, the semiconducting oxide gas sensor, field effect transistor-based gas sensors, surface acoustic wave devices, as well as infrared spectroscopic and fibre optic sensors. For application in a mass-produced vehicle, low cost and small size, combined with fast response, high-temperature operation and good selectivity for the gas to be detected, are essential. None of the techniques, except the galvanic oxygen sensor, meet all these criteria, and as far as CO, HC and NO_x detection are concerned, there appear to be only two candidate methods which, if suitably developed, could meet the majority of the criteria: semiconductor gas sensors and infrared gas sensors.

10.3.1 Semiconductor Gas Sensors

Semiconductor gas sensors use semiconducting oxides which change their resistance when exposed to air containing any combination of a wide range of gases.

Tin oxide has been the most widely used material, and there has been some limited development of sensors using other oxides. Unfortunately all these oxides, although extremely sensitive to relatively low amounts of reactive gases in air, lack selectivity between different gases. They also

suffer from relatively low response rates — of the order of minutes for a step change of gas input. The presence of water vapour is a complicating factor, as it can cause considerable changes to the response of some oxides.

In principle the semiconducting gas sensor adsorbs oxygen initially at the surface of the oxide, and electrons are abstracted from the bulk material to form surface oxygen ions. This process forms a surface charge which, in the case of n-type oxides, inhibits the adsorption of further oxygen and leads to rapid saturation and slow response to future changes in gas concentration. In the presence of a reactive gas, the surface coverage of oxygen ions would decrease, with a consequent decrease in n-type semiconductor resistance.

In the case of p-type oxides, a decrease in the coverage of oxygen ions would lead to an increase in semiconductor resistance, and since in this case the surface coverage of oxygen ions is not limited by the supply of electrons, saturation takes place at a much higher level.

Most oxides show a response to a range of gases, but this is quite variable with significant selectivity in some cases.

Sensitivity is significantly affected by temperature, with a tin oxide sensor immersed in oxygen with 100 PPM of CO reported as changing percentage conductance from less than 10% at 200 °C to a peak of 40% at 400 °C, falling again to under 20% at 600 °C [10]. This peak occurs at different temperatures for different gases, but it would require some method of temperature cycling to define the peaks and therefore the gases concerned, which does not seem to be practical in a low-cost sensor.

A more likely way of using such sensors would be to have an array of sensors using different oxides with some 'smart' electronics to analyse the changes in resistance with temperature, with the aim of correlating the signals to produce an identifiable signal for each gas of interest.

In the case of automotive exhaust emissions CO, HC and NO_x could probably be measured by these multiple sensor devices, but the exhaust gases would have to be allowed to cool from the 800 °C or so at which they leave the engine to somewhere around 400 °C before the measurements could be made. Response time would continue to be a major problem and much work would need to be done to reduce this, at least to below 100 ms, before practical use could be considered.

10.3.2 Infrared Gas Sensors

Another type of gas sensor which seems to show some potential for the rapid analysis of exhaust gases is the infrared gas analyser. Currently available only as a large rack-mounted instrument for analysis in the emissions laboratory, progress has been made in making these instruments

more portable for mobile use, and infrared sensing systems have been developed for measuring exhaust gas emissions of passing vehicles from the roadside [11].

Moseley *et al* [10] quote five methods of infrared spectral analysis:

(i) fingerprint identification;
(ii) correlation spectroscopy;
(iii) high-resolution spectral measurements;
(iv) interferometric techniques;
(v) acoustic-optic filters.

All these systems rely on the identification of the selective wavelengths of absorption of infrared radiation identified with each gas of interest. For emission gases these are: CO, 4.6 μm; NO, 5.3 μm; and HC, 2.3 or 3.3 μm.

Of the methods quoted only correlation spectroscopy seems likely to offer a relatively low-cost rapid analysis technique. In this technique an infrared beam is passed through the gas to be measured and then alternately through two similar cells, one containing a known sample of the gas to be measured and the other a gas which is non-absorbing at the infrared wavelength of interest. The remaining infrared beam is then collected by a wide-band infrared sensor. The resultant signal is differentially amplitude modulated depending on which cell it passes through, and the depth of this modulation is a function of the amount of absorption caused by the gas being measured and therefore of its concentration.

There appears to be a possibility of developing a low-cost gas sensor using this technique if semiconductor infrared emitters and sensors are used combined with a dual-path optical system using optical fibre techniques, or extending the idea further with a multipath system using samples of CO, HC and NO to give absorption cells for all three emission exhaust gases.

10.4 FUEL COMPOSITION SENSORS

One method of improving engine emissions which has been extensively investigated is by the use of alternative fuels, such as methanol, ethanol, propane, butane and methane. Of these, methanol and ethanol are of particular interest because of the reduced level of CO, HC and NO_x emissions which they produce and their capability of operating alternately or mixed with gasoline in a conventionally adjusted engine.

They do, however, have significant disadvantages compared with gasoline and diesel due to their higher latent heat of evaporation and lower useful energy content per unit volume. To reduce these problems methanol and ethanol are mixed with gasoline in varying proportions in what is known as a 'flex-fuel' vehicle.

To enable engines fuelled in this way to operate under optimum conditions during both cold and hot running, it is necessary to provide the engine control system with information on the relative percentage of gasoline and either methanol or ethanol at all times during operation, since this proportion may vary considerably depending on the operating conditions, the quality of the fuel used, the availability of each fuel and recent refuelling actions.

Most work has been done on methanol/gasoline mixtures and sensors. A sensor for this use has to take into account the variation in aromatic content of commercial fuels, the insolubility of methanol in gasoline at low temperatures and in the presence of water, and the variation in methanol density with temperature. It is also essential that the analysis of the fuel composition takes place in the fuel line to ensure that it is the true composition of the fuel reaching the engine which is measured and not composition variations within the fuel in the tank.

Three methods have been considered for making fuel composition measurements under engine operating conditions:

(i) measurement of infrared absorption at two different wavelengths;
(ii) measurement of refractive index;
(iii) combined measurement of capacitance, conductivity and temperature.

10.4.1 Infrared Absorption Fuel Composition Sensor

The combined near-infrared absorption spectrum of a methanol/gasoline mixture is shown in figure 10.11, for proportions of methanol mixed with the gasoline from 100% to 0%. It can be seen that the peak absorption between the wavelengths of 1550 and 1600 nm varies directly, although somewhat nonlinearly, with the percentage of methanol in the mixture, whereas in the 1300 nm region the absorption is low and virtually independent of the proportions of the mixture. The relative absorption at 1300 and 1600 nm has been shown to be relatively unaffected by changes in temperature or by the presence of a fuel/water emulsion. However,

changes in fuel density due to temperature do have a significant effect and need to be corrected for electronically.

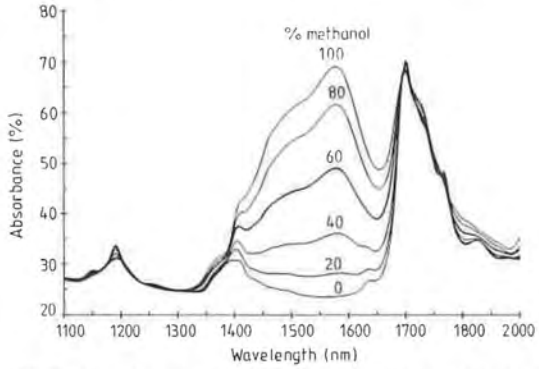

Figure 10.11 Combined near-infrared absorption spectrum of a methanol/gasoline mixture.

The configuration of a suitable twin-path infrared spectrometer capable of making absorption measurements in the fuel line at 1600 and 1300 nm is shown in figure 10.12. In this arrangement an infrared source illuminates a fibre optic bundle which has a small gap through which the flowing fuel passes and absorbs infrared radiation characteristically. The fibre optic bundle is then split and the two infrared beams pass through narrow-band interference filters transmitting at 1600 and 1300 nm respectively. The resulting radiation is detected by infrared sensors, and the two signals are passed into a comparator where the reading is linearised and corrected for variations in fuel density with temperature as measured in the fuel sensing cell, before being transmitted to the engine control system.

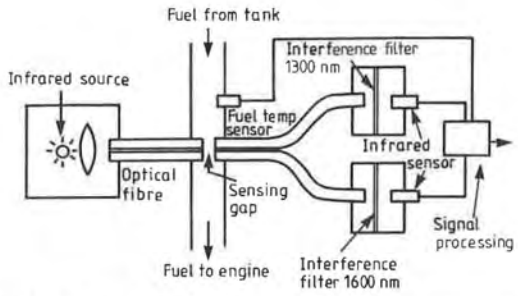

Figure 10.12 Twin-path infrared spectrometer for making absorption measurements in the fuel line.

10.4.2 Refractive Index Fuel Sensor

Refractive index fuel composition sensors use the fact that as fuel composition in a gasoline/methanol or gasoline/ethanol mixture changes, so does the refractive index of the mixture; this causes a rapid change of reflectivity with composition near the critical angle of reflection.

The layout of such a sensor is shown in figure 10.13. A light source illuminates the surface of the fuel mixture at a suitably small angle and the reflected light is measured by an area photodetector which is masked to accurately control the effective aperture of the optical system to just encompass the critical angles for light reflection at the extremes of the composition range of the fuel mixture. Under those conditions the amount of light reaching the photodetector is a direct measure of the refractive index of the fuel mixture with gasoline showing minimum reflectivity and methanol showing maximum.

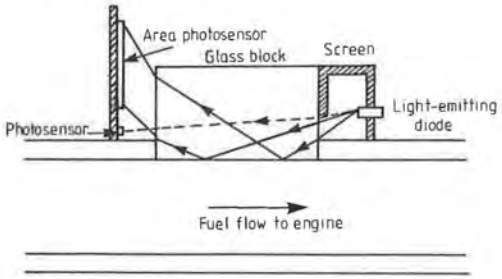

Figure 10.13 Layout of refractive index fuel composition sensor.

In a practical sensor the fuel would flow in a cylindrical glass tube and the space around the tube would be filled with a transparent medium of similar refractive index to glass to avoid problems caused by the high reflectivity at the glass/air interface.

It is important to shield the sensor from any direct illumination by the source as only a small percentage of the source emission reaches the detector via the controlled reflective path. It is also useful to have a separate photodetector to measure the source output directly to permit compensation for changes in source output.

10.4.3 Fuel Composition Sensing by Integrated Measurement of Capacitance, Conductivity and Temperature

It has been reported that at the 1991 Detroit Auto Show, VW demonstrated a multifuel Jetta which used a Siemens methanol/gasoline fuel mixture composition sensor which makes use of simultaneous

measurements of fuel dielectric constant, conductivity and temperature in order to determine mixture composition. No details of this combined sensor are available at the time of writing, so it is not possible to describe it in detail or comment on its accuracy compared with the infrared and refractive index sensors already described.

10.5 SUMMARY

A very large amount of work has gone into the development of EGO sensors, and the stoichiometric EGO sensors described in section 10.1 have made possible the accurate feedback control of automotive gasoline engines to a level which would otherwise not have been possible, and have enabled them to meet ever tougher emission regulations throughout the world, as well as being a prime driver for the introduction of automotive electronic systems.

In all the developed countries severe emission controls have become, or are becoming, compulsory so that almost all of the world production of 30 million gasoline-powered vehicles per year will in future have one of these sensors fitted.

The next generation of devices seem likely to be the lean burn sensors described in section 10.2, since the move towards mixed-cycle stoichiometric/lean burn engines appears to be the best way of making further progress to improve both emissions and fuel consumption (and therefore CO_2 emissions). The advent of lean burn catalysts, recently reported from Japan, would make lean burn sensors essential.

In the longer term more sophisticated engine controls can be expected for the measurement of individual exhaust gas constituents, and this is discussed in section 10.3, while the use of alternative fuels such as ethanol and methanol, and the fuel composition sensors required to make their use efficient and practical, are covered in section 10.4.

Gas and fuel composition sensors have had and will have in the future a major part to play in improving the performance of the motor vehicle and its acceptability to the public worldwide.

REFERENCES

[1] Katoh K, Iguchi S and Okano H 1992 Toyota lean burn engine — recent developments *Proc. 13th Vienna Int. Motor Symp.* pp 249–56

[2] Shulman M A and Hamburg D R 1980 Non-ideal properties of ZrO_2 and TiO_2 exhaust gas oxygen sensors *Society of Automotive Engineers Paper* 800018

[3] Pfeifer J L and Wertheimer H P 1985 Heated oxygen sensor based on thick-film titania *Proc. 5th Int. Automotive Electronics Conf. (IMechE Conf. Publication 1985-12)* pp 357–63

[4] Hetrick R, Fate W and Vassell W 1982 Oscillatory mode oxygen sensor *IEEE Trans. Electron. Devices* **29** 129–32

[5] Haaland D 1977 Internal-reference solid electrolyte oxygen sensor *Anal. Chem.* **49** 1813–17

[6] Velasco G, Schnell J and Croset M 1982 Thin solid-state electrochemical gas sensor *Sensors and Actuators* **2** 371–84

[7] Kamo T, Chujo Y, Akatsuka T and Nakamo J 1985 Lean mixture sensor *Society of Automotive Engineers Paper* 850380

[8] Wiedenmann H-M, Raff L and Noack R 1984 Heated zirconia oxygen sensor for stoichiometric and lean air–fuel ratios *Society of Automotive Engineers Paper* 840141

[9] Moseley P T and Tofield B C (ed) 1987 *Solid State Gas Sensors* (Bristol: Adam Hilger)

[10] Moseley P T, Norris J O W and Williams D E (ed) 1991 *Techniques and Mechanisms in Gas Sensing* (Bristol: Adam Hilger)

[11] Bishop G A, Starkey J R, Ihlenfeldt A, Williams W J and Steadman D H 1989 IR long path photometry: a remote sensing tool for automobile emissions *Anal. Chem.* **6** 671A–677A

11

Liquid Level Sensing

One of the earliest applications of electronic sensing to be adopted for automotive use was that developed for fuel level measurement. In early vehicles the fuel tank was often placed behind the dashboard, allowing the engine to be gravity-fed. As long as the tank remained in this position mechanical sensing devices such as a dipstick, gauge glass or (on more expensive vehicles) manometer could be used. However, for safety reasons the fuel tank was soon moved to its current position at the rear of the vehicle, and a pumped fuel supply was adopted. This introduced a need for electrical methods to measure and display the fuel level.

By the late 1920s remotely operated electrical petrol gauges were becoming standard. The approach almost uniformly embraced was the one still current today, in which a float at the free end of a pivoted arm in the petrol tank moves with changes in the fuel level. Displacement of the arm moves the wiper of a wire-wound or hybrid potentiometer, which in turn controls a remote (dashboard-mounted) meter. In early versions a voltmeter provided with a pair of windings was used. The windings were known as the *deflection* and *control* windings, and the control coil replaced the hairspring found in an ordinary voltmeter. The two windings were interconnected in such a way that changes in the battery voltage do not affect the reading, as shown in figure 11.1. For example, a decrease in battery voltage decreases the deflecting force, but also decreases the controlling force. The meter reading which results is a function of the ratio of the two forces, and is independent of battery voltage.

Modern fuel gauges are not usually voltmeters, since these have too fast a response and give a reading which fluctuates due to the fuel sloshing on hills and when cornering. Instead a meter is used which contains a bimetallic component and a heating coil as shown in figure 11.3 and discussed in the next section. The deflection of the pointer depends on the current in the heating coil, which is a function of the resistance of the float sensor in the fuel tank. A voltage regulator is usually fitted to remove the effect of variations in supply voltage. The

thermal inertia of the system is made sufficiently large to smooth out most of the effects of fuel slosh.

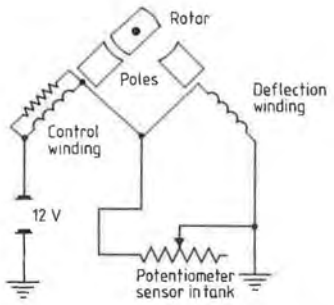

Figure 11.1 Circuit diagram for moving-iron meter fuel tank sensing system.

Many alternative ways of measuring fuel level have been proposed. One of the most promising uses the capacitance change which occurs when a pair of conducting plates are dipped into the fuel. Other reported systems have used optical, resonance and thermal techniques. A number of aircraft fuel gauging systems use ultrasound for tank level sensing. This development was introduced because it offers the possibility of non-invasively (and therefore safely) measuring the amount of liquid in a fuel tank. Sophisticated systems have been developed in which measurements from a number of transducers are integrated to give a reading which is independent of the motion of the fuel ('sloshing'). It seems likely that, as has happened before, a technology originally developed for aerospace will eventually be used in automotive engineering.

Liquid level sensors have now become standard on vehicles for a number of applications other than fuel level sensing. These include the fitting of transducers to measure the levels of windscreen washer liquid, engine coolant, engine and transmission oil, and hydraulic fluid. For these applications a full analogue measurement is not usually necessary, since what is required is a simple 'go/no-go' indication. The sensing techniques employed depend to some extent on the physical characteristics of the fluid involved, although systems in which a reed switch, Hall effect device or potentiometer is used to sense the displacement of a float are by far the most common.

11.1 POTENTIOMETRIC LEVEL SENSORS

To measure the quantity of fuel in a fuel tank the potentiometer, arm and float are usually mounted so that the float remains as close as possible to the centre of the tank for all fuel levels. This is partly because modern fuel tanks are often shallow and of uneven shape, and partly because the central position is less affected by vehicle attitude changes (such as climbing or descending a gradient).

Normally a nonlinear indication of fuel quantity is required, with the greatest sensitivity being given to the 'empty' end of the scale. This is achieved in two ways. First, an electrical approach can be adopted, by using a suitably nonlinear potentiometer. Second, mechanical means may be used involving careful design of the float and arm. Figure 11.2 shows a typical arrangement in which the float is connected to the arm at 90°, and is loaded with a weight, which has the effect of varying the float submersion throughout its travel. A circuit diagram for the sensor and dashboard indicator is shown in figure 11.3.

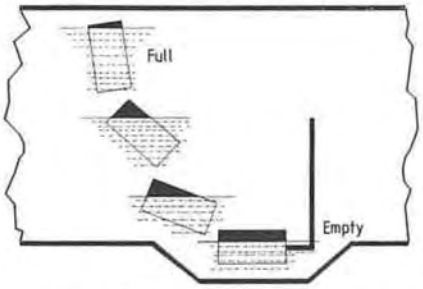

Figure 11.2 Diagram of fuel tank float.

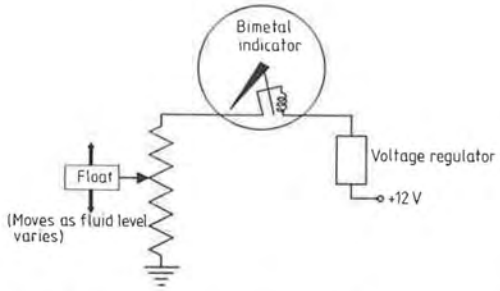

Figure 11.3 Potentiometer fuel tank sensing system.

Potentiometer and float transducers for monitoring the level of liquids other than fuel follow the same basic approach, but in general do not drive a level-indicating display. Instead a warning lamp or message is illuminated to alert the driver to the need for replenishment. In essence this is because hydraulic fluid, engine oil and coolant should not need attention more than once or twice a year.

The other major application for liquid level sensing, in the windscreen washer fluid reservoir, is non-critical, and does not justify the cost of an indicating gauge.

Potentiometers suffer from a number of problems, mainly concerned with wear and poor electrical contact, which were discussed in detail in chapter 8. However, they are very low-cost devices, and as a result are almost universally used for liquid level sensing. A number of alternative techniques have been proposed or demonstrated which do not rely on a sliding contact, and some of these are described in the following sections. However, at the time of writing none of them have made the step to mass production from being research prototypes or demonstration systems.

11.2 CAPACITIVE LIQUID LEVEL SENSORS

Petrol and diesel are non-conducting and have a relative permittivity of about 2. The temperature coefficient of permittivity is very low, typically $-0.06\%\ °C^{-1}$. This property has led to capacitance fuel gauging being common in aircraft applications. The sensor used in aircraft consists of a parallel-plate capacitor (in the form of a pair of concentric tubes) running vertically from top to bottom of the fuel tank. The relative permittivity of air is 1, so the capacitance of the probe doubles when the tank is filled with fuel. A high-frequency signal is applied to one plate of the capacitor, and the AC current flow to earth is measured. In aircraft the signal is also applied to a reference capacitor which always has air as the dielectric, and the current flow through this is subtracted from that through the sensor, so that the final current flow is zero when the tank is empty.

The nominal capacitance of this type of sensor is small, and it has been the practice in the aircraft industry to measure the capacitance change by frequency bridge techniques. Systems of this kind have to be individually calibrated for each installation, and it is this, together with the relative cost and complexity of the electronics involved, which has probably prevented the introduction of capacitive fuel gauging on road vehicles. The cost of electronics continues to fall, however, and a number of designs have been published in which the need for individual calibration is greatly reduced if not eliminated [1,2].

A basic capacitive fuel level sensor is shown in figure 11.4. It consists of a pair of concentric tubes, the outer of diameter $2R$ and the inner of diameter $2r$. The capacitance of the assembly is given by

$$C = \frac{2\pi\varepsilon_0\varepsilon_r}{\log(R/r)} \tag{11.1}$$

where ε_0 is the permittivity of free space and ε_r is the relative permittivity of the medium filling the dielectric space between the tubes. A capacitive transducer of this type effectively compares the dielectric constants of air and fuel respectively. If the sensor is partly immersed in a dielectric liquid such as fuel, the resulting capacitance may be shown to be a linear function of the depth of immersion.

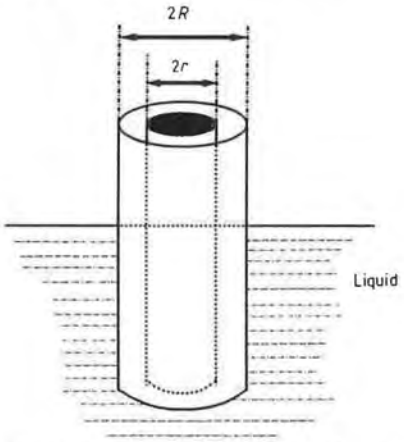

Figure 11.4 Concentric tube capacitive level sensor.

11.2.1 The Effect of Dielectric Variation

If the permittivity of a range of petroleum-based automotive fuels is measured it is found to vary by around 5%. In addition additives such as those used in unleaded petrol can increase these variations by very large factors. To be successful therefore a capacitive fuel sensor must include some form of compensation for variations in the dielectric properties of fuel. A typical compensated system is described in [1]. Two probes are used, one of which acts as the reference for fuel type, and one which measures the height of fuel in the tank. The reference probe is fitted in the fuel line so that it is completely immersed and has a continuous fuel flow through it. This is so that the fuel permittivity can be constantly checked,

and it should correspond closely to that in the tank. The main problem with this system is reported to be the difficulty of ensuring that the reference probe is completely immersed at all times. If, for example, the tank is run dry, some reliable means of refilling the reference probe is required.

11.3 OPTICAL LIQUID LEVEL SENSING

Very few optical fuel level sensors have been described in the research literature, although a number of fuel flow sensors based on optical techniques have been reported [3]. Obviously, the quantity of fuel remaining can be derived from a flow measurement by integration if the initial conditions are known.

Optical sensors for engine oil level have been more frequently reported, however. A typical example is described in [4]. The device depends for its operation on the modification to total internal reflection which occurs when an optical component is immersed in a liquid. As shown in figure 11.5, a light-reflecting 90° prism will return all of the input energy to a detector if it is in air, but loses some of the incident light when immersed in liquid. The signal from the photosensor therefore indicates whether the device is immersed, providing a simple go/no-go output.

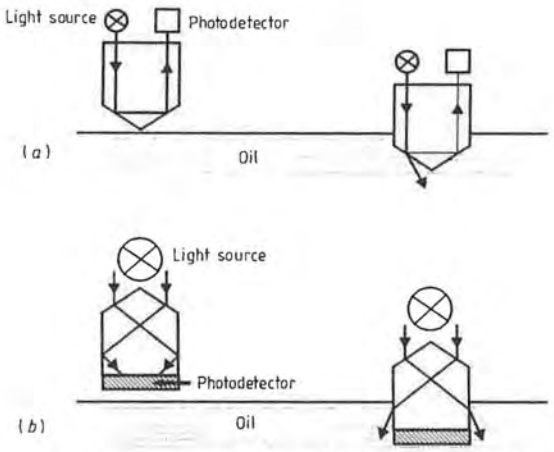

Figure 11.5 Principles of optical fluid level detection: (*a*) simple prism liquid level sensor; (*b*) prismatic light guide level sensor.

A number of problems have been reported with this approach. Residual oil films and droplets on the base of the prism can introduce monitoring errors. The solution proposed is to use the sides of the device for reflection, rather than the base, as shown in figure 11.5(*b*).

An alternative approach is to replace the prism with an exposed optical fibre as shown in figure 11.6. A single fibre (a monofilament) is bent through 180° with a fixed radius at the sensing point. If the sensor is in air total internal reflection ensures that almost all of the energy returns to the photodetector. If it is immersed in a liquid some or all of the light will emerge and be lost in the fluid, reducing the detector output signal amplitude.

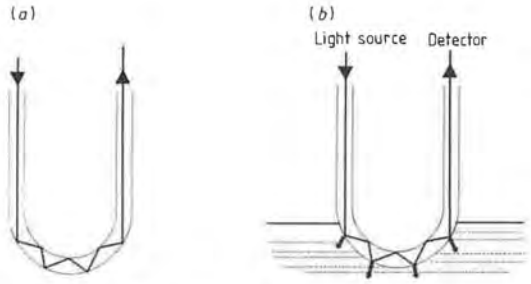

Figure 11.6 Optical fibre liquid level detector: (*a*) total internal reflection; (*b*) light lost to liquid.

The advantage of the optical fibre technique is that the light source and detector can be positioned away from the engine, unlike the system shown in figure 11.5(*b*). The main disadvantages, however, are that the use of a monofilament requires precision connectors and sensitive photocouplers. In addition the technique appears vulnerable to error if an oil film or droplets form on the sensor.

11.4 RESONANT STRUCTURES FOR LIQUID LEVEL SENSING

An electrically driven mechanical oscillator in the form of a solid rod can be used to sense the depth to which the rod is immersed in a liquid. Fluid loading and viscous damping effects cause changes in the resonance frequency of the rod, and these can be measured and used to deduce the

liquid level. A number of liquid level sensors of this type have been proposed (see, for example, [3]). However, other reports suggest that it is difficult to distinguish between a moistened (i.e. covered with a liquid film) and an immersed sensor of this type over the full temperature range found in automotive applications [5].

A typical sensor construction is shown in figure 11.7. A rod of aluminium or stainless steel is used, which acts as an electromechanical resonator. Two piezoelectric transducers are attached to the orthogonal axes of the rod. One of these provides the actuation which sets the rod into bending resonance, while the other transduces the motion back into an electrical signal which is fed to a phase-locked loop (PLL) circuit. As the depth of immersion is varied the resonance frequency changes, and the output from the PLL circuit is a function of this depth.

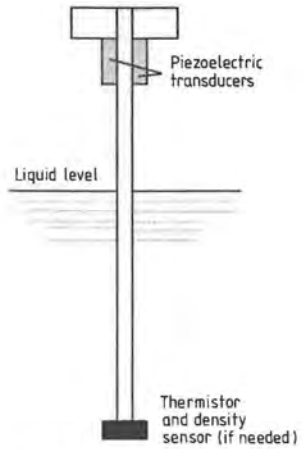

Figure 11.7 Resonating rod level sensor.

The amplitude of the flexural vibrations of the rod is very small. Reference [3] gives as a typical figure ± 1 μm. The change in resonance frequency is therefore mainly a function of the fluid loading, rather than being due to viscous or shear effects. It is claimed that sensors of this type can measure fluid mass, if the geometry of the surrounding container is known and if the fluid density is constant. The density dependence probably means that this technology is restricted to fuel tank level sensing, since other automotive liquids (and in particular engine oil and hydraulic fluid) experience quite large density changes as their temperature rises.

According to [3], devices of the type described above are commercially available for sensing the level of fuel in bulk delivery vehicles. However, widespread automotive use of the resonator principle has not been reported. This is probably because of technical difficulties such as the density dependence mentioned above, and may also be because the technology required is relatively expensive and cannot compete in cost with the simple float/potentiometer designs described earlier.

11.5 THERMAL LIQUID LEVEL SENSORS

Thermistors (see chapter 5) are sometimes used to give a crude indication of liquid level. Since a thermistor is self-heating, its equilibrium temperature when subjected to a constant current depends on the heat transfer conditions between the device and its environment. The temperature of a thermistor obviously affects its resistance, and a step change in resistance is seen if a previously submerged thermistor is exposed to the air (or vice versa). This phenomenon can be exploited in the design of sensors for engine oil, hydraulic fluid and windscreen washer water level. No analogue indication of the liquid level is given, but a simple 'yes/no' level warning can readily be obtained.

Thermistor level sensors of this type are most commonly used as dipstick engine oil detectors. This is because the small physical size (see chapter 5) of a thermistor allows it to be mounted on a conventional engine dipstick and inserted through the narrow opening provided for access.

A number of difficulties are commonly experienced with this type of level detection. First, there is a delay in reaching steady-state conditions after the current is switched on. In severe cases this may mean that the car can be driven away before an indication of dangerously low oil level is given.

Other problems are that the 'switching point' is partially dependent upon the temperature of the working fluid (normally oil). The switching point may also be affected by any liquid film adhering to the sensor. Since the impedance of a thermistor is relatively low, special care is needed to ensure that low-resistance cable and interconnections are used.

REFERENCES

[1] Clark A J and Sikka J 1979 A new capacitive fuel level transducer *Proc. 2nd Int. Conf. on Automotive Electronics* (London: IEE) pp 159–65 ISBN 0 85296 210 X

[2] Huddart J 1979 An alternative approach to automotive fuel gauging

Proc. SAE Symp. SP-441 (Warrendale, PA: Society of Automotive Engineers) pp 1–7

[3] Goodyer E N 1989 An overview of a range of novel automotive sensors *Proc. 7th Int. Conf. on Automotive Electronics* (London: IMechE) pp 79–86 ISBN 0 85298 697 1

[4] Zuckmantel E, Heddergott H and Holland C R 1985 Optical engine oil level sensor *Proc. 5th Int. Conf. on Automotive Electronics* (London: IMechE) pp 325–30 ISBN 0 85298 569 X

[5] Zuckmantel E 1983 Engine oil level monitoring *Proc. 4th Int. Conf. on Automotive Electronics* (London: IEE) pp 27–31 ISBN 0 85296 283 5

12

Smart Sensors

The term 'smart sensor' is not yet in widespread use, although a number of somewhat conflicting definitions have been promulgated. Indeed, at the time of writing it is claimed that there are no integrated smart sensors on the market [1]. It is sensible therefore to try and define what a smart sensor is, and what it can do, before moving on to examine the potential automotive applications of such a device.

A smart sensor is an electronic device which combines all of the functions needed for measurement and subsequent data conversion to produce a bus-compatible digital output signal. To create a smart sensor many silicon micromachining and fabrication techniques are exploited: a silicon sensing element is combined with analogue signal conditioning circuits, temperature or other compensation systems, linearisation circuits, and analogue to digital conversion.

Smart sensors do not necessarily have to take the form of a single-chip silicon device. Studies have been reported in which two or three silicon devices are interconnected using thick-film hybrid construction techniques. Thick-film sensing elements may also be used. The resulting assembly is called a smart sensor by some workers, although the package size is of course larger than that achieved by a single-chip silicon device.

The output of a smart sensor is configured in a standard digital form, suitable for connection to a microprocessor bus. Many smart sensor designs include a microprocessor on the chip to handle tasks such as compensation, linearisation and bus communication. Self-testing, calibration or diagnostic functions can also be built-in. However, in at least one study it has been found that there is an unacceptable increase in silicon area and complexity, and hence a prohibitive cost penalty, when a microprocessor is included on the chip [2]. The optimum division of responsibility between the smart sensor and the main system microprocessor appears to be a very difficult matter to determine.

12.1 THE NEED FOR SMART SENSORS IN AUTOMOTIVE ENGINEERING

Sensors are essential in any automatic control system. In general control of any automated process (such as management of the ignition timing or fuel injection in a car engine) is achieved by three functional blocks, as shown in figure 12.1. Sensors are used to acquire information about the process to be controlled, a microprocessor is used to decide what action should be taken, and finally actuators are used to carry out the actions determined by the microprocessor.

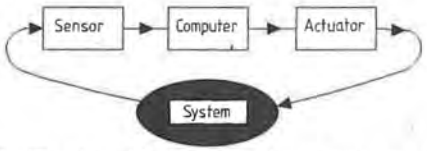

Figure 12.1 The three blocks comprising an automated process.

The costs associated with these three blocks have changed in a markedly different manner over the past 20 years. For actuators, such as electric motors, the price–performance ratio has improved by a factor of approximately 10. For microprocessors and computing power in general the improvement is much more marked, and has been estimated to be of the order of 1000 [1]. For an average sensor, however, the price–performance ratio has only improved by a factor of three. Moreover, each kind of sensor requires its own specialised signal conditioning system, which makes the application of electronic instrumentation expensive. This lag in the progress of sensor technology hampers the process of automation, and its application to motor vehicles, and it is this deficiency which the 'smart sensor' approach is attempting to address.

The modern car contains on average about 30 sensors. These include, for example, the transducers necessary to ensure efficient and clean operation of the engine. Typical automotive measurands are described elsewhere in this book, and include temperature, air flowrate, pressure, chemical concentration and so on. In a conventional data acquisition system these transducers are connected to the central microprocessor in a 'star' configuration, as shown in figure 12.2(*a*). Each sensor is connected to its own local signal conditioning circuit and has its own wiring. The size and cost of the wiring looms is rapidly becoming unacceptable, and the problem is likely to grow, since it is predicted that the number of sensors on a vehicle will exceed 100 by the end of the century [3].

The wiring problem outlined above is solved by the adoption of the smart sensor approach shown in figure 12.2(*b*). Each sensor is now provided with its own integrated circuit, which conditions the sensor

signal, compensates and linearises it as necessary, and interfaces with a digital transducer bus which links all the sensors in a 'ring-main' arrangement as shown. This systematic and standardised approach has obvious advantages and greatly simplifies the wiring. A similar strategy has been applied in aircraft for a number of years, but separate sensing, signal conditioning and digital modules are used. The approach used in aircraft has until now been far too expensive for automotive engineering, but now seems likely to become cost-effective with the introduction of single-chip smart silicon or hybrid microelectronic sensing devices. The manufacturing technology which has made possible a thousand-fold reduction in the cost of computing power can also be used to manufacture smart sensors, in which the sensor and its associated electronics are integrated on one chip, without the need for any external components.

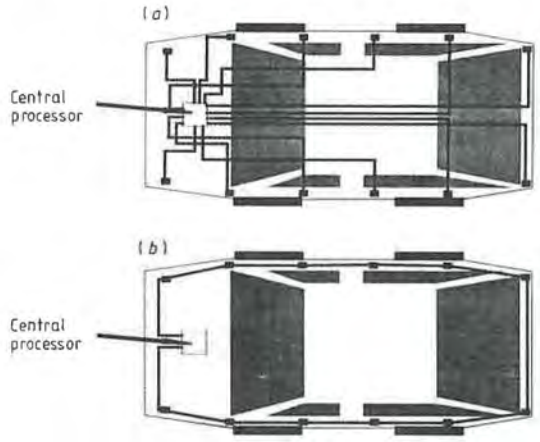

Figure 12.2 Wiring simplification achieved by smart sensors. (*a*) Conventional star connection for automotive sensing. (*b*) Distributed smart sensing system with 'ring main' digital bus. Sensors are shown as ■.

12.2 MANUFACTURING SMART SENSORS

To make a smart sensor the appropriate transducers and circuits have to be combined on a single piece of silicon, or else fabricated using thick-film techniques as a hybrid circuit. The manufacturing processes used can severely restrict the sensing techniques available. The feasibility of manufacturing an integrated (silicon or hybrid) sensor may be assessed by answering the following questions:

(i) Is the device to be made in sufficient quantities to warrant the expense of setting up a silicon production facility? As a very rough rule of thumb, silicon devices only become cost-effective if at least 300 000 are to be made.

(ii) Can a suitable transducer be made using standard silicon micromachining techniques? Or is it possible to use thick-film piezoresistors as the sensing element?

(iii) What signal conditioning and conversion circuits are needed, and can they be implemented without external components (which dramatically increase the cost)?

(iv) If the answers to questions (i), (ii) and (iii) rule out a single-chip silicon solution, is a hybrid (surface-mounted silicon components with thick-film interconnections and/or sensing element) solution feasible?

(v) Do any new possibilities result from combining sensors and circuits on one device?

(vi) Is a large research effort necessary to design the device?

12.2.1 Restrictions Inherent in Silicon Processing

Standard IC fabrication techniques have been used to produce sensors for the six energy domains summarised in table 12.1, which is taken from [1]. It should of course be borne in mind whenever the use of silicon is contemplated that the upper operating temperature is limited to about 200 °C. This restriction does not apply to thick-film sensing elements, which in the form of piezoresistors will operate up to about 400 °C.

The silicon optical sensors are mainly diode-based. Up to 10^6 diodes can be placed on one silicon chip for use in applications such as charge-coupled device (CCD) solid state television cameras. Phototransistors are used for sensing low light levels, and position-sensitive photodetectors are also available for measuring the position of a light spot. The main difficulty encountered in fabricating optical sensors from silicon is that a transparent window is required.

Silicon temperature sensors are also relatively easy to make. As discussed in chapter 5 (section 5.4), a PN junction makes a good temperature sensor. The ambient temperature usually penetrates a standard IC package relatively quickly, although care has to be taken to ensure that the on-board circuitry does not generate so much self-heat that an error is introduced. Thermocouple arrays can also be deposited by standard thin-film techniques, and with this approach quite small temperature changes (of the order of ±0.1 °C) can be detected. Heat-flux sensors (again see chapter 5) can also be made by adapting the packaging technique used.

Table 12.1 Sensors which can be fabricated in silicon by standard IC processes.

Sensor type	Main problems
1 Optical	
Photodiode for sensing	Encapsulation
brightness	
colour	
radiation	
Charge-coupled devices:	Encapsulation
Phototransistor	
Position sensitive devices	
2 Mechanical	
Piezoresistors for sensing	Micromachining
force	
pressure	
acceleration	
Capacitive techniques for	Encapsulation
displacement	
tactile sensing	
3 Thermal	
PN junctions for temperature	Self-heating
Thermocouples for temperature and heat-flux	Encapsulation
4 Magnetic sensors	
5 Hall effect for proximity, rotation rate	None
6 Chemical sensors	
Chemocapacitors for moisture	Special processing
ISFET for chemical concentration, pH	Special processing
SAW (surface acoustic wave) devices for gas composition	Special processing

Much more difficult problems have to be overcome in designing mechanical sensors in silicon. The measurand here is usually displacement (or a derived quantity such as force, pressure, acceleration etc). Piezoresistors and piezojunctions can be made using standard processes, but micromachining techniques are required if membranes, cantilevers or other mechanical structures are required.

Chemical sensors are also difficult to manufacture in silicon. Integrated circuits are normally encapsulated to shield the device from chemical attack by water or other aggressive agents. However, this requirement is in direct contravention to the need to expose the device to chemical agents for measurement purposes. Nevertheless, measurands such as the hydrocarbon concentration in a car exhaust have been addressed

successfully using transducers such as ion-sensitive field effect transistors (ISFETS).

The final sensor group likely to be of interest to automotive engineers is magnetic transducers. Silicon magnetic sensors are almost always based on the Hall effect [4], and can easily be integrated using standard fabrication processes for applications such as contactless switches, proximity sensors or rotation rate transducers. A magnetic field will readily penetrate most standard IC packages.

12.2.2 Restrictions Inherent in Thick-film Hybrid Manufacturing

Thick-film materials are fired at a temperature of around 800–900 °C. They must therefore be laid down onto a suitable substrate. The materials most commonly used are alumina-based ceramics or stainless steel [5]. Although this requirement may appear to impose a severe restriction, it has the benefit of allowing thick-film transducers to be operated at temperatures up to 300 °C or more without damage. Obviously, with a hybrid construction in which silicon devices are placed on the same substrate it is not possible to operate at such high temperatures without resorting to special cooling or insulating techniques.

Thick-film sensors are usually based on piezoresistors, and are therefore routinely used in applications such as force, acceleration or pressure measurement. Research is currently under way into other areas, such as chemical sensing and the production of piezoelectric thick-film inks, and it seems likely that thick-film transducers will eventually be available for use with measurands other than force/displacement.

12.3 CONCLUSIONS

Combining the sensing element and some or all of the signal processing and conversion circuits on one chip or hybrid assembly has a number of obvious advantages. The device is low cost if sufficient numbers are produced, and since batch processing is used the device-to-device variation is very small. The interconnections are very short, so the sensitivity to electromagnetic interference is greatly reduced. This may be a very considerable advantage in automotive applications. A considerable wiring reduction is achieved, with typically only two wires being required to connect the smart device to a central controller. Cross sensitivities may be reduced or eliminated by adding sensing elements which respond only to

the unwanted parameters, enabling appropriate corrections to be made at the signal processing stage. Similarly, offsets and nonlinearities can be adjusted. Smart sensors can be made very small, and may therefore be placed in positions which are not practical when 'standard' sensors are used. In automotive engineering this may for example make it possible to perform much more detailed cylinder-by-cylinder monitoring of the combustion process than is possible at present.

In considering any engineering system it is seldom possible to present a list of advantages without compiling a corresponding list of disadvantages, and smart sensors are no exception. The main drawbacks of smart sensors are that the initial tooling-up costs of silicon fabrication are high, and may only be justified when a large number (>100 000) of devices are required. If smaller numbers are needed thick-film hybrid construction may be a better alternative. The close proximity of sensing elements and circuitry can lead to unexpected feedback problems, such as self-heating. It is difficult to make the transducing element sensitive to a given measurand without also giving the adjacent circuitry some sensitivity. In many smart sensors made from silicon the circuitry is exposed to the measurand, which compounds the problem.

REFERENCES

[1] Huijsing J H 1992 Integrated smart sensors *Sensors and Actuators* A **30** 167–74

[2] Haskard M R 1990 An experiment in smart sensor design *Sensors and Actuators* A **24** 163–9

[3] Hofflinger B *et al* 1990 Integrated electronics for automotive applications in the EUREKA programme PROMETHEUS *Proc Essirc '91*

[4] Turner J D 1988 *Instrumentation for Engineers* (London: Macmillan) ISBN 0 333 47295 0

[5] White N M 1989 An assessment of thick-film piezoresistors on insulated steel substrates *Hybrid Circuits* no 20 (September) 1–5

13

Sensors for Intelligent Vehicles on the Road

For almost its entire existence the car has been considered to be a self-contained people-transporting system, with its control and communication with the outside world handled entirely by the driver, perhaps with help from the passengers on some occasions! The driver views the traffic and the road scene ahead and controls the vehicle appropriately for the situation in that scene. The driver listens to the engine noise and changes gear accordingly, and even monitors normal operation by subconsciously listening for unusual noises as an indicator of any incipient fault. However, as traffic congestion has increased, road systems have become more complicated and practical vehicle speeds have increased, the sensory and control load on the driver has become greater and the driver's ability to react in an emergency has been compromised.

Together with this increase in driver load has been combined the long-present problem of navigation of the vehicle in what may be an unfamiliar road system to a destination identified only as a point on a street map. This frequently results in dangerous situations where the driver attempts to drive and read a street map at the same time, usually in a congested traffic situation.

Because of the heavy traffic now commonly existing, the driver also has difficulty in reacting quickly enough to rapidly changing traffic and anticipating dangerous situations ahead such as obstacles (particularly in fog), accidents and the dangerous overtaking conditions which may exist on high-speed motorways.

This unsatisfactory situation has led to extensive development of vehicle navigation systems, and of methods of communication between individual vehicles and between vehicles and the road infrastructure, giving the ability to provide vehicle route guidance and navigation systems and warnings to the driver of dangerous situations, accidents and traffic congestion which can then be avoided. Much of this development is currently taking place within the major joint European research programmes PROMETHEUS and DRIVE, where it is known as Road

Transport Informatics (RTI). Similar work is also in progress in the USA (known there as Intelligent Vehicle Highway Systems (IVHS)) and in Japan.

The PROMETHEUS programme was started as an initiative of the European automotive industry and has been operating since 1986 within the framework of the Europe-wide EUREKA research project; it is dedicated to considering vehicles as active elements integrated into the road traffic and transport system. The DRIVE programme was started in 1987 and is a European Community (EC) project concerned more with the development of the infrastructure associated with road traffic and transport than with the vehicle itself.

The two projects have operated interdependently from the start, but have become more closely integrated since 1989 through a coordinating committee (SEFCO), and since 1991 through a more comprehensive coordinating organisation (ERTICO), with the task of defining a suitable implementation strategy for the results of the developments from PROMETHEUS, DRIVE and other relevant European R&D projects.

The major PROMETHEUS and DRIVE projects which involve the significant use of sensors in vehicles are:

(i) Dual-mode route guidance — the combination of an autonomous route guidance system using an on-board digital road map and dead-reckoning navigation equipment combined with an infrastructure-based system.

(ii) Cooperative driving — self-organised cooperation between vehicles based on direct vehicle communication and vehicle infrastructure–vehicle communication supporting the control of manoeuvres within traffic for lane changing and overtaking, intelligent cruise control for harmonising speed and distance between vehicles, intelligent intersection control and medium-range pre-information to inform the driver in advance of safety and traffic-related occurrences.

(iii) Vision enhancement — methods of enhancing visibility in adverse environmental conditions by the use of technologies such as infrared and/or ultraviolet illumination of the scene ahead of the driver, thermal imaging and gated intensified video camera technology combined with a head-up display.

(iv) Collision avoidance — perceiving the environment and traffic situations using multiple sensors, predicting possible collisions with objects and other vehicles and providing appropriate driver information and possible intervention where appropriate.

Similar issues and systems of vehicle information and control are being addressed in the USA through the more recently started IVHS

programme and also in Japanese cooperative projects. Some of the equipment developed for use in these and other related projects and the systems and the sensors required are described in detail in the remainder of this chapter.

13.1 NAVIGATION AND ROUTE GUIDANCE SYSTEMS

Navigation systems for road vehicles have been of interest for a number of years, but it is only recently that the technology has become sufficiently developed to make possible reliable systems at an acceptable cost.

The first practical system was developed by Blaupunkt (Bosch) and is known as ALI; it makes use of the inductive loop sensors in the road, which are commonly used to control traffic lights, to communicate between an inductive loop receiver/transmitter mounted in the vehicle and a central computer which coordinates traffic information for the area. Drivers input their required destination using a keyboard in the car; this is transmitted via the inductive loop system to the central computer which calculates the best route to the required destination avoiding traffic jams, road works, etc. It then gives the driver this routing information in the form of turning and direction instructions at each junction.

Since ALI was developed two other methods, using short-range radio or infrared, have been used to communicate with vehicles at junctions, particularly to permit the transfer between vehicle and infrastructure of larger amounts of information. Development of self-contained on-board (autonomous) navigation systems has also made considerable progress in the last few years, making use of low-cost precision magnetic field sensors to obtain bearing heading information, and differential wheel rotation counting for turn information and distance travelled, together with comprehensive map information on compact disc and the capability of identifying the true position of the vehicle by matching the turns made by the vehicle to this map information.

Systems making use of navigational information from satellites have also been developed, and the capability of some systems to use traffic information transmitted on the new radio data system (RDS) currently being introduced in Europe is under investigation, as is the possibility of using cellular radio.

A large number of companies and organisations are developing systems, and significant development is taking place at Bosch/Blaupunkt, Philips and Siemens/GEC in Europe and ETAK in the USA. There is also a joint programme in Japan including all the major motor manufacturers and electronics suppliers.

The Bosch/Blaupunkt system is called Travelpilot and is modelled on a design by ETAK. It is self-contained and based on the storage of map information over a large area in digital form on a compact disc in the vehicle. This is combined with a dead-reckoning system which continuously measures distance travelled, differential wheel rotation and vehicle directional bearing. The driver selects a destination by punching in a suitable code on a keyboard in the vehicle and is then given a map of the area displayed on a cathode-ray tube with the positioning of the vehicle shown and the position and general direction of the destination, it is then up to the driver to follow the route on the displayed map which continually changes position to maintain the vehicle's position at the centre of the screen.

The system overcomes the problems of accumulating error by identifying the true position of the vehicle by matching the turns made by the vehicle with the map stored on the compact disc. There is currently no capability for feeding traffic information into the system, although inputs from RDS or radio beacons could be added at additional cost.

The Philips system, known as CARIN, is very similar in method of operation to the Bosch system, but as an alternative to the map display it can also provide a route guidance display as well as using voice synthesis to give information to the driver. It also requires a digitised map which is stored on compact disc in the vehicle, and in the same way as Bosch employs map matching to overcome the dead-reckoning accumulated error problem. Similarly to the Bosch system, there is currently no capability for feeding external traffic information into the system, but this could be provided at additional cost. Philips have been working closely with Renault on developing their system for in-vehicle use.

The Siemens/GEC system, originally known as 'Autoguide' in the UK (now as APPLE) and 'ALI-SCOUT' in Germany, operates by using infrared transmitter/receiver beacons at selected junctions to communicate directly with vehicles equipped with similar infrared transmitter/receivers. Those beacons connect each vehicle to a central computer. A driver entering the area selects a destination by punching in a suitable code on a keyboard and the system displays the direction in which to turn at each junction, gives information on the distance from the turn and voice information on the necessary road lane to adopt. Route guidance is by a combination of the map information stored in the central computer and dead-reckoning based on measuring directional bearing and distance travelled, any resetting required taking place at infrared beacons. No compact disc storage of map information is required in the vehicle.

A major advantage of the system in the urban situation is that the central computer automatically records the speed with which equipped

vehicles travel between beacon-equipped junctions, and therefore can immediately redirect following vehicles on alternative routes if that speed drops below a certain specified level due to traffic jams, roadworks or other hold-ups. The capability to react in real time to road traffic conditions is the major strength of the Siemens/GEC system compared to the autonomous Bosch and Philips systems.

Pilot schemes have been run in London and Berlin and there are plans, which are still under review, to extend these to give full-scale operation with extensive network infrastructures; in the London case this will cover a substantial section of the area within the M25 orbital motorway.

As has already been described the European PROMETHEUS and DRIVE joint research programmes are particularly involved with developing navigation and route guidance systems to operate within Europe-wide standards and to develop dual-mode systems using a combination of autonomous and infrastructure-based technology to provide a Europe-wide compatible system by the late 1990s.

In the USA early work was done by ETAK on the system eventually licensed by Bosch/Blaupunkt but interest in these systems has increased as road congestion has become an increasing problem, with the establishment in 1990 of the Integrated Vehicle Highway Systems (IVHS) programme. Both GM and Ford are now developing systems.

In Japan, however, the Ministry of International Trade and Industry (MITI) has worked with industry to promote the development and use of automobile navigation aids since the early 1980s. Systems developed jointly by Japanese government agencies and industry include RACS (Road and Automobile Communication System), which uses proximity microwave beacons to connect vehicles to the infrastructure combined with dead-reckoning, and AMTICS (Advanced Mobile Traffic Information and Communications System), which uses cellular radio to communicate traffic data to vehicles. The Japanese automotive companies have also been extremely active, with Mazda, Toyota and Nissan all developing systems using dead-reckoning with map matching and, in the case of Mazda, with navigation satellites being used to give position updates.

All these systems use similar sensors, in particular those for directional bearing, distance travelled using wheel rotation measurement and turn direction using differential wheel rotation measurement. These sensors are common to virtually all systems since they provide the basic dead-reckoning information to enable sophisticated map matching techniques to be used, or to enable a vehicle to navigate between successive position updates from the infrastructure. In the case of systems using infrared methods of communication with the infrastructure, the infrared sensor technology and information processing used in the transmitter/receiver is of interest from the point of view of sensor development.

13.1.1 Directional Bearing Measurement

All current vehicle navigational systems use a device for sensing the direction in which the vehicle is pointing relative to the earth's magnetic field, this being essential for dead-reckoning and map-following navigation.

Current methods of sensing the direction of the earth's magnetic field, using flux gate magnetometers, are unfortunately not as accurate as is really required and they are severely affected by the proximity of the vehicle to large steel-framed buildings and large metal vehicles. The possible solution to this is the use of gyroscopic methods of directional bearing measurement and this alternative for the future will be discussed later.

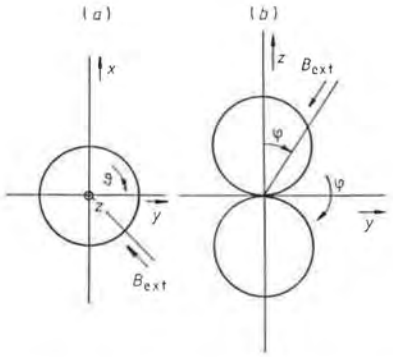

Figure 13.1 Directional sensitivity of flux gate magnetometer to an external field (B_{ext}) (*a*) perpendicular to probe axis (*z*) and (*b*) along probe axis showing the cosine response with minimum sensitivity at an angle of 90° to the longitudinal axis of the element and maximum sensitivity on the longitudinal axis.

The flux gate magnetometer actually measures field strength, but with suitable probe design has a directional sensitivity as shown in figure 13.1. It operates by comparing the magnetic field to be measured with a magnetic reference field generated in a premagnetisation winding within the magnetometer (figure 13.2). The magnetic reference field normally has a triangular form, and this is applied to a coil surrounding the probe to superimpose an induced magnetic field on the externally applied earth's field. Figure 13.3(*a*) shows the principle. A pick-up coil also surrounding the probe picks up the combined magnetic field variations and by detecting the zero cross-over points, as shown in figure 13.3(*b*), a pulse

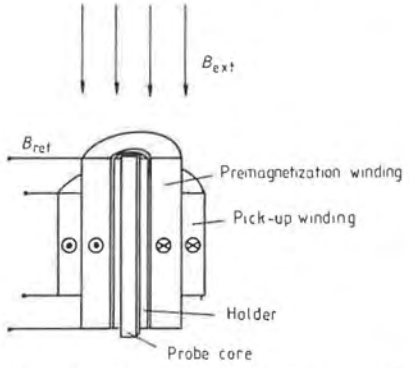

Figure 13.2 Flux gate magnetometer showing premagnetisation and configuration of pick-up winding.

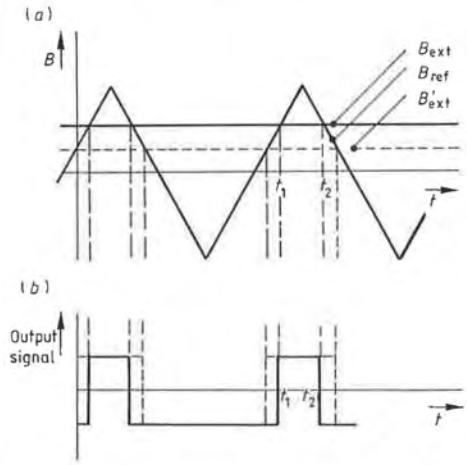

Figure 13.3 (*a*) Triangular reference flux (B_{ref}) superimposed on two different external flux levels (B_{ext} and B'_{ext}) and showing cross-over times (t_1, t_2). (*b*) Pulse width modulated signal derived from (*a*) with the width being proportional to the external flux level.

signal can be produced with a width which is a linear function of the external field strength. Since the probe has a maximum sensitivity along its magnetic axis and minimum sensitivity at 90°, directional information can be obtained, although as can be seen from figure 13.1 this directional sensitivity varies as a cosine function. The ratio of sensitivities between the two axes is reported as being about 10^{-5} [1]. By arranging three probes

orthogonally on x, y and z axes it is possible to sense the magnetic field vector independently of the orientation of the probe system and reasonable accuracies of bearing measurement ($\pm 1°$) can be obtained in ideal conditions. In practice, however, in the vehicle, the distortion of the earth's magnetic field by substantial metal objects on or close to the road reduces this accuracy considerably and makes it essential to combine the output of the flux gate magnetometer with an independent measure of the amount the vehicle has turned. This is obtained by measuring the difference between the rotations made by two (ideally undriven) wheels of the vehicle.

On vehicles equipped with anti-lock braking (ABS), the wheel rotation sensors already used for the ABS system may conveniently be used. For those vehicles not so equipped, suitable sensors would need to be added. The ABS sensors are described in detail in section 8.7.1. The wheel rotation sensors also give distance travelled to a reasonable degree of accuracy.

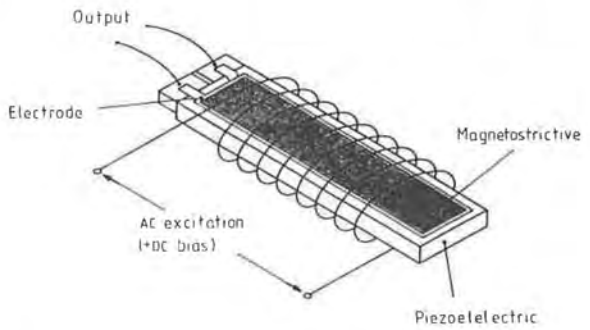

Figure 13.4 Magnetic field sensor using a magnetostrictive/piezoelectric laminate with DC and AC excitation.

An alternative method of sensing the direction of the earth's magnetic field has been described [2] in which the magnetic field sensing element consists of a laminate of magnetostrictive and piezoelectric materials in the form of a thin strip which is placed inside a coil through which AC excitation and optional DC bias conditions are passed (see figure 13.4). The magnetostrictive material is directly mounted onto the piezoelectric element, and when subjected to a magnetic field its length changes by an amount proportional to the square of the magnetic field. If in the device proposed AC excitation is applied so that the magnetostrictive material is subjected to an AC magnetic field then the magnitude and phase of the extension is dependent on the DC magnetic bias.

The magnetic field, H_m, experienced by the magnetostrictive material consists of three separate elements

$$H_m = H_a + H_b + H_1 \cos \omega t$$

where H_a is the external magnetic field to be measured, H_b is the applied DC magnetic bias, and $H_1 \cos \omega t$ is the AC magnetic bias. The extension of the magnetostrictive material is therefore proportional to

$$H_m^2 = [(H_a + H_b)^2 + H_1^2/2] + [2(H_a + H_b)H_1 \cos \omega t] + [(H_1^2 \cos 2\omega t)/2].$$

This extension is directly converted into an electrical signal by means of the piezoelectric element directly attached to it, producing a signal proportional to H_m^2. Figure 13.5 illustrates the principle. At the AC excitation frequency ω, it can be seen that the magnitude and phase of the output voltage, for a constant AC excitation, H_1, depends on the total DC magnetic bias, $H_a + H_b$, as shown in figure 13.5 for different DC bias conditions A, B and C.

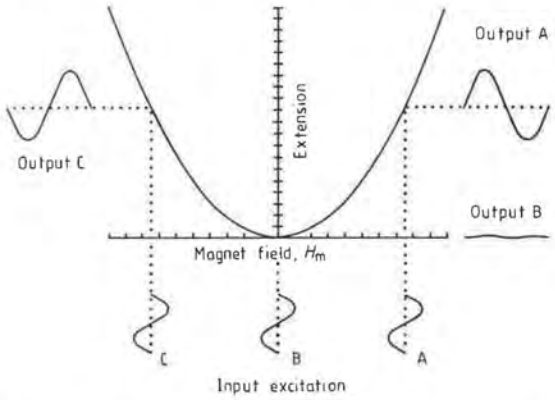

Figure 13.5 Magnetic field (H_m) combining external magnetic field (H_a) with applied DC magnetic bias (H_b) and showing output proportional to H_m obtained from the piezoelectric layer for three different DC bias conditions with superimposed AC excitation.

When the magnetometer is used to sense the direction of the earth's magnetic field, two methods of use are possible with a single element. In the open loop method no DC bias is applied; the magnetostrictive material is extended by the earth's magnetic field as the device is rotated, the AC signal output moves from the minimum at the DC bias B position in figure 13.5 towards either position A or C depending on the direction of the

external field, and there is a corresponding increase in the AC signal output at the appropriate phase. In the closed loop method, a DC bias field is used to cancel the earth's magnetic field so that $H_b = H_a$. The device is kept in this condition while it is rotated by feedback from the output to maintain zero output. The level of the feedback signal then represents the strength of the external field.

If two orthogonal axis elements are used, then if the first element is at an angle 0 to the maximum magnetic flux, its output is proportional to $H_a \cos\theta \cos\omega t$. If the second element is energised in antiphase ($\sin\omega t$), then its output V_2 is proportional to $H_a \sin\theta \sin\omega t$. Subtraction of these output signals gives a voltage proportional to $H_a \cos(\omega t + \theta)$ and hence θ can be determined from the phase relative to the drive signal. It is this arrangement that would probably be most suitable for use in vehicles for navigation purposes, although a three-axis system, which can provide absolute field strength and direction, may prove necessary.

The possible bearing accuracy obtainable with the device has not yet been fully established and largely depends on the methods used for signal processing, but it would appear to offer a small-size low-cost device with comparable performance to the flux gate magnetometer when used in the vehicle.

An alternative for the future, which does not suffer from the accuracy problems caused by having to detect the direction of the earth's magnetic field, is to make use of a *gyroscopic device*.

Conventional mechanical gyroscopes with high-speed rotating masses are too expensive and complex to be practical for use in mass-produced vehicles, but a vibratory gyroscope, which relies on the effect induced in a vibrating structure — usually a cylinder when it is rotated — is a possible, relatively low-cost alternative. GEC-Ferranti have been working on such a device [3] which uses a piezoelectric (PZT) ceramic cylinder on which electrodes are plated, as shown by the hatched areas in figure 13.6. The

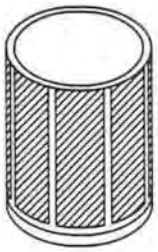

Figure 13.6 Piezoelectric vibratory gyro, showing the vibrating cylinder made from piezoelectric material. Plated electrodes are shown by the hatched areas (from [3]).

cylinder can be excited into oscillation by the application of appropriate voltages to alternate electrodes and the static vibration mode is shown by the full curves of figure 13.7. If the cylinder is then rotated, rotation-induced secondary modes appear as shown by the broken curves, their relative amplitude being a function of rotation rate. The voltages generated in the piezoelectric ceramic cylinder by these secondary modes can be picked off by the electrodes between those being used to induce the oscillation and show a direct relationship to rotation rate; for greater accuracy a feedback signal can be developed to cancel these rotation-induced secondary modes, the feedback signal then providing a measurement of high angular rotation rates up to 5000° s^{-1}.

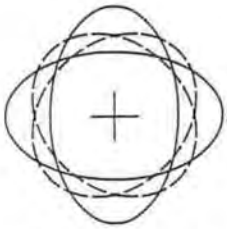

Figure 13.7 These cylindrical gyroscope modes show a primary vibration mode (full ellipses) and a secondary rotation-induced vibration mode which is an oscillation between the two broken ellipses.

The device is aimed particularly at suspension control systems rather than navigation; however, if the accuracy could be made sufficient, it might be possible to use it for both functions.

Another approach is to use a solid state version of the optical gyroscope using optical fibres. This has the promise of a low-cost device of sufficient sensitivity and fast enough response for use in vehicle navigation.

The *optical gyroscope* works on the principle shown in figure 13.8(*a*) and is known as the Sagnac interferometer. In this a narrow frequency band optical source, normally a laser, projects light onto a half-silvered double-sided mirror, A. This splits the light into two separate paths x and y which are then reflected in opposite directions round 360° by the mirrors B, C and D. At mirror A they recombine and are reflected/transmitted to a detector. When the system is static the time for the light to traverse the two equal-length light paths is the same and the returning signals are perfectly in phase; if, however, the system is rotated the transit time of the light in one direction is increased by a very small amount and the transit time in the opposite direction decreased by a very small amount. It is

possible to detect the destructive interference caused by this rotation and the consequent phaseshift, and hence obtain a measure of the rotation rate.

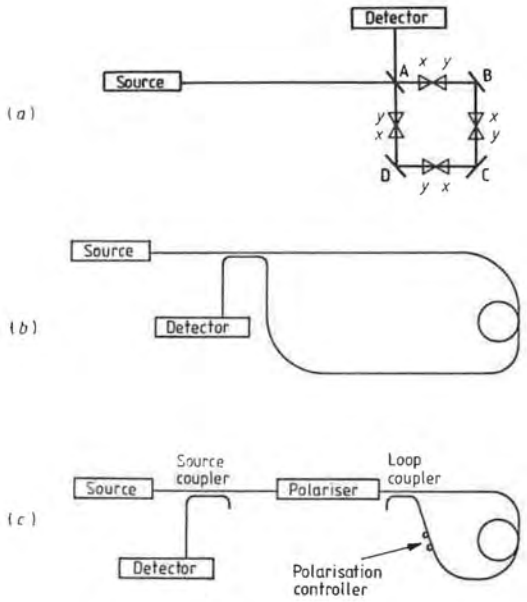

Figure 13.8 (*a*) Sagnac interferometer using mirrors and a beam splitter. (*b*) Sagnac interferometer using optical fibre. (*c*) Practical optical fibre gyroscope configuration.

With a single interferometer as described the sensitivity is very low, but it can be increased by substituting a multiturn optical fibre for the arrangement of mirrors as shown in figure 13.8(*b*). It is of course necessary to couple the optical fibres at the start and finish of the coil to obtain the appropriate transfer of optical energy both as the light enters from the source and as it returns to the detector. The phaseshift between the two light paths is multiplied directly by the number of turns on the optical fibre coil.

A more practical form of this fibre optic gyroscope has been realised using the arrangement shown in figures 13.8(*c*) and 13.9, in which a polariser is incorporated to reduce the scatter caused by interference between the different planes of polarisation of the laser light. It has proved possible to realise the parts of the system within the broken line

box shown in figure 13.9 using integrated optics etched from a silicon wafer, this improves the control over the optical path lengths and makes possible independent modulation of the optical energy in the two paths.

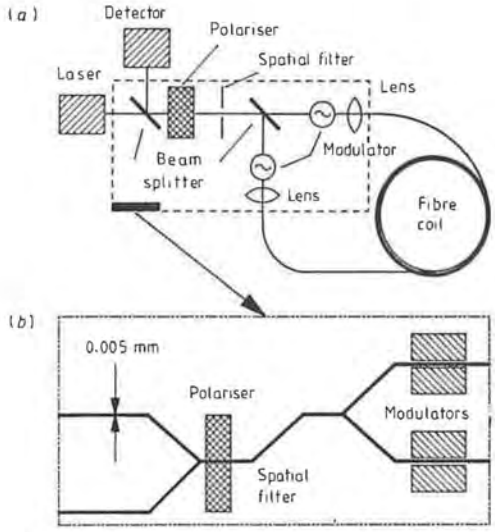

Figure 13.9 (*a*) Bulk optical fibre gyroscope. (*b*) Equivalent structure with integrated optics.

Hitachi in Japan have been working on a low-cost automotive gyroscope thought to be using the above principles, and a vehicle navigation system using a gyroscope to replace a flux gate magnetometer has been developed by Mazda. However, at the time of writing all other navigation systems use the flux gate sensor and we will now consider how it is used in a complete system.

The information obtained from the flux gate directional sensor and the wheel sensors are combined in a position-finding device which when integrated with a navigation system provides dead-reckoning navigation position and direction information relative to the point at which the system started or was reset. This is the arrangement used in infrastructure-based systems to navigate between reset points. The information can also be used in an autonomous navigation system such as the Bosch Travelpilot or Philips CARIN to compare the distance and direction travelled continuously with map information shared within the navigation system, usually on a compact disc. This map-following technique makes it possible to relate turns made by the vehicle to turns in

the road on the recorded map and this information can in turn be used to reset the dead-reckoning system to the coordinates of that particular piece of road. It is on this principle that most autonomous vehicle navigation systems work.

13.1.2 Communication Systems

An alternative approach to that discussed in the previous section is to use an infrastructure which is able to communicate with the vehicle at intervals to reset the dead-reckoning system. This is the principle used in the Autoguide/ALI-SCOUT (APPLE) system.

If the communication is made two-way, so that the infrastructure and the vehicle can exchange messages, then much more detailed information than just position can be given, for example alternative routes, traffic hold-ups, parking information and so on. However, to achieve this it is necessary to find a medium of communication which can be used for a number of vehicles at a junction simultaneously, but is of short enough range not to cause interference.

In Japan use has been made of microwave radio, whereas in Europe infrared communication has been preferred, and one of the reasons for this is that if more than one microwave beacon of the same frequency is required at a junction to avoid 'shadowing' by large vehicles of the vehicle receiving or transmitting information, then problems with interference and cancellation due to the coherent nature of the radiation may occur. In the case of infrared radiation, incoherent radiation is transmitted and cannot interfere in the same way. Information is transmitted on the infrared link by pulsing all the beacons at a particular junction simultaneously, infrared emitting diodes being used as transmitting devices. The pulsing helps in reducing power dissipation and makes possible relatively high emitted power levels. Reception is by means of an infrared-sensitive photodiode matched to the infrared frequency being used.

The full system arrangement for communication and signal processing [4] is shown in figure 13.10 which shows the Siemens ALI-SCOUT/GEC Autoguide (APPLE) system (now known as EUROSCOUT).

To ensure the best fault-free communication in both directions between the vehicle and the beacons a protocol for pulsing the infrared transmitters known as High Level Data Link Control (HDLC) is used. The pulse signals this uses are shown in figure 13.11, and as can be seen a binary code is used in which the HDLC protocol output retains its previous status when a '1' is output and changes its signal state only when a '0' appears. To synchronise the receiver a '0' is introduced automatically after five consecutive '1' states. The pulsed infrared flashes are emitted only on the transitions from '0' to '1' and '1' to '0'.

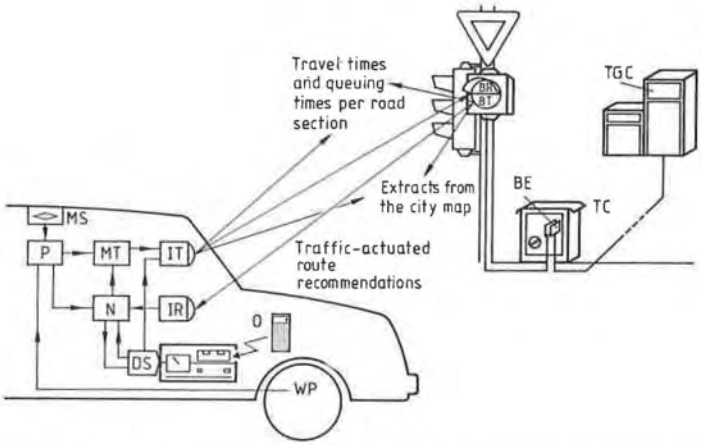

Figure 13.10 Siemens EUROSCOUT infrastructure-based route guidance and information system. P, position-finding device with magnetic field sensor MS and wheel pulser WP; N, navigation device; O, operation board with input keyboard direction indicator and direction store DS; MT, measuring device for travel time; IT, infrared transmitter; IR, infrared receiver; TGC, traffic guidance computer; TC, traffic signal controller; BE, beacon electronic device; BR, infrared-beacon receiver; BT, infrared-beacon transmitter.

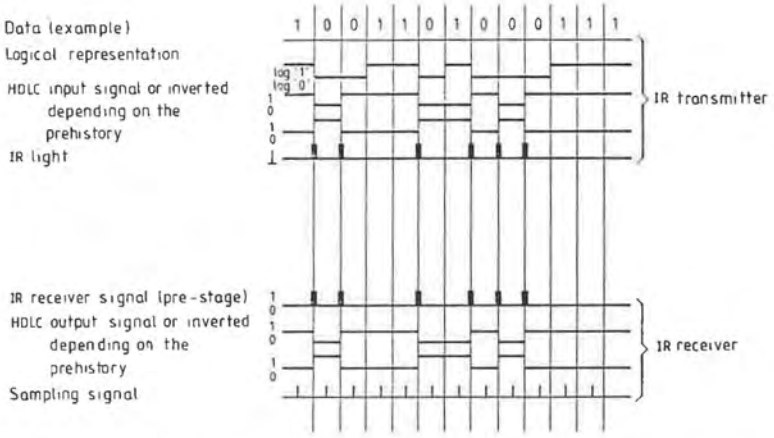

Figure 13.11 Data transmission between beacons and vehicle using modulated infrared transmission.

The system requires the synchronisation of the transmitter and receiver, and this is achieved by controlling the receive clock from the receive data by the use of a digital phase lock loop (DPLL) circuit which operates with a clock signal frequency 32 times greater than the data transmission rate. This provides a sampling pulse located in the centre of the receive data bit elements as shown in figure 13.12 with a maximum deviation of only 1/32 of the width of the bit elements due to deviations between the frequencies of the transmitter and receiver crystals.

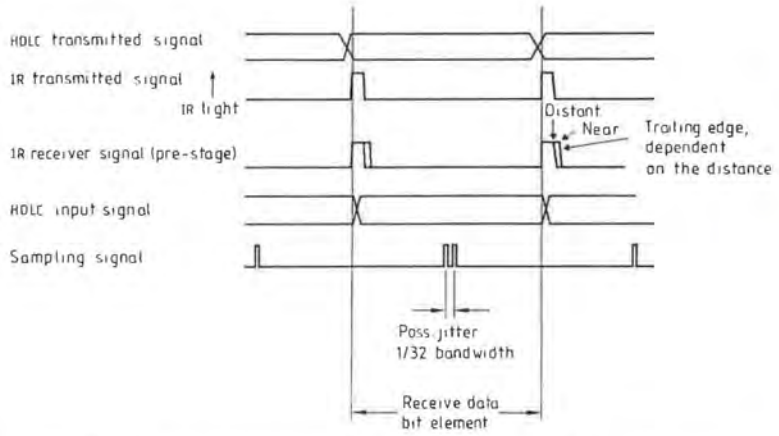

Figure 13.12 Bit synchronisation during data transmission between beacon and vehicles.

The information from the beacons to the vehicles, currently transmitted at 10 kbytes, is split into 40 blocks of 256 bytes. In the case of information from the vehicles to the beacons, this is transmitted in blocks of 128 bytes during the beacons' pauses in transmission. With this system every vehicle can make a number of attempts to transmit its information at each beacon with a probability of success of 99.4% after four attempts. Higher data rates up to 64 kbytes at transmission rates of 500 kbytes are possible for future applications, but the system as currently designed has shown high capability in field trials in Berlin.

The infrared sensor has its maximum sensitivity in the wavelength range 830 mm to 940 mm and matches the output of the infrared-emitting diode used in the transmitter. This wavelength is less affected by weather than visible light, although because the car transmitter/receiver is normally mounted on the inside of the windscreen, there can be attenuation by dirt and frost and the use of heat-radiation-reducing windscreens. However, these problems seem to have been overcome by a suitable increase in the

sensitivity of the system and the use of low-cost optical components to concentrate both the radiation and the sensor sensitivity in the direction required.

Thus sensors play a critical part in making navigation/route guidance systems practical.

13.2 COLLISION AVOIDANCE AND AUTONOMOUS DRIVER WARNING SYSTEMS

An area which is only just beginning to receive the attention it deserves is that of collision avoidance and driver warning.

A well known curve (see figure 13.13) relates the reduction in frequency of collisions to the increase in driver anticipation time of the collision situation. It indicates that if a collision can be anticipated 1 s earlier than by the unaided driver then 90% of the collisions can be avoided at intersections and in unidirectional traffic and 65% can be avoided in the case of two-directional traffic. It is this increase in the anticipation time achieved by automatic driver warning of dangerous situations that has prompted a significant part of the Europe-wide PROMETHEUS precompetitive research programme. The situations which should trigger a driver warning range from a straightforward obstruction in the road, perhaps a large object which has fallen from a truck, to the warning of a fast car approaching when overtaking is about to take place or of an accident which has just taken place in front.

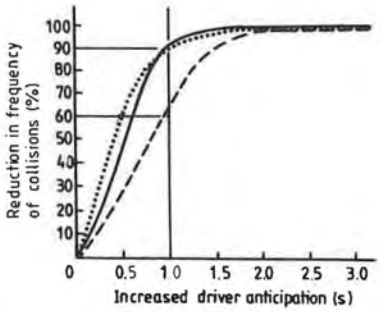

Figure 13.13 Reduction in frequency of collisions at intersections (full curve), in two-directional traffic (broken curve) and in one-directional traffic (dotted curve) with earlier driver anticipation.

13.2.1 Obstacle Detection

To cover many of the most likely collision situations requires the capability to detect — particularly under adverse weather conditions — objects in front of the vehicle and in the same road lane which are either stopped or travelling at a significantly lower speed, or to detect vehicles which are in an oncoming or passing situation and are travelling at substantially different speeds to the warned vehicle.

The first of these situations requires some form of vehicle radar, probably combined with an infrared camera sensor to make it possible to relate the information obtained from both sources by intelligent electronics — perhaps using neural nets — to give a driver warning at a high level of reliability.

The second condition requires a different approach, in which other vehicles are effectively used as remote sensors of the speed and direction of adjacent vehicles, this information then being transmitted to all the vehicles in the vicinity by short-range high-frequency radio. In this second case a vehicle might transmit messages which effectively mean 'I have just been passed by another vehicle travelling at 150 $km\,h^{-1}$ in the outside lane', so warning other vehicles in front in the inside lane not to pull out to pass at that time. The use of vehicles as remote sensors makes possible the provision to following vehicles, by successive transmission of the information, of accident warnings automatically triggered from a crash sensor when a vehicle is involved in an accident.

For short-range back-up detection of the presence of objects or other vehicles, radar, infrared or *ultrasonic* methods of sensing have been used. The transmitter/receiver for a typical ultrasonic system with a range of about 3.5 m for people and about twice this for a 0.5×0.5 m object, is shown in figure 13.14. For use for back-up obstacle detection the transmitter/receiver unit is mounted on the rear bumper at a suitable angle and the transmitter unit is pulsed with a short-frequency burst. This ultrasonic burst is directed over a wide angle to behind the vehicle where it will be reflected by any obstruction and the reflected burst picked up by the ultrasonic sensor/receiver. The pulse envelope, after detection, is then compared with the pulse envelope of the transmitted signal, the difference in timing giving the distance of the obstruction. Conversion of the timing difference to a voltage by integration makes it possible to derive the distance of the obstruction as a simple analogue or digital signal, display it to the driver and to sound a warning buzzer.

Because of the relatively long wavelengths used, very small objects, or objects with edge angles at or near 45° from the direction of sensing, can be difficult to detect. However, the system is simple and of relatively low cost.

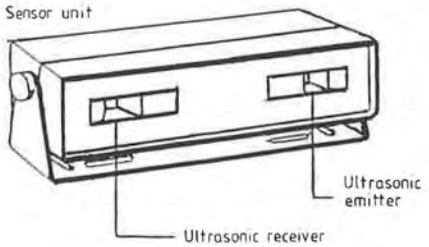

Figure 13.14 Transmitter/receiver for back-up obstacle detection.

Another application of ultrasonics, in this case to blind-spot sensing, has been described [5]. (The blind spot is the area to the side and rear of the car and not within the viewing angle of the internal and external mirrors (see figure 13.15).) Use was made in this experimental system of a low-cost ultrasonic ranging device developed for automatic camera focusing. A device of this type is capable of detecting the distance of objects from 0.3 to 10 m with a resolution of about 30 mm. Operation is similar in principle to the back-up device already described. A short burst of ultrasound is transmitted and the delay of the resulting echo measured. (Continuous-wave systems which measure the phase change between transmitted and received signals have also been used but the range is limited since phase changes of more than about 70° become much more difficult to measure by simple means.)

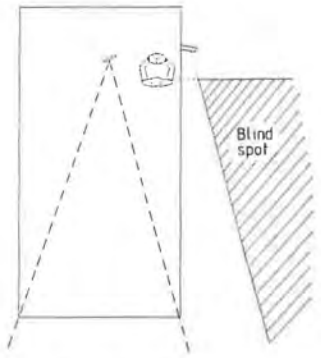

Figure 13.15 Area of blind spot not covered by either external or internal mirror field of view.

This system requires two transducers mounted at the rear of the vehicle on the driver's side and they must be arranged so that both the presence of a vehicle and its direction of travel is detected. This is done by measuring the interval between detection of the vehicle by each of the two devices.

The transducer in the system acts as both transmitter and receiver. In this transducer a foil is stretched over a grooved plate to form a capacitor which when charged and discharged in the transmitting mode electrostatically deflects the foil and produces acoustic energy from electrical energy. In the receiving mode the returned energy deflects the foil and this deflection is measured by measuring the consequent capacitance change. In normal operation as a camera ranging device, the transducer is pulsed with a 1 ms burst of 56 pulses and then becomes a microphone to detect the return burst. Because of the 20° radiation angle of these transducers it is necessary to operate them alternately so that each transducer receives only the return signal from its own transmission. It is also necessary to pulse them at a rate fast enough to permit detection of the shortest vehicles at the highest speeds likely to be encountered, but since this also reduces the detection range by limiting the time available for the received pulse to be returned a compromise is necessary. For a range of 2.5 m (which was determined in practice to be a suitable detection distance), the frequency of pulsing had to be not more than 20 Hz, but this meant that under certain circumstances fast motorcycles were not always detected.

An alternative solution to blind-spot monitoring is to use an infrared system. Using this medium transit time is not a problem. An experimental passive system has been described [6] which uses two infrared detectors mounted at the front and rear of the vehicle on the driver's side and detects the infrared radiation from the warm engine compartment. The time between detection by rear and front detectors and the relative speeds of vehicles travelling in the same and opposite directions ensures that only the vehicles travelling in the same direction trigger the warning.

An active infrared system developed in the USA, and which has reached the production stage, has also been described, although little technical information is available. The system makes use of 21 pulsed infrared beams in four sensor modules mounted on the rear half of the vehicle which scan the 'blind zone' for the presence of vehicles out to 2.5 m.

For longer-range obstacle detection it is necessary to use either *radar* or optical methods. Optical methods may use visible or infrared laser rangefinders and/or scanning television systems to determine lateral obstacle spacing and distance. Generally infrared systems require some cooling for adequate sensitivity and may require mechanical scanning. Costs seem unlikely to become low enough to enable realistic systems to be installed in production vehicles. Systems working in the visible spectrum, however, have potential for low-cost production but do not

work effectively in periods of atmospheric obscuration or in darkness, when it can be argued that obstacle detection is most needed. It seems then that millimetric wave radar is likely to be the most effective medium for all-weather operation.

A low-cost 94 GHz system developed by Philips has been described [7] and is considered to be suitable for use in vehicles. It uses a frequency-modulated continuous-wave (FMCW) technique which involves the transmission of a continuous frequency-modulated signal rather than the short pulses normally used. The frequency is modulated by a switch at 1 kHz. The time delay between the transmitted and reflected signals then produces a frequency difference between the transmitted and received signal which is a function of the target range, the relationship being given by $f = 2\ ar/c$ where f is the frequency, a the frequency sweep rate, r the target range and c the speed of light. This technique makes it possible to achieve an accuracy of about 2 m at ranges up to 150 m using a 2°, 10 mW beam. The development has been particularly aimed at use in an intelligent cruise control system which automatically controls the speed of the vehicle so as to follow the vehicle in front at a safe distance and at a matched speed by controlling the acceleration and braking of the vehicle according to the speed and distance measured.

Low cost has been achieved by the use of injection-moulded plastic with a flash coating of gold metallisation to make the waveguide components. These components are made in two halves with the waveguides formed by mating grooves in each half. Active components such as Gunn diode oscillators and gallium arsenide Schottky diode mixers are suspended in the waveguides by pieces of dielectric.

The major problem with vehicle radar is overcoming false reflections from vehicles in different road lanes or from road furniture or bridge pillars. This is a particular problem on curves, but is likely to be much less of a problem when used in an intelligent cruise control system than in general obstacle detection because of the shorter distances involved. Recognition of another vehicle could of course be greatly eased if all vehicles were equipped with a small corner reflector on the rear of the vehicle, but without legislation this is unlikely ever to be possible.

Infrastructure-based driver warning systems are as old as motoring. Permanent or temporary roadside signs giving instructions or information to the driver have been used for most of the motoring era. Apart from traffic lights, it is only in the last 20 years that roadsigns have become active with the ability to be changed remotely to fit traffic circumstances. The police-controlled signs on motorways are typical examples. Their usefulness is closely related to the accuracy of the information input and the speed with which changes are made. We are all aware that there can be considerable delays in changing signs after an accident or obstruction has happened and an even longer delay after it has been cleared.

The most recent development in this area is the appearance of infrastructure-based systems which can be accessed by electronics within the vehicle to give direct information to the driver. An example of this is the Trafficmaster system which is currently operational on the M25 motorway around London and is scheduled to be extended to other UK motorways; it is also under review for use in the USA. It makes use of bridge-mounted detectors using active infrared technology to illuminate two strips across each outside lane of a motorway with coded infrared radiation through which vehicles must pass (see figure 13.16). The reflected infrared radiation is detected and processed by a bridge-mounted microprocessor to determine the speed of the vehicles passing underneath. If the vehicle speed drops below 40 km h^{-1} a signal is sent to a control via the Paknet radio-accessed packet data network. The data showing sensor location, speed on both carriageways and vehicle count are then sent to the subscriber's vehicle by means of the Air Call radio-paying network and shown on an LCD map display to give a complete picture of the motorway network with an indication of where traffic is slow moving or stationary.

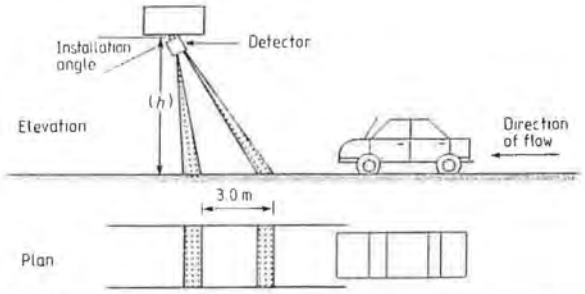

Figure 13.16 Infrared transmitters and receivers used in the Trafficmaster system to sense the speed of passing motorway vehicles.

The information obtained from a system of this type could clearly be used to control road signs and could be used as a real-time input to a vehicle navigation system, which could then provide alternative route information when congestion was indicated.

The ability to detect congestion, accidents and other road obstructions is further extended by the use of each vehicle as a sensor of the conditions existing at its location with the capability of automatically communicating that information by a suitable microwave radio link both to other vehicles and to the infrastructure for wider distribution. The development of the sensors and systems which make this possible is one of the prime concerns of the European PROMETHEUS and DRIVE research programmes.

REFERENCES

[1] Bornhofft W and Trenkler G 1989 Magnetic field sensors: flux gate sensors *Sensors — A Comprehensive Survey* vol 5 (VCH) ch 5
[2] Lynch B J and Gallantree H R 1990 A new magnetic sensor technology *GEC J. Res.* **8** 13–20
[3] Nuttall J D 1991 The development of a solid-state gyroscope *IMechE Conf. on Technology of Inertial Sensors and Systems, April 1991*
[4] von Tomkewitsch R 1990 *Dynamic route guidance and interactive transport management with infra-red beacons* private communication from Siemens Plessey Controls Ltd, November 1990
[5] McWilliam S W 1988–9 Automotive blind spot monitor *Final Year Project Report* Dept of Mechanical Engineering, University of Southampton
[6] Wai W O 1987 Blind spot monitor *IEE/IMechE 6th Int. Conf. on Automotive Electronics* (*IEE Conf. Publication 280*) pp 51–3
[7] Shelley T 1990 Low cost radar reads the road ahead *Eureka* (August) 26–8

14

Future Developments

Future developments expected in the field of automotive sensors are concerned particularly with making sensors more reliable, of higher quality, easier to use, cheaper to make and more accurate; this will be achieved by various means such as the use of improved materials, better processing and greater use of smart electronics to compensate for the unavoidable inaccuracies in the different devices. In particular, the use of silicon-based sensors is expected to increase dramatically, and the use of such devices with added intelligence will be widespread within the next few years.

There are three areas of general development in the use of sensors which are going to be of most importance, the first two in the 1990s and the third after the turn of the century. These areas are:

(i) integrated system sensors;
(ii) self-calibrating sensors;
(iii) embedded simulation.

14.1 INTEGRATED SYSTEM SENSORS

Up to the present automotive sensors have been 'stand-alone' devices providing a specific output to an electronic control unit in which the information from the sensor was then processed together with that from other sensors to control a system in the vehicle. If more than one system needs to use the same information, for example when throttle position and movement is required by the engine control system and the suspension system, it has generally been the practice to use two separate sensors measuring the same physical quantity. System designers have been reluctant to feed the signal from a common sensor into their electronics and allow the second system requiring the information to take a parallel

signal from the sensor, because of the difficulty of determining which system is responsible in a fault situation and the risk of increased electromagnetic interference (EMI) between systems. However, with the advent of in-car multiplex wiring systems, each control system can be independently fed with coded signals from a single sensor for each parameter of interest.

The original concept of multiplex wiring was to use a single, large gauge ring circuit to distribute electrical current to all parts of the vehicle (see figure 14.1), then at each take-off point where current is required (for rear lamps, stop lamps, direction flashers, heated backlight etc), there is an integrated circuit control module which includes an electronic switch capable of controlling the current to the component, plus a sensing circuit module that can detect and act on a digital control information code passed down a separate control wire. The dashboard carries a low-current switch which incorporates a code pulse generator and can be as physically small as required by the car designer.

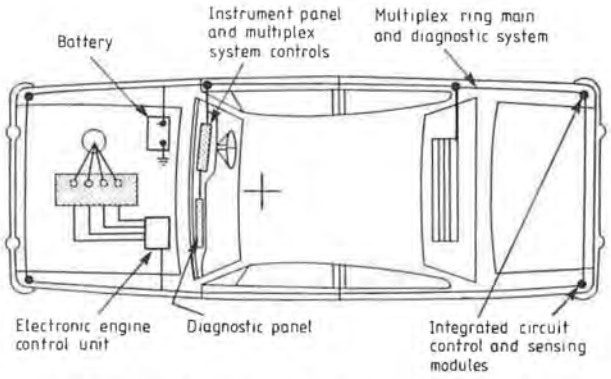

Figure 14.1 A multiplex wiring ring circuit.

The sensing circuit also has the capability to transmit diagnostic information back down the control wire indicating whether the component to be switched on is in a normal state, short circuited or open circuit. Thus, for example, a diagnostic display of bulb condition can be easily provided without additional wiring or sensors.

It is also possible to transmit sensor information on the multiplex link from one part of the vehicle to another. For example, information about engine speed may be needed by a number of the control systems and this can be made available to all systems on the multiplex link. By the use of very high-frequency data codes on the control link it is even possible to

interconnect operating control systems to provide complex real-time interactive control functions.

Recent multiplex systems have moved away from the single ring main around the vehicle and all systems now in use or development split the vehicle into zones fed on a 'star' network by power cables and data links, so that failure in one zone can be isolated from the rest of the system.

Early multiplex systems made use of simple repetitive digital control codes (protocols). However, all recent systems make use of what is known as a 'contention' protocol. In this, any remote logic circuit module can generate a digital 'word' describing the information or instruction to be transferred at any time, but each source has a predetermined priority, so that if two or more messages coincide, whichever message has the highest priority is allowed to proceed to its destination and the lower-priority message source has to try again.

The Bosch CAN system in Europe and the SAE J1850 system in the USA are examples of multiplex systems which use contention protocols. In these systems, it is only necessary to send the control instruction once and have it acknowledged to switch a particular load on or off or to send a particular piece of information. The transmission of sensor information is therefore possible provided that each sensor has associated with it a suitable electronic module to represent its (usually) analogue signal at a sufficient number of levels of that signal to give the required accuracy and to convert those levels to a series of binary digits which can be electronically incorporated into the full digital word required by the multiplex system. This full word will include an indication of the priority of the input from each sensor, so that changes in output can rapidly be transmitted to the control system making use of the information. For example a change in throttle position would have very high priority as an input signal to both an engine control system and a suspension control system, whereas a change in engine temperature, or fuel level, because they can only change comparatively slowly, would not require a high-priority designation in the multiplex system.

It is essential that the pulse repetition rate of the multiplex system is sufficiently high to ensure that rapid changes in sensor input are transmitted to the control system in a time which is short compared with the response rate of the system being controlled. This requires pulse repetition rates in the 50–200 kHz range for the transmission of sensor data, and pulse repetition rates up to 1 MHz if the sensor is included in a real-time control loop and system instability is to be avoided. This requirement has led to the division of multiplex systems into three different ranges:

(i) Class A. Low-frequency pulse repetition rates of 5–30 kHz, suitable for basic switching functions in the vehicle.

(ii) Class B. Medium-frequency pulse repetition rates of 50–200 kHz, suitable for transmission of sensor and other real-time data.

(iii) Class C. High-frequency pulse repetition rates of 500 kHz–2 MHz, suitable for the inclusion of sensors and actuators within the control system direct feedback loop.

In the multiplex systems expected to be used during the 1990s, some combination of Class A and B functions, possibly in a two-level multiplex system, seems most probable. In the longer term Class C multiplex will be used to produce an integrated total vehicle system which can be looked upon as a complete vehicle control system, with all systems being fully interactive.

The use of optical fibres instead of 'twisted pair' copper cables for multiplex interconnection seems a strong possibility after the turn of the century and would lead to further developments in the multiplex protocol to take advantage of the much greater bandwidth which would become available. This technology requires that there is a multiplex module for each inlet and outlet port to the system. So that each sensor has its own electronic module. This arrangement is a halfway-house to the 'smart' sensors described in chapter 12, where the electronics instead of being in a separate module is integrated onto the same substrate or even the same chip as the sensor itself.

As was said in chapter 12, very few sensors have so far been developed to this level since it is necessary that the sensor element itself should be suitable for integration; that is, it is either a semiconductor sensor such as a Hall effect magnetic field probe, or a substrate sensor such as a thick-film strain gauge. However, this is an area in which there is considerable development, and when these smart sensors are fully developed they will include not only the standardised coding/decoding functions required for multiplex wiring but also electronics with the capability for processing the sensor information to provide checks on operation within range, repeatability and linearity. This capability leads on to a second long-term development likely in automotive sensors, that of self-calibration.

14.2 SELF-CALIBRATING SENSORS

The ability of a sensor to be self-calibrating is likely to be a potent means of reducing sensor costs in future.

Many low-cost sensors suffer from poor linearity over their operating range, although their repeatability is good. The fuel level sensor is an excellent illustration of this, for in spite of careful calibration during development, the inherently low cost of the sensor means that only very simple methods of linearity correction are possible, with the consequence

that significant nonlinearity exists under operational conditions. This inaccuracy has been compounded by the susceptibility of the fuel tank sensor to distortion of the float arm during assembly and the small depth of modern fuel tanks, with the result that fuel level readings in vehicles are a constant source of customer complaint, although the repeatability of the reading in each individual vehicle is good.

In a self-calibrating system the system would be cycled under controlled conditions after assembly of the sensor into the fuel tank; this would be done by filling and emptying the tank by known increments, while the sensor smart electronics was instructed, through the multiplex controller, to record the readings at each increment of filling and to use these recorded data as the calibration against which all future measurements are read.

Providing the sensor has good repeatability, then wide variations in linearity and range between nominally similar sensors can be accepted, giving the opportunity for the increasing use of low-cost devices.

The third and longer-term development which in the authors' view will, in the early part of the next century, cause dramatic changes in the comprehensive vehicle control systems that will then be in use, is what we call 'embedded simulation'.

14.3 EMBEDDED SIMULATION

The development of computer simulation of vehicle systems such as the engine, transmission, suspension etc is currently proceeding apace, and further developments of the technique will involve combining those simulations to make it possible to represent the complete vehicle. Currently this requires substantial computing power and memory, but it seems certain that by early in the next century computing power of this order will be available in low-cost small-package devices, sufficiently compact to fit into a vehicle.

At this point it seems probable that control systems for vehicles (and probably other self-contained real-time control systems around the home and in industry) will change in character, so that a full simulation of the system being controlled is embedded in the control system. With the availability of this simulation within the control system it will be possible to compare the actual performance of the vehicle, as measured by suitable sensors, with an ideal as specified by the embedded simulation. This not only provides the opportunity for comprehensive feedback control of all controllable functions, but also provides a target against which the performance of all these functions can be compared, hence making fully active diagnostics possible.

Another benefit of such a system is the ability to continue operating the

vehicle satisfactorily even when failure of major parts of the main control system, in particular sensors, occurs. In fact one benefit might well be the ability to dispense with many of the existing sensors, since given information on for example only engine speed, torque and temperature, the simulation — particularly if it is designed to be adaptive — may well be able to specify the full operating conditions of the engine. Then, as major slowly varying conditions such as wear, fuelling or altitude of operation change, the simulation is automatically modified to take this into account.

An example of how this system might work in practice is shown in the diagram of figure 14.2. In this a comprehensive vehicle control system is shown in which three major control units are used for power train, chassis systems and body/information systems. These control units are connected to a high-speed multiplex data link and are capable of operating together through the multiplex link with their associated sensors and actuators as direct real-time feedback control systems for their designated functions.

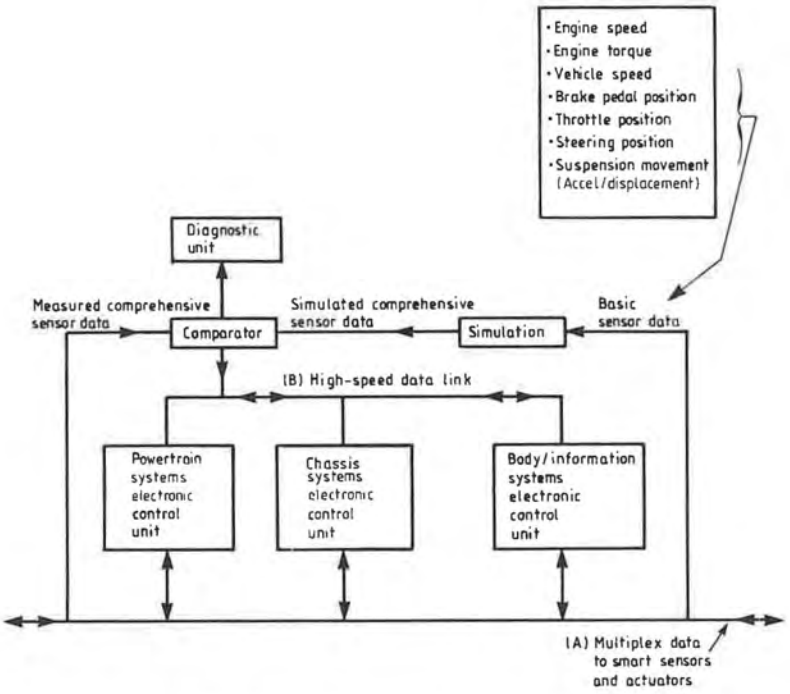

Figure 14.2 A comprehensive embedded simulation system.

The three control systems are also directly interconnected by a high-speed data link, along which sensor data can be rapidly exchanged. This data link is also fed with data which can modify the sensor data under the control of the simulation/comparator system shown at the top of the diagram.

The simulation/comparator system contains a full simulation of the vehicle systems, and is fed with basic sensor data, the suggested list of parameters required being shown. These inputs control the operating conditions and parameters of the simulation which then transmits the resulting comprehensive sensor data for an accurate simulation to the comparator where it is compared with the measured comprehensive sensor data from the multiplex data link (A). Any differences are transmitted via data link (B) to the three system control units and weighted according to their importance and priority and then used to modify the control loops to correct the system towards the optimum as defined by the simulation. The output of the comparator is also transmitted to a diagnostic unit which can assess the importance of the deviations from the simulated optimum sensor readings and flag substantial deviations which could indicate a serious malfunction in either sensors or actuators.

In the case of a gross disruption of sensor information, the simulation could be used to reconstitute sensor signals to keep the vehicle running under the best practicable conditions, consistent with the extent and severity of the failure.

In the longer term, there is the possibility of reducing the number of sensors to those basic parameters listed on the diagram, and relying on the simulation to reconstitute the comprehensive sensor data required for full system control, with a corresponding dramatic reduction in cost and improvement in reliability.

This technique of embedded simulation is of course applicable to a whole range of complex control systems, not only in automotive applications but in many other areas where complex real-time control is essential.

14.4 THE CAR OF THE FUTURE

The car of the future seems likely to be comprehensively controlled and optimised by sophisticated real-time controls so that it is intelligent enough to both compensate for minor aging and deterioration as well as being able to warn the driver — and through the external driver

communication network the nearest repair and rescue services — of major failures or accidents.

In many ways, the vehicle system can be seen to approach the way in which human beings operate as a complex control system [1], since a human being uses all the techniques we have described to operate effectively on a day to day basis. A person requires a comprehensive sensor set to enable them to operate, just as does a motor vehicle. In fact cutting either system off entirely from their sensors is a sure way of obtaining rapid and total system failure!

The combination of sophisticated internal controls with the ability to communicate with the outside world through the driver communication, collision avoidance, vision enhancement and navigation systems described in chapter 13, makes the car of the future part of a complex and sophisticated transportation system which will optimise road travel while retaining the individual control of destination, speed and direction that people consider to be of great importance. The sensors described in this book are essential in making all this possible.

REFERENCE

[1] Westbrook M H 1984 The human body and its transducers: a real time, real life computing and control system *Proc. IEE* **131** A 10–16

Index