AUTOMOTIVE ELECTRONICS HANDBOOK

Related McGraw-Hill Books of Interest

Handbooks

Avallone and Baumeister • MARK'S STANDARD HANDBOOK FOR MECHANICAL ENGINEERS
Benson • AUDIO ENGINEERING HANDBOOK
Brady • MATERIALS HANDBOOK
Chen • COMPUTER ENGINEERING HANDBOOK
Considine • PROCESS/INDUSTRIAL INSTRUMENTS AND CONTROL HANDBOOK
Coombs • PRINTED CIRCUITS HANDBOOK
Coombs • ELECTRONIC INSTRUMENT HANDBOOK
Di Giacomo • DIGITAL BUS HANDBOOK
Fink and Beaty • STANDARD HANDBOOK FOR ELECTRICAL ENGINEERS
Fink and Christiansen • ELECTRONICS ENGINEERS' HANDBOOK
Ganic • MCGRAW-HILL HANDBOOK OF ESSENTIAL ENGINEERING INFORMATION
Harper • ELECTRONIC PACKAGING AND INTERCONNECTION HANDBOOK
Harper and Sampson • ELECTRONIC MATERIALS AND PROCESSES HANDBOOK
Hicks • STANDARD HANDBOOK OF ENGINEERING CALCULATIONS
Hodson • MAYNARD'S INDUSTRIAL ENGINEERING HANDBOOK
Johnson • ANTENNA ENGINEERING HANDBOOK
Juran and Gryna • JURAN'S QUALITY CONTROL HANDBOOK
Kaufman and Seidman • HANDBOOK OF ELECTRONICS CALCULATIONS
Lenk • MCGRAW-HILL ELECTRONIC TESTING HANDBOOK
Lenk • LENK'S DIGITAL HANDBOOK
Mason • SWITCH ENGINEERING HANDBOOK
Schwartz • COMPOSITE MATERIALS HANDBOOK
Townsend • DUDLEY'S GEAR HANDBOOK
Tuma • ENGINEERING MATHEMATICS HANDBOOK
Waynant • ELECTRO-OPTICS HANDBOOK
Woodson • HUMAN FACTORS DESIGN HANDBOOK

Other

Boswell • SUBCONTRACTING ELECTRONICS
Gieck • ENGINEERING FORMULAS
Ginsberg • PRINTED CIRCUIT BOARD DESIGN
Johnson • ISO 9000
Lenk • MCGRAW-HILL CIRCUIT ENCYCLOPEDIA AND TROUBLESHOOTING GUIDE, VOLS. 1 AND 2
Lubben • JUST-IN-TIME MANUFACTURING
Markus and Sclater • MCGRAW-HILL ELECTRONICS DICTIONARY
Saylor • TQM FIELD MANUAL
Soin • TOTAL QUALITY CONTROL ESSENTIALS
Whitaker • ELECTRONIC DISPLAYS
Young • ROARK'S FORMULAS FOR STRESS AND STRAIN

To order or to receive additional information on these or any other McGraw-Hill titles, please call 1-800-822-8158 in the United States. In other countries, please contact your local McGraw-Hill office. **BC14BCZ**

AUTOMOTIVE ELECTRONICS HANDBOOK

Ronald K. Jurgen Editor in Chief

McGraw-Hill, Inc.

New York San Francisco Washington, D.C. Auckland Bogotá
Caracas Lisbon London Madrid Mexico City Milan
Montreal New Delhi San Juan Singapore
Sydney Tokyo Toronto

Library of Congress Cataloging-in-Publication Data

Automotive electronics handbook / Ronald Jurgen, editor in chief.
p. cm.
Includes index.
ISBN 0-07-033189-8
1. Automobiles—Electronic equipment. I. Jurgen, Ronald K.
TL272.5.A982 1994
629.25'49—dc 94-39724
CIP

6 7 8 9 10 FGRFGR 9 9 8

ISBN 0-07-033189-8

The sponsoring editor for this book was Stephen S. Chapman, the editing supervisor was Virginia Carroll, and the production supervisor was Suzanne W. B. Rapcavage. It was set in Times Roman by North Market Street Graphics.

This book is printed on acid-free paper.

This book is dedicated to Robert H. Lewis and to the memories of Douglas R. Jurgen and Marion Schappel.

CONTENTS

Part 4 Displays and Information Systems

Part 5 Safety, Convenience, Entertainment, and Other Systems

CONTRIBUTORS

Robert E. Bicking *Honeywell, Micro Switch Division* (CHAP. 4)
Tracy Blake *Arizona State University* (CHAP. 30)
David S. Boehmer *Intel Corporation* (CHAP. 11)
Werner Brehm *Robert Bosch GmbH* (CHAP. 13)
Wolfgang Bremer *Robert Bosch GmbH* (CHAP. 22)
Wolfgang Bullmer *Robert Bosch GmbH* (CHAP. 13)
Jerry L. Cage *Allied Signal, Inc.* (CHAP. 15)
Tom Chrapkiewicz *Philips Semiconductor* (CHAP. 25)
Armin Czinczel *Robert Bosch GmbH* (CHAP. 16)
Jeffrey N. Denenberg *Noise Cancellation Technologies, Inc.* (CHAP. 31)
William C. Dunn *Motorola Semiconductor Products* (CHAP. 7)
Randy Frank *Motorola Semiconductor Products* (CHAPS. 2, 5, 32)
Robert L. French *R. L. French & Associates* (CHAP. 29)
Frieder Heintz *Robert Bosch GmbH* (CHAP. 22)
Gary C. Hirschlieb *Robert Bosch GmbH* (CHAP. 12)
Raymond S. Hobbs *Arizona Public Service Company* (CHAP. 30)
Gerhard Hötzel *Robert Bosch GmbH* (CHAP. 6)
Robert Hugel *Robert Bosch GmbH* (CHAP. 22)
Ronald K. Jurgen *Editor* (CHAPS. 1, 20, 21)
George G. Karady *Arizona State Univeristy* (CHAP. 30)
Donald B. Karner *Electric Transportation Application* (CHAP. 30)
Shinichi Kato *Nissan Motor Co., Ltd.* (CHAP. 24)
Bernhard K. Mattes *Robert Bosch GmbH* (CHAP. 23)
Fred Miesterfeld *Chrysler Corporation* (CHAP. 26)
Salim Momin *Motorola Semiconductor Products* (CHAP. 32)
James P. Muccioli *JASTECH* (CHAPS. 27, 28)
Klaus Müller *Robert Bosch GmbH* (CHAP. 10)
Kurt Neuffer *Robert Bosch GmbH* (CHAP. 13)
Harald Neumann *Robert Bosch GmbH* (CHAP. 6)
Paul Nickson *Analog Devices, Inc.* (CHAP. 3)
Johann Riegel *Robert Bosch GmbH* (CHAP. 6)

Makoto Sato *Honda R&D Co., Ltd.* (CHAP. 18)
Gottfried Schiller *Robert Bosch GmbH* (CHAP. 12)
Shari Stottler *Robert Bosch GmbH* (CHAP. 12)
Richard Valentine *Motorola Inc.* (CHAPS. 14, 19)
Helmut Weyl *Robert Bosch GmbH* (CHAP. 6)
Hans-Martin Wiedenmann *Robert Bosch GmbH* (CHAP. 6)
William G. Wolber *Cummins Electronics Co., Inc.* (CHAPS. 8, 9)
Akatsu Yohsuke *Nissan Motor Co., Ltd.* (CHAP. 17)

PREFACE

Automotive electronics as we know it today encompasses a wide variety of devices and systems. Key to them all, and those yet to come, is the ability to sense and measure accurately automotive parameters. Equally important at the output is the ability to initiate control actions accurately in response to commands. In other words, sensors and actuators are the heart of any automotive electronics application. That is why they have been placed first in this handbook where they are described in technical depth. In other chapters, application-specific discussions of sensors and actuators can be found.

The importance of sensors and actuators cannot be overemphasized. The future growth of automotive electronics is arguably more dependent on sufficiently accurate and low-cost sensors and actuators than on computers, controls, displays, and other technologies. Yet it is those nonsensor, nonactuator technologies that are to many engineers the more "glamorous" and exciting areas of automotive electronics.

In the section on control systems, a key in-depth chapter deals with automotive microcontrollers. Without them, all of the controls described in the chapters that follow in that section—engine, transmission, cruise, braking, traction, suspension, steering, lighting, windshield wipers, air conditioner/heater—would not be possible. Those controls, of course, are key to car operation and they have made cars over the years more drivable, safe, and reliable.

Displays, trip computers, and on- and off-board diagnostics are described in another section, as are systems for passenger safety and convenience, antitheft, entertainment, and multiplex wiring. Displays and trip computers enable the driver to readily obtain valuable information about the car's operation and anticipated trip time. On- and off-board diagnostics have of necessity become highly sophisticated to keep up with highly sophisticated electronic controls. Passenger safety and convenience items and antitheft devices add much to the feeling of security and pleasure in owning an automobile. Entertainment products are what got automotive electronics started and they continue to be in high demand by car buyers. And multiplex wiring, off to a modest start in production cars, holds great promise for the future in reducing the cumbersome wiring harnesses presently used.

The section on electromagnetic interference and compatibility emphasizes that interference from a variety of sources, if not carefully taken into account early on, can raise havoc with what otherwise would be elegant automotive electronic designs. And automotive systems themselves, if not properly designed, can cause interference both inside and outside the automobile.

In the final section on emerging technologies, some key newer areas are presented:

- Navigation aids and intelligent vehicle-highway systems are of high interest worldwide since they hold promise to alleviate many of vehicle-caused problems and frustrations in our society.
- While it may be argued that electric vehicles are not an emerging technology, since they have been around for many years, it certainly is true that they have yet to come into their own in any really meaningful way.
- Electronic noise cancellation is getting increasing attention from automobile designers seeking an edge over their competitors.

The final chapter on future vehicle electronics is an umbrella discussion that runs the gamut of trends in future automotive electronics hardware and software. It identifies potential technology developments and trends for future systems.

Nearly every chapter contains its own glossary of terms. This approach, rather than one overall unified glossary, has the advantage of allowing terms to be defined in a more application-specific manner—in the context of the subject of each chapter. It should also be noted that there has been no attempt in this handbook to cover, except peripherally, purely mechanical and electrical devices and systems. To do so would have restricted the number of pages available for automotive electronics discussions.

Finally, the editor would like to thank all contributors to the handbook and particularly two individuals: Otto Holzinger of Robert Bosch GmbH in Stuttgart, Germany and Randy Frank of Motorola Semiconductor Products in Phoenix, Arizona. Holzinger organized the many contributions to this handbook from his company. Frank, in addition to contributing two chapters himself and cocontributing a third, organized the other contributions from Motorola. Without their help, this handbook would not have been possible.

Ronald K. Jurgen

AUTOMOTIVE ELECTRONICS HANDBOOK

P · A · R · T · 1

INTRODUCTION

CHAPTER 1
INTRODUCTION

Ronald K. Jurgen
Editor

1.1 THE DAWN OF A NEW ERA

In today's world of sophisticated automotive electronics, it is easy to forget how far the technology has come in a relatively short time. In the early 1970s, other than radios and tape players, the only standard electronic components and systems on most automobiles were alternator diodes and voltage regulators.[1] By the fall of 1974, "there were twelve electronic systems available, none of which were across the board standard production items. . . . The twelve electronic systems or subsystems were: alternator diodes, voltage regulators, electronic fuel injection, electronic controlled ignition, intermittent windshield wipers, cruise control, wheel lock control, traction control, headlamp control, climate control, digital clocks, and air bag crash sensors."[1]

1.1.1 Car Makers and the Electronics Industry: Friendly Adversaries

In the early days of automotive electronics, the automotive industry and the electronics industry were often at odds. Carmakers needed inexpensive components and systems that would operate reliably in the extremely harsh automotive environment. The electronics industry, on the other hand, used to producing high-quality but expensive parts and systems for the military, was skeptical about its ability to produce the components the automobile industry wanted at the prices they demanded. But both industries realized that electronics could provide the capability to solve automotive problems that defied conventional mechanical or electromechanical approaches.

Some of the leading electronics engineers who worked in the automotive industry—as well as their counterparts in the electronics industry—realized that this existing friendly adversarial relationship had to be converted to a mutual effort to find cost-effective and reliable solutions to urgent automotive problems.

Thus it was in 1973 that Trevor Jones (then with General Motors), Joseph Ziomek (then with Ford), Ted Schaller (Allen Bradley), Jerry Rivard (then with Bendix), Oliver McCarter (General Motors), and William Saunders (Society of Automotive Engineers), proposed that a new conference be held in 1974.[1] Dubbed Convergence to signify the coming together of the two industries, the first conference was successful and, sponsored alternately by the Society of Automotive Engineers and the Institute of Electrical and Electronics Engineers, it has been held successfully every other year ever since.

1.1.2 The United States Government Forces the Issue

One of the major problems facing the automotive industry at the time of the first Convergence conference was upcoming stricter government-mandated exhaust emissions controls. When the United States government first mandated emissions standards for all United States cars, car makers met the challenge through the use of catalytic converters for hydrocarbon and carbon monoxide emissions and exhaust gas recirculation techniques for nitrogen oxides emissions. But they knew that in 1981, when the standards would be tightened from the previous limit of 2.0 grams per mile to 1.0 gram per mile, those approaches would no longer in themselves be sufficient. A new approach was necessary and it involved use of a three-way catalyst for all three emissions together with a closed-loop, engine control system.[2]

Tighter emissions control solved one problem but created another—fuel economy. The two seemed to be mutually exclusive. Charles M. Heinen and Eldred W. Beckman, writing in *IEEE Spectrum* in 1977,[3] said, "The simple truth is that there is very direct interaction between emissions and fuel economy. Probably the clearest example of that interaction is the fact that automobiles equipped to meet California's tight emissions control regulations have consistently demonstrated about 10 percent poorer fuel economy than have comparable cars equipped to meet the less stringent Federal U.S. standards." As a result of this interdependence, emissions and fuel economy measures tended to be compromises. Greater fuel economies could be achieved if emissions levels were not a problem.

1.2 THE MICROCOMPUTER TAKES CENTER STAGE

The microcomputer, introduced in 1971, had yet to make major inroads in automobiles. But it became increasingly obvious that it was the key to meeting government exhaust emission and fuel economy demands while also providing car buyers with cars that performed well. Meeting these needs necessitated precise engine control in such areas as the air/fuel ratio and idle speed.

1.2.1 Early Applications of Microcomputers

One of the first microcomputer applications in cars was an advanced ignition system built by Delco-Remy for the 1977 Oldsmobile Toronado. Called the MISAR (microprocessed sensing and automatic regulation) system, it controlled spark timing precisely no matter what load and speed conditions prevailed while meeting emissions control requirements and providing good driveability. Input signals from sensors provided data on crankshaft position, manifold vacuum, coolant temperature, and reference timing.[4] The microprocessor used had a capacity of 10,240 bits.

Early applications such as the MISAR paved the ground for what would later become the prolific use of microcomputers in cars. Once reliable microcomputers met the cost restraints of carmakers, there was no end in sight to microcomputer applications in cars. In the late 1970s, total engine control with microcomputers became widespread and, as time went on, use of microcomputers spread to other controls for transmission, braking, traction, suspension, steering, lighting, air conditioning, and so forth.

1.2.2 The Bells and Whistles Period

There was also a time in the early 1980s when carmakers, heady with success with microcomputers in other areas, went through a period of electronic overkill. Notable in this regard were voice commands and warnings that tended to wear out their welcome quickly with car drivers

and elaborate and flashy information displays that also turned off many car buyers. It was a period of doing things with microcomputers because they could be done rather than doing them because they were needed.

That overindulgent microcomputer period quickly waned as car buyers made their feelings known. Voice commands were all but totally abandoned and displays were made less garish. There was even a return to analog displays for speedometers, for example, albeit electronically based rather than the old mechanical or electromechanical system. Carmakers returned to using microcomputers in truly functional ways to answer real needs.

1.3 LOOKING TO THE FUTURE

The future for automotive electronics is bright. Electronic solutions have proven to be reliable over time and have enabled carmakers to solve problems otherwise unsolvable. But what does the future hold? Some predictions for the future have been discussed in the following pages by contributors.

1.3.1 Contributors' Predictions

Although there have been many significant automotive electronics advances over the years, the end is certainly not in sight. The final chapter in this handbook describes many upcoming advances in detail. Authors Frank and Momin, for example, state that a likely future scenario "will be a combination of centralization and distributed intelligence where the centralization would be based along the lines of body, chassis and safety, powertrain, and audio/entertainment and communications. Within these centralized systems would be distributed intelligence based on multiplex wiring with smart sensors, switch decoders, and smart actuators all controlled by a central intelligence."

Here are additional selected future developments cited by contributors in other chapters:

- Expansion of the air bag system to include side impact protection (Dunn, Chap. 7)
- Magnetic transistors and diodes that can be directly integrated with signal conditioning circuits (Dunn, Chap. 7)
- Electronic switched stop lamps involving a rate-of-closure detector system to determine if the vehicle's speed is safe for objects ahead of it. If the closure rate is unsafe, the stop lights could be activated to alert trailing drivers to a pending accident (Valentine, Chap. 14)
- The integration of watchdog and failsafe functions onto a microcontroller (Boehmer, Chap. 11)
- Microcontrollers that operate at frequencies of 24 MHz or 32 MHz to allow more code to be executed in the same amount of time (Boehmer, Chap. 11)
- In the mid-90s, cars will have twice the electronic content of today's cars but will be easier to manufacture because there will be half the number of modules due to feature content integration. The data network interconnecting the modules will reduce the size and number of cables and cut the number of circuits by 50 percent (Miesterfeld, Chap. 26)
- A move from switching units to stepped operation actuators and the substitution of continuous for discrete time control (Müller, Chap. 10)
- Electrorheological and magnetorheological fluid actuators (Müller, Chap. 10)
- Micromechanical valves as actuators for converting low control power as in regulating the flow of fluids in hydraulic or pneumatic systems (Müller, Chap. 10)

REFERENCES

1. Trevor O. Jones, "Convergence—past and future," *Proceedings,* 1992 International Congress on Transportation Electronics (Convergence), P-260, Society of Automotive Engineers, Inc., Warrendale, Pa., Oct. 1992, pp. 1–3.
2. George W. Niepoth, and Stonestreet, Stephen P., "Closed-loop engine control," *IEEE Spectrum,* Nov. 1977, pp. 52–55.
3. Charles M. Heiner, and Beckman, Eldred W., "Balancing clean air against good mileage," *IEEE Spectrum,* Nov. 1977, pp. 46–50.
4. Trevor O. Jones, "Automotive electronics I: smaller and better," *IEEE Spectrum,* Nov. 1977, pp. 34–35.

P · A · R · T · 2

SENSORS AND ACTUATORS

CHAPTER 2
PRESSURE SENSORS

Randy Frank
Technical Marketing Manager
Motorola Semiconductor Products

2.1 AUTOMOTIVE PRESSURE MEASUREMENTS

Various pressure measurements are required in both the development and usage of vehicles to optimize performance, determine safe operation, assure conformance to government regulations, and advise the driver. These sensors monitor vehicle functions, provide information to control systems, and measure parameters for indication to the driver. The sensors can also provide data log information for diagnostic purposes.

Depending on the parameter being measured, different units for indicating pressure will be used. Since pressure is force per unit area, basic units are pounds per square inch (psi) or kilograms per square centimeter. For example, tire pressure is usually indicated in psi. Manifold pressure is typically specified in kiloPascals (kPa). A Pascal, which is the international unit (SI or Systems Internationale) for pressure, is equal to 1 Newton per meter^2 or 1 kg $\cdot$ m^{-1} $\cdot$ s^{-2}. Other common units of pressure measurement include: inches, feet, or centimeters of water; millibars or bars, inches, or millimeters of mercury (Hg), and torr. The conversion constants as defined per international convention are indicated in Table 2.1.

Pressure can be measured by a number of devices that provide a predictable variation when pressure is applied. Sensors used on vehicles range from mechanical devices—with position movement when pressure is applied—to a rubber or elastomer diaphragm, to semiconductor-based silicon pressure sensors. Various pressure-sensing techniques are explained in Sec. 2.3.

The type of pressure measurement that is made can be divided into five basic areas which are independent of the technology used for the measurement: gage, absolute, differential, liquid level, and pressure switch.

2.1.1 Gage Pressure Measurements

The silicon pressure sensor technology explained in Sec. 2.3.5 is used to visualize the difference between gage, absolute, and differential pressure (refer to Fig. 2.1). For gage pressure measurements, the pressure is applied to the top of a (silicon) diaphragm (Fig. 2.1), creating a positive output. The opposite (back) side of the diaphragm is exposed to atmospheric pressure. Gage vacuum is a special case of gage pressure. For gage vacuum measurements, vacuum is applied to the back side of the diaphragm resulting in a positive output signal. Gage and gage vacuum are single-sided pressure measurements. Gage pressure is frequently indicated by psig.

TABLE 2.1 Pressure Unit Conversion Constants

(Most commonly used per international conventions)

	psi*	In $H_2O^{\dagger}$	In $Hg^{\ddagger}$	K Pascal	Millibar	cm $H_2O^{\S}$	mm $Hg^{\P}$
psi*	1.000	27.680	2.036	6.8947	68.947	70.308	51.715
In $H_2O^{\dagger}$	3.6127×10^{-2}	1.000	7.3554×10^{-2}	0.2491	2.491	2.5400	1.8683
In $Hg^{\ddagger}$	0.4912	13.596	1.000	3.3864	33.864	34.532	25.400
K Pascal	0.14504	4.0147	0.2953	1.000	10.000	10.1973	7.5006
Millibar	0/01450	0.40147	0.02953	0.100	1.000	1.01973	0.75006
cm $H_2O^{\S}$	1.4223×10^{-2}	0.3937	2.8958×10^{-2}	0.09806	0.9806	1.000	0.7355
mm $Hg^{\P}$	1.9337×10^{-2}	0.53525	3.9370×10^{-2}	0.13332	1.3332	1.3595	1.000

* PSI = pounds per square inch
† At 39 °F
‡ At 32 °F
§ At 4 °C
¶ At 0 °C

2.1.2 Absolute Pressure Measurements

An absolute pressure measurement is made with respect to a fixed (usually a vacuum) reference sealed within the sensor (Fig. 2.1). For a 100-kPa-rated absolute unit, the diaphragm is fully deflected with standard atmospheric pressure. Application of a vacuum restores the diaphragm to its undeflected (flat) position. The result is a high-level output with no vacuum applied and a low-level signal at full vacuum unless the zero is established at the full-scale deflection of the diaphragm. Pressure can also be applied to absolute units with appropriately designed diaphragms to withstand the additional applied stress. Absolute pressure is frequently indicated by psia.

2.1.3 Differential Pressure Measurements

Differential or Delta-P measurements are also shown in Fig. 2.1. The higher pressure is applied to the top of the diaphragm and the lower pressure, possibly a reference pressure, is applied to the opposite side. The diaphragm's deflection is a result of the pressure difference.

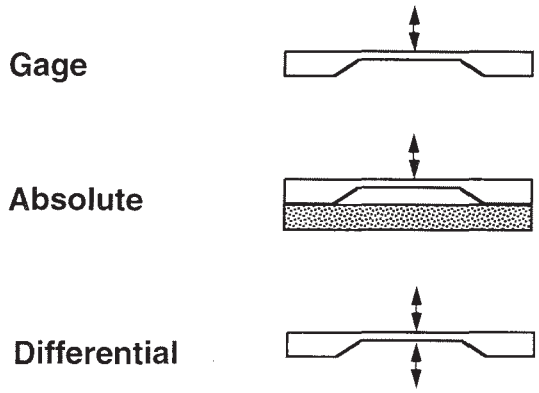


FIGURE 2.1 Types of pressure measurements: (*a*) gage, (*b*) absolute, and (*c*) differential.

Typically, the pressure differential is only a small percentage of the total line pressure and a system fault can expose one side of the sensor to the full line pressure. This must be taken into account when choosing the sensor and determining the rated pressure range that will be required. Differential pressure is frequently indicated by psid.

2.1.4 Liquid Level Measurements

The height of a column of liquid can be measured by a pressure sensor. The term *head* is frequently used in hydraulics to denote pressure. Measurements of inches or feet of water and centimeters of mercury are direct indications of the effect of pressure on liquid level. Other liquid levels are dependent on their specific weight and can be calculated by $h = (P_L - P_H)/w$, where $(P_L - P_L)$ is the pressure differential caused by the height of the fluid column and w is the specific weight of the liquid. The vapor pressure in a sealed enclosure will have an effect on the measurement of liquid height. Returning the reference side of a differential pressure sensor to the top of the enclosure will compensate for vapor pressure.

2.1.5 Pressure Switch

A pressure switch is typically achieved by mounting an electric contact on a diaphragm (rubber or any elastic material). The application of sufficient pressure (or vacuum) on one side of the diaphragm causes the movable contact to meet a stationary contact and close the circuit.

A pressure switch can also be achieved by any of the previously described techniques merely by establishing a reference threshold voltage that is calibrated to indicate the point that the pressure changes from an acceptable to unacceptable (or low to high) level. Once the threshold voltage is achieved, additional electronic circuits can be used to produce an electronic switch that can control loads such as an indicator lamp.

2.2 AUTOMOTIVE APPLICATIONS FOR PRESSURE SENSORS

Automotive requirements for pressure measurements range from the basic—oil pressure—to the sophisticated—air pressure differential from one side of the vehicle to the other. This section elaborates on the various possibilities for pressure measurements that exist either in the development, laboratory, or pilot phases of the vehicle, to actual volume production. Table 2.2 lists a number of potential pressure measurements versus vehicle systems and provides an indication of the pressure range and type of measurement.

Automotive specification and testing guidelines have been developed and published by the Society of Automotive Engineers (SAE) specifically for manifold absolute pressure (MAP) sensors. These documents are intended to assist in establishing test methods and specifications for other sensors. Other SAE documents that may apply to sensors are summarized in Table 2.3.

The packaging and testing requirements for automotive sensors can represent 50 to 80 percent of the sensor cost and over 90 percent of the warranty and in-service problems. The pressure-sensing applications that are presented in the following sections will include packaging requirements that are of particular concern.

2.2.1 Existing Applications for Pressure Sensors

A late twentieth century production vehicle is likely to have a number of pressure sensors for measurements such as manifold pressure and engine oil pressure, and has the potential for

TABLE 2.2 Pressure Sensing Requirements for Various Vehicle Systems

System	Parameter	Pressure range	Type
Engine control	Manifold absolute pressure	100 kPa	Absolute
	Turbo boost pressure	200 kPa	Absolute
	Barometric pressure (altitude)	100 kPa	Absolute
	EGR pressure	7.5 psi	Gage
	Fuel pressure	15 psi—450 kPa	Gage
	Fuel vapor pressure	15 in H_2O	Gage
	Mass air flow		Differential
	Combustion pressure	100 Bar, 16.7 Mpa	Differential
	Exhaust gas pressure	100 kPa	Gage
	Secondary air pressure	100 kPa	Gage
Elect transmission	Transmission oil pressure	80 psi	Gage
(continuously variable transmission)	Vacuum modulation	100 kPa	Absolute
Idle speed control	AC clutch sensor/switch	300–500 psi	Absolute*
	Power steering pressure	500 psi	Absolute*
Elect power steering (also elect assisted)	Hydraulic pressure	500 psi	Absolute*
Antiskid brakes/ traction control	Brake pressure	500 psi	Absolute*
	Fluid level	12 in H_2O	Gage
Air bags	Bag pressure	7.5 psi	Gage
Suspension	Pneumatic spring pressure	1 MPa	Absolute*
Security/keyless entry	Passenger compartment pressure	100 kPa	Absolute
HVAC (climate control)	Air flow (PC) Compressor pressure	300–500 psi	Absolute*
Driver information	Oil pressure	80 psi	Gage
	Fuel level	15 in H_2O	Gage
	Oil level	15 in H_2O	Gage
	Coolant pressure	200 kPa	Gage
	Coolant level	24 in H_2O	Gage
	Windshield washer level	12 in H_2O	Gage
	Transmission oil level	12 in H_2O	Gage
	Tire pressure	50 psi	Gage/absolute
	Battery fluid level	1–2 in below	Optical
Memory seat	Lumbar pressure	7.5 psi	Gage
Multiplex/diagnostics	Multiple usage of sensors		

* Gage measurement but absolute sensors used for failsafe

TABLE 2.3 SAE Specifications That Effect Pressure Sensors

Recommended environmental practices for electronic equipment design	SAE J1211
Performance levels and methods of measurement of electromagnetic radiation from vehicles and devices	SAEJ551
Performance levels and methods of measurement of EMR from vehicles and devices (narrowband RF)	SAEJ1816
Electromagnetic susceptibility procedures for vehicle components (except aircraft)	SAE J1113
Vehicle electromagnetic radiated susceptibility testing using a large TEM cell	SAE J1407
Open-field whole-vehicle radiated susceptibility 10 kHz–18 GHz, electric field	SAE J1338
Class B data communication network interface	SAE J1850
Diagnostic acronyms, terms, and definitions for electrical/electronic components	SAE J1930
Failure mode severity classification	SAE J1812
Guide to manifold absolute pressure transducer representative test method	SAE J1346
Guide to manifold absolute pressure transducer representative specification	SAE J1347

several other pressure measurements. Tighter emissions control and improved efficiency may necessitate further sensor use in future systems.

Manifold, Barometric and Turbo Boost Pressure. Manifold absolute pressure (MAP) is used as an input to fuel and ignition control in internal combustion engine control systems. The speed-density system that uses the MAP sensor has been preferred over mass air flow (MAF) control because it's less expensive, but stricter emission standards are causing more manufacturers to use mass air flow for future models.

Higher resolution from 32-bit engine controllers, with greater analog-to-digital (A/D) conversion capability and higher operating frequencies, will provide greater accuracy for a given MAP sensor during the critical transitions of the engine cycle. As shown in Fig. 2.2, previous changes from 8-bit to 16-bit controllers have resulted in a two-time improvement in resolution in the digital conversion for the intake manifold pressure. The 8-bit control unit performed the A/D conversion on a 4-ms timer interrupt in order to maintain a balance with other controls, with the resulting 1.1-ms lag time (worst case) during periods of overlapping interrupts. The 16-bit microcontroller performs the A/D conversion every 2 ms, which reduces the lag time to 0.3 ms. The actual system improvements that can result from using the higher performing microcontrol units is a result of other factors such as more precise and faster control of fuel injectors and sparkplugs, and additional and/or more accurate sensors and control algorithms.

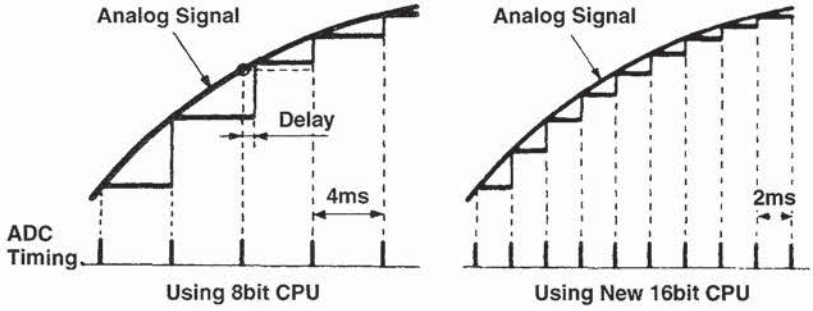


FIGURE 2.2 Effect of A-D on pressure measurements.

The MAP error band is also being tightened with a goal of 1 percent accuracy over the entire automotive temperature range. Existing specifications allow tolerances to increase as shown by the bowtie specification in Fig. 2.3 with associated multiplier(s).

The need for barometric pressure is often desirable in MAF systems to provide altitude information to the engine control computer. The barometric pressure range is typically from 60 to 115 kpA with accuracy on the order of 1.5 percent over the operating pressure range. The error band tolerance increases by a temperature multiplier of up to 2× outside the 0 to 85 °C range. MAP and barometric pressure sensors are frequently mounted inside control modules making the mounting technique a key consideration for manufacturability. The increased usage of surface mount technology, and the need to reduce space so that additional features can be included in control modules are factors that will affect next-generation sensor designs.

A typical turbocharger can provide boost pressure in the range of 80 kPa over the naturally aspirated internal combustion engine. This increases the maximum rating for the sensor to 200 kPa absolute, but other requirements are scaled appropriately.

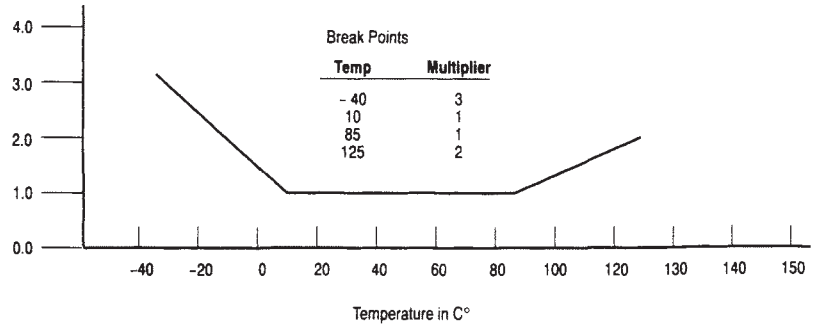
MPX4100 • MPX4101 SERIES

Transfer Function

Nominal Transfer Value: $V_{out} = V_S (0.01059 \times P - 0.1518)$
+/– (Pressure Error x Temp. Mult. x 0.01059 x V_S)
$V_S = 5.1\ V \pm 5\%$ P_{in} kPa

Temperature Error Multiplier

MPX4100 Series



Pressure Error Band

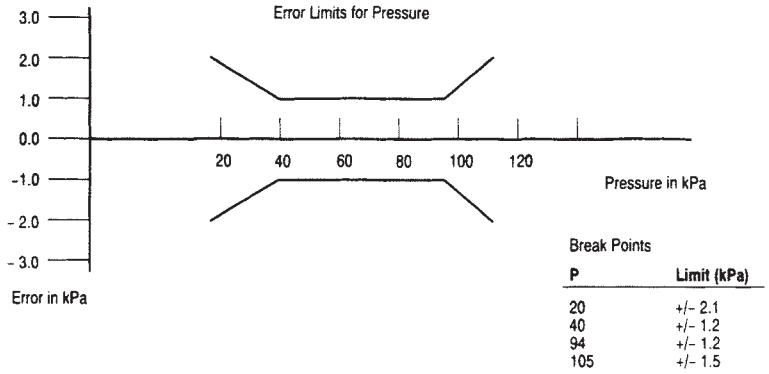


FIGURE 2.3 Error band for MAP sensor.

Oil Pressure. Oil pressure on automobiles has traditionally been measured by the deflection of a rubber diaphragm which closes a set of contacts (switch) providing a lamp indication with low oil pressure or moves a potentiometer to provide an analog signal for a gage.

A replacement for the conventional electromechanical oil pressure sending unit is an electronic device such as the one designed by Chrysler's Acustar Division. In addition to a silicon

piezoresistive pressure sensor, the unit contains transient protection circuitry, signal amplification for the sensor output, and output drivers for both an electromagnetic gauge and a fuel pump. The FET output drivers are capable of handling 10 A based on the heat dissipated though a heatsink that is integral to the sensor package.

The unit utilizes a supply voltage from 9 to 16 V and operates over a media temperature range of −40 to +150 °C. The overall accuracy is ±3.25 percent with linearity ≤ ±0.25 percent over the entire operating pressure range of 200 psi. The switch point for the low pressure indication is 4 psi ± 1.5 psi.

The sensor package was specially designed for easy assembly. The housing interfaces to the sensor with an extremely reliable O-ring seal that can withstand a burst pressure over 400 psi. Special materials were used for both the package and the protective gel that covers the sensor, which allow it to survive qualification tests with over 1 million pressure cycles, including portions conducted at high temperatures. This exceeds the number of cycles that can be achieved with traditional diaphragm-driven potentiometers that have been used for providing the indication of oil pressure. The sensor has been designed for a 10-year/100,000-mile life that could be required for future vehicle warranties.

Media Compatibility in Automotive Measurements. Pressure sensors frequently have to interface with an environment that is more demanding than other electronic components. For example, the measurement of engine oil pressure, transmission oil pressure, fuel pressure, and so on, or fluid level (oil, gas, coolant, etc.) requires the sensor package to be exposed to one or more fluids that are detrimental to the operation of semiconductor circuitry. Each of these media interface problems is addressed separately, depending on the application. Automotive cost requirements usually limit the usage of stainless steel as the isolation technique. Instead, more cost-effective protective polymers and chemically tolerant plastic and rubber materials have been developed for sensor packages.

2.2.2 New Applications for Pressure Sensing

The list of potential applications for sensing pressure in the automobile includes several new applications. These measurements are frequently made during the development of new vehicle systems. Their actual use on the vehicle is determined by factors such as cost, legislated requirements, need for diagnostics, and value added to the system. Applications in this section will identify areas of concern, range of pressure measurement, and factors that affect the usage of a pressure sensor.

Transmission Oil Pressure and Brake Pressure. Transmission pressure is required as an input in computer-controlled transmission shift points. This pressure can be measured with sensors similar to those developed for engine oil pressure.

Pressure in a hydraulic system, such as the master cylinder of an antilock brake system (ABS), is much higher than transmission oil pressure typically requiring a sensor with minimum rating of 500 psig. Pressures in other locations in the ABS system can be lower. The dynamic pressures in brake tubing can be of interest during the development phase of passenger vehicles and may be of interest in heavy duty commercial vehicles. These pressures can be below 150 psig.

The ABS system controls front and rear tire slip. Tradeoffs that exist in developing an ABS system for a particular vehicle include stopping deceleration to achieve the shortest possible stopping distance versus more steering control. Increased yaw stability can be obtained by reducing the deceleration rate of the rear wheels. The addition of traction control to the system improves stability during acceleration and provides independent control of each wheel during a variety of driving maneuvers for improved vehicle performance.

Passenger vehicles may have a single pressure sensor to monitor the pressure in the hydraulic system. One system, General Motors' ABS-VI, provides information on the brake pressure by detecting the current going to motors in the system. For the ABS-VI system, a pres-

sure sensor is not required to provide optimum brake pressure at each wheel. However, other systems rely on the rate of brake application and release to control lockup. Commercial vehicles may have several sensors for sensing brake pressures. Sensors close to brake cylinders report the actual pressure, which is compared to the reference value stored in the control unit.

Tire Pressure. The continuous monitoring of tire pressure offers increased fuel economy and safety to passenger cars or commercial vehicles. Underinflated tires create excessive rolling friction and therefore decrease fuel economy. Overinflated tires have excessive stress that can result in failure during operation. Improperly inflated, either over- or underinflated, tires have uneven wear patterns which decrease tire life. Available tire pressure systems consist of a tire pressure sensor (or switch) at each wheel, wheel speed indicator, temperature sensor, a radio frequency transmitter, electronic receiver/controller, and a display unit. The dashboard display provides an indication to the driver that the tires are improperly inflated. Tire pressure increases with temperature approximately 1.5 psi for every 10 °C of tire air temperature rise, so the system must provide correction for this effect. Abrupt increases in temperature and pressure can be sensed by these systems and provide a warning of eminent tire failure providing an additional safety factor.

Another tire pressure system utilizes a separate hand-held reader to easily verify the proper tire inflation when the vehicle is stationary. Yet another commercial vehicle system for trucks with dual tires operates while the vehicle is stationary and employs a visual indication for the driver that adequate pressure exists. This system provides a single fill point for the dual tires, maintains equal pressure under normal conditions, and provides an isolation valve in the event that a blowout or slow leak develops in one tire.

EGR Pressure. EGR (exhaust gas recirculation) back pressure and a pressure differential exist across the EGR valve used to control NO_x emissions in the engine control system. The valve is modulated by a vacuum which lifts a pintle from its seated position to allow exhaust gas to be recirculated. A change in vacuum pressure from 50 to 90 mm Hg is sufficient to fully open the valve, and a pressure differential across the valve of 200 mm Hg is typical. Pressure measurements are made during the development phase to determine system operating characteristics. However, a position sensor is typically used to measure the EGR valve's position and control NO_x emissions during normal vehicle operation.

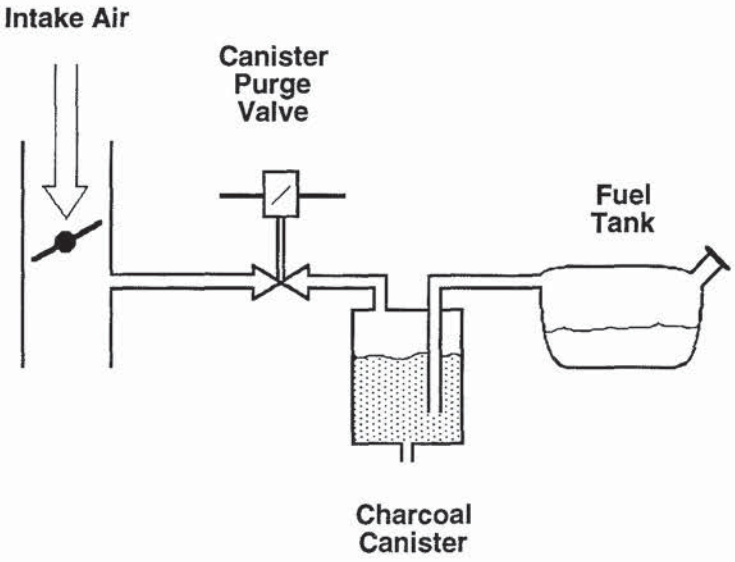


FIGURE 2.4 Canister purge system.

Fuel Rail and Vapor Pressure. Evaporative emissions that occur when the engine is off are currently stored in an activated charcoal canister of about 850 to 1500 cc until the engine is running, as shown in Fig. 2.4. The vapors are then consumed by the combustion chamber and catalytic converter. One implementation of on-board vapor containment of refueling hydrocarbons (on-board refueling vapor recovery or ORVR) would require refueling canisters on the vehicle that could be three to four times the volume of existing canisters. If leaks need to be detected in this system, a diagnostic pressure sensor may be required.

One approach to fuel routing, employed in the 5.9-liter Dodge Magnum engine, is to mount the fuel filter and pressure regulator at the fuel tank. The fuel pump is mounted inside the tank. Therefore, since only a fuel supply line to the fuel rail and a line to the evaporative canister are necessary, the fuel return line is eliminated. This system maintains lower fuel tank temperatures, resulting in lower evaporative emissions.

Monitors required for on-board diagnostic (OBD) systems per California Air Resources Board's (CARB's) OBDII legislation were originally targeted to be phased in between model year 1994 and model year 1996. A Bosch fuel injection system with on-board diagnostics is shown in Fig. 2.5 that identifies a differential air pressure sensor for tank vapor pressure.

Overpressure Occurrences. Fuel supply pressures in automobile fuel injection systems normally operate at pressures below 75 psi; however, fuel pumps develop pressures up to 3200 psi to open injectors. Pressure spikes can be reflected back through the fuel supply system that measure up to 300 psi during each fuel injection pulse.

Overpressure created by backfire can also apply a positive pressure of 75 psi and higher to the intake manifold absolute pressure sensor. Techniques used to prevent failure from over-

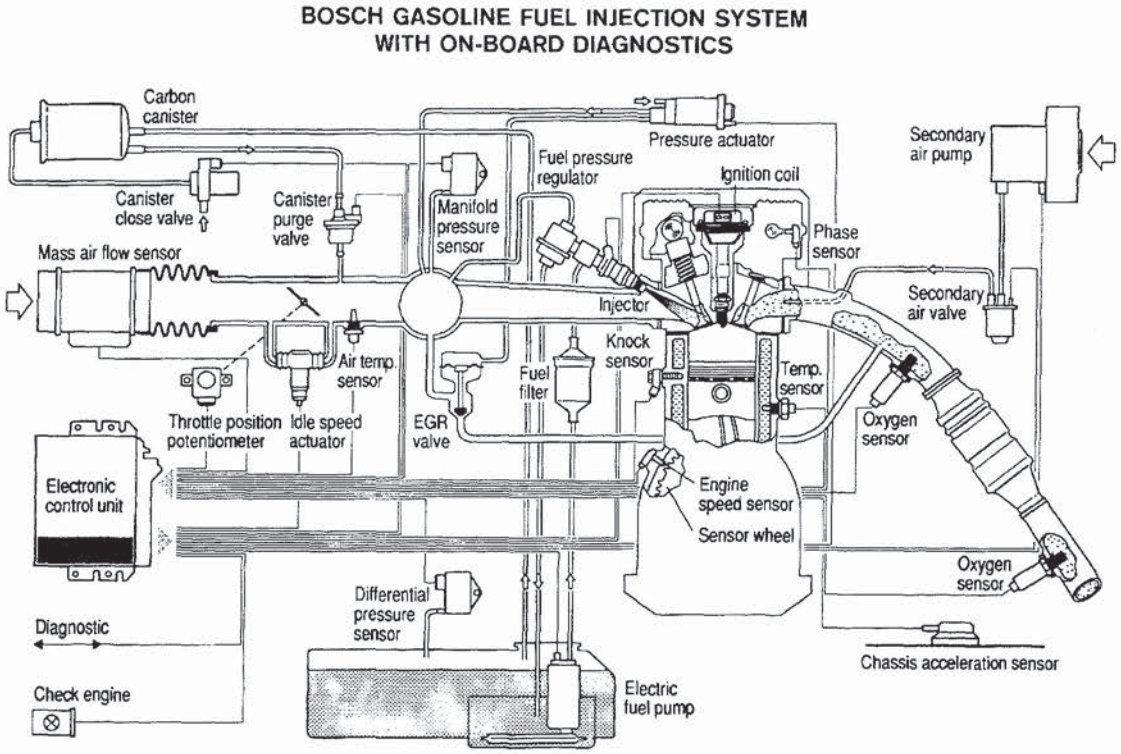


FIGURE 2.5 Fuel vapor control in electronic fuel injection system. *(Courtesy Robert Bosch, GMBh)*

pressure include overpressure stop built into the transducer, mechanical pulse filtering, and a sensor designed to operate within the overpressure range.

Mechanical stops have been a traditional protection technique for mechanical pressure sensors. This is possible where the amount of diaphragm deflection is large. Silicon pressure sensors have a modulus of elasticity that is the same as steel (30×10^6 psi) and a yield strength (180,000 to 300,000 psi) that is higher than steel, allowing high overpressure without diaphragm damage. However, the sensor package itself must be designed to handle the maximum pressure safely.

Snubbing, or mechanical filtering, is commonly used for static pressure measurements. A small diameter tube reduces the dynamic variation in applied pressure. If dynamic measurements are desired, the ac component of the desired signal may also be attenuated.

Increasing the diaphragm thickness of the sensor to safely handle the full range of pressure within normal operating range will also result in a lower sensitivity.

Alternate Fuel and Alternate Engine Implications to Pressure. Legislation that requires a percentage of the vehicles sold in California to be LEV (low emissions vehicles) and even ZEV (zero emissions vehicles) is increasing the demand for alternate fuel and electric vehicles. Among the alternate fuel vehicles, CNG (compressed natural gas) and hydrogen cells most likely would require sensors for pressure measurements. CNG is pressurized at 3000 psi and the distribution system includes pressure regulators, a transducer, valves, and idle air solenoids. Before the natural gas enters the engine, a regulator reduces the fuel pressure to near atmospheric pressure. Sensing may be required in both low- and high-pressure portions of the system. Development of alternative engines, such as the two-stroke engine for vehicle applications, will utilize electronics for control functions similar to four-cycle engines. However, the range and necessity for pressure measurements will differ from four-cycle engines. The pressure range for direct fuel injection is considerably higher for a two-stroke engine. The need to control the oil pump may necessitate pressure sensing in two-stroke systems as well.

Hydrogen fuel cells are another potential source of energy for use in electric vehicles. In one design, the proton-exchange membrane (PEM) design, a turbocompressor is used to pressurize the system and maintain hydration of the membrane. A pressure of at least three atmospheres (0.3 MPa) is required to remove the water. This pressure or the pressure drop across the membrane may need to be monitored during operation.

2.2.3 Other Applications for Pressure Sensors

Pressure sensors can be used on vehicles for measuring flow through pressure differential, or delta-P, measurements and for determining liquid level.

Delta-P (Flow-Sensing) Measurements. Applications on the vehicle for flow sensing include mass air flow; heating compartment flow; oil, fuel, and cooling liquid flow; and vehicle flow in an air stream. Mass air flow is typically accomplished by hot-wire anemometer or Karman vortex flow meters which do not use pressure-sensing techniques. Other vehicle flow requirements, including the pressure drop across the air filter, could be sensed and monitored by a differential pressure sensor. In addition to requirements such as media compatibility (Sec. 2.2.1), the lower-level signals require higher-sensitivity pressure sensors and/or additional amplification and must tolerate faults that could apply full line pressure to only one side of the sensor.

One of the more interesting applications of differential pressure measurements applied to flow analysis is the flow of the vehicle itself relative to crosswinds. A rear-wheel steering system developed by Daimler-Benz uses two pressure sensors to measure the pressure caused by wind on the vehicle's sides. An electronic control unit analyzes the pressure difference and inputs from other sensors, and alters the rear-wheel setting according to the wind strength. The system that measures the crosswinds directly is faster than yaw sensors, which are reactive and measure the change in attitude and direction of the vehicle.

Other laboratory measurements of airflow and crosswind force have also been made by Daimler-Benz that utilize 10 differential pressure gages with a range of ±3700 kPa. All pressure sensors were connected to a single pressure vessel to have a common reference. The reference pressure was measured by a 100-kPa absolute pressure sensor. The measurements were used to determine aerodynamic forces and moments and to compensate for wind effects in an active steering system.

Fluid Level Sensing. Various liquid levels can be measured in a vehicle, as shown in Table 2.4. All of these requirements, except fuel level, could be satisfied by a switch that simply detects that a predefined minimum level of liquid has been reached so that a driver indication can be provided. This can be accomplished by directly illuminating a dash lamp or through a microcontroller in a body computer which activates an output driver.

TABLE 2.4 Liquid Level Measurements

Level	Type	Range
Engine oil	Switch	38.1 cm of water
Transmission oil	Switch	30.5 cm of water
Coolant	Switch	61 cm of water
Windshield washer fluid	Switch	30.5 cm of water
Battery	Refraction switch	5.1 cm below reference
Power steering fluid	Switch	7.6 cm below reference
Brake fluid	Switch	30.5 cm of water
Fuel	Sensor and switch	38.1 cm of water

Sensing the fuel level has traditionally been performed by a float to sense the fluid level and a variable resistor with the wiper arm connected to the float. Configuration for the sensor depends upon the specific tank for which it was designed, necessitating a unique sensor for each vehicle. Manufacturers with several different vehicle models have the additional impetus to replace the existing techniques with a nonwearing, more accurate, self-calibrating electronic alternative. However, media compatibility for fuel level is among the harshest requirements for a pressure sensor. In addition to gasoline, oxygenated fuels containing ethanol, methanol, benzene, MTBE, engine additives, and even sour gas must be tolerated by the sensor. Nonintrusive differential sensors isolate the liquid from the sensor interface but must still tolerate fuel vapors. Also, the sloshing of the fuel in a vehicle's tank requires a time amplitude filter to smooth out the indication provided to the driver.

2.2.4 Combustion Pressure

The direct measurement of combustion pressure is being investigated for detecting misfire to meet CARB OBDII requirements. The high pressure (≥16 MPa) and temperature ranges combined with other environment factors make the design of a pressure sensor for this application extremely expensive. As a result, other techniques are being developed as alternatives to direct pressure measurement. These technologies include optical, fiberoptics measuring luminous emissions from combustion, and noncontact torque sensors. Section 2.3.7 explains a fiber optic technique.

The operation of the reciprocating-piston, internal combustion engine is represented by a constant volume process and the engine power cycle is analyzed by using pressure-versus-volume and pressure-versus-crank angle diagrams. To obtain these measurements in a laboratory environment a number of techniques have been developed. Direct (in-cylinder) pressure measurements have been performed with small diameter piezoresistive sensors placed in (or near) the sparkplug and piezoelectric washers placed under the spark plug. A high natural fre-

quency is required for these sensors based on the dynamic measurement involved in the combustion process. Indirect measurements with shaft torque and optical phase shift are additional possibilities. The need to determine misfire due to component failure during vehicle operation is part of OBDII requirements. A sensor used on production vehicles will be required to survive the high pressure and temperature environment for the life of the vehicle, which could be 100,000 miles and 10 years. It must also have no need for periodic zeroing or calibration and be available at a low cost.

2.2.5 Other Pressure Measurements

An adaptive suspension system (see Chap. 17) can be accomplished with an air pressure controlled shock absorber damping, such as Mitsubishi's Active-ECS (Electronically Controlled Suspension). This system has two air pumps and nine solenoids that regulate air pressure based on inputs from sensors including an air pressure sensor in the rear of the vehicle that measures the passenger and cargo load. The driver can select soft, medium, or hard suspension. Another system utilizes an air reservoir charged to a pressure of 1 Mpa by a compressor. A pressure switch monitors when the pressure drops below 760 kPa to recharge the reservoir. Air springs operate at 300 kPa unloaded and at 600 kPa in the rear when fully loaded.

HVAC (heating, ventilating, and air conditioning) changes are occurring as manufacturers are required to convert from refrigerant CFC-12 to alternatives such as HFC-134a. The theoretical performance of these two refrigerants will mean about a 6 percent loss in efficiency, compressor discharge pressure that is 175 kPa higher, and a discharge temperature that is about 8 °C lower. Measuring the compressor discharge pressure (almost 1900 kPa for the HFC-134a system) is desirable for electric load control as vehicles add more requirements to the 12-V charging system. Also, the effect on engine performance and fuel economy when the A/C is used could make the refrigerant pressure sensor a standard vehicle sensor in the future. An absolute sensor used to measure gage pressure of the refrigerant provides a deadhead effect to prevent refrigerant loss in the event of a sensing diaphragm failure.

The measurement of the pressure developed when the air bag is inflated is part of the evaluation, qualification, and lot acceptance criteria of air bag inflating techniques. Time-to-peak pressure and peak tank pressure measurements require measurements in the tens of milliseconds range. Inflated bag pressures are in the range of 100 kPa or less. Hybrid inflators use a stored inert gas, such as argon, in place of sodium-azide propellant that requires a squib for ignition. The hybrid uses a pressure sensor to monitor the status of the stored gas.

Special heavy duty/commercial truck measurements require pressure measurements that are quite different from those made on passenger cars. Accumulator-type fuel injection systems for direct injection diesel engines have fuel pressurized to 20 to 100 MPa in the accumulator by a high-pressure pump. The accumulator pressure is monitored in this approach to reduce particulates. Another method to reduce diesel particulates utilizes a ceramic fiber as a filter in a canister. A pressure sensor monitors the backpressure and allows the full filter to be regenerated by burning off the accumulated particulates in the filter. A heater element in the trap has power supplied from the power-switching module. A temperature approaching 1300 °F (700 °C) is reached inside the filter cartridge to incinerate the particles.

Lumbar support systems utilize a pressurized system with a pressure sensor ($\leq$7.5 psig) as the feedback element controlling the air pump to provide additional support to the driver's lower back in luxury vehicles. Pressure-sensitive grids have been used in the development process to automatically measure up to 3600 contact points for visual display and weight distribution analysis.

2.2.6 Partial Pressure Measurements

The oxygen (or lambda-) sensor in engine control systems is a chemical sensor that utilizes partial pressures to provide a feedback signal for the closed-loop control system. Lambda is

defined as the actual air/fuel ratio divided by the stoichiometric (14.6) fuel ratio. The operation of the oxygen sensor is defined by the Nernst equation: $U_L = RT/4F \cdot 1n(P_{O2}''/P_{O2}')$, where U_L is the unloaded output voltage of the sensor, R is the universal gas constant, T is the absolute temperature, F is the Faraday constant, P_{O2}'' is the oxygen partial pressure of the air (about 2.9 psi), and P_{O2}' is the equilibrium partial pressure of the oxygen in the exhaust gases. Equilibrium occurs due to the catalytic activity of the platinum electrodes used to coat the inside and outside of the Y_2O_3 stabilized ZrO_2 ceramic electrode. The oxygen partial pressure changes by a factor of 10^7 at 900°C (or 10^{19} at 500°C) when the exhaust gas changes from a reducing environment (lambda = 0.999) to an oxidizing environment (lambda = 1.001).

2.3 TECHNOLOGIES FOR SENSING PRESSURE

A number of technologies have been used for on-vehicle measurements of static and dynamic pressure: diaphragm-potentiometer, linear variable differential transformer (LVDT), aneroid, silicon or ceramic capacitive, piezoresistive strain gage, piezoelectric ceramic or film, and optical phase shift (combustion pressure). Recent advances in sensing have focused on transducers that provide an electric signal easily interfaced to microcontrollers. Mechanical devices are frequently used in the laboratory for calibration and component development or on the vehicle during the development phase of the vehicle systems. Common mechanical devices include the Bourdon tube, bellows, diaphragms, deadweight gage, and manometer.

The Bourdon tube is a curved or twisted, flattened tube with one end closed that acts as a force collector. When pressure is applied at the open end, the tube tends to straighten and the resulting motion is used as an indication of the applied pressure.

The bellows, or pressure capsule, is a chamber that expands with applied pressure. Absolute pressure can be measured by sealing a reference pressure (e.g., vacuum) inside a closed unit and applying pressure to the outside. The movement of the chamber is proportional to the applied pressure.

Diaphragms are the most common force collector used in modern pressure sensors. The diaphragm material can be rubber, elastomer, stainless steel, silicon, ceramic, or even sapphire. Diaphragm shapes are circular or square and can be supported or clamped around their periphery.

A deadweight tester or piston gage can withstand extreme pressure changes and high overpressure occurrences. The piston is sealed in a cylinder using O-ring or Teflon seals. Pressure on the piston causes a deflection that can be measured by position-sensing techniques. Precision weights allow calibration for high-accuracy measurements. The deadweight tester is one of the few pressure-sensing techniques that measures pressure in terms of its fundamental units—force and area. Errors associated with the measurement are air-bouyancy corrections, gravity error, surface tension, fluidhead, and thermal expansivity. These errors are normally small but should be taken into account when high accuracy is required.

The manometer is used both as a pressure-measuring instrument and a standard for calibrating other instruments. Its simplicity and inherent accuracy result from it being the measurement of the height(s) of a column of liquid. Three basic types of manometers are the U-tube, well (cistern), and slant-tube.

Other measurement devices including McLeod, Pirani, Alphatron, and thermocouple gages which can measure vacuum in the range of 10^{-5} mm Hg.

Sensing techniques that provide a transducer for conversion from mechanical to electrical units include resistive, LVDT, capacitive, piezoresistive, and piezoelectric.

2.3.1 Potentiometric Pressure Sensors

Prior to electronic engine control systems, carburetor dashpots and distributor vacuum advance units used the distance that a rubber diaphragm traveled when pressure was applied

as a mechanical indication of pressure. A diaphragm which moves a potentiometer (resistor with a sliding element or wiper) provides an electric signal that can be applied to a remote gage such as an oil pressure gage. Potentiometric sensors inherently have some level of noise and wear associated with their operation due to the contact of the wiper to the resistive element. Stiction or static friction is also a potential concern with these devices, especially if control of ≤0.5 percent of the total resistance is required. The finite resolution of wirewound potentiometers is overcome by the use of newer thin-film plastic resistor designs.

2.3.2 Linear Variable Differential Transformer

One of the earliest pressure inputs for engine control systems was provided by the LVDT pressure sensor. The principle of operation is demonstrated in Fig. 2.6. An LVDT pressure sensor consists of a primary winding and two secondary windings positioned on a movable cylindrical core. The core is attached to a force collector which provides differential coupling from the primary to the secondaries resulting in a position output that is proportional to pressure.

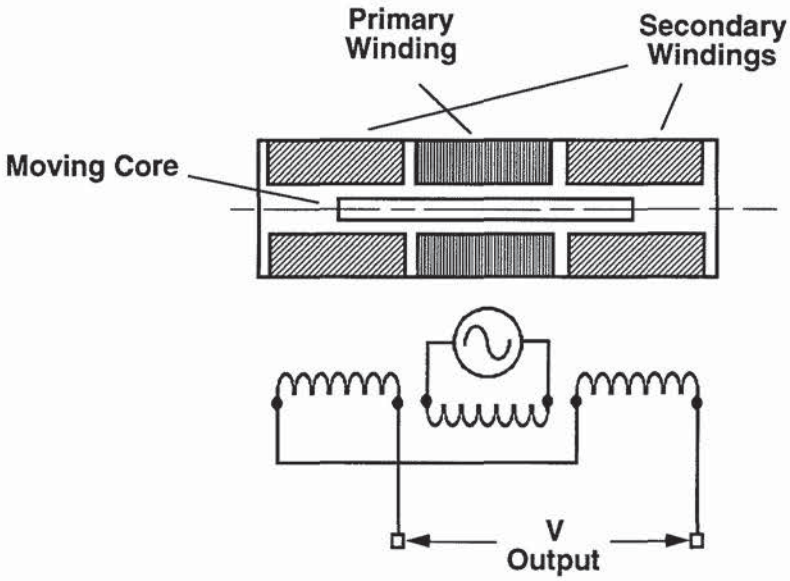


FIGURE 2.6 LVDT pressure sensor.

An alternating current is used to energize the primary winding, which results in a voltage being induced in each of the secondary windings. The windings are connected in series opposing, so the equal but opposite output of each winding tends to cancel (except for a small residual voltage called the null voltage). A pressure applied to a Bourdon tube or diaphragm causes the core to be displaced from its null position and the coupling between the primary and each of the secondaries is no longer equal. The resulting output varies linearly within the design range of the transducer and has a phase change of 180° from one side of the null position to the other. Since the core and coil structures are not in physical contact, essentially frictionless movement occurs.

Electronic devices necessary to signal condition the output of an LVDT consist of an oscillator for the supply voltage, circuitry to transform the constant voltage to a constant current, an amplifier with high input impedance for the output, a synchronous demodulator, and a filter with characteristics designed for quasistatic or dynamic measurements.

2.3.3 Aneroid Diaphragms

Another early design for sensing automotive manifold pressure consisted of dual sealed aneroid diaphragms. The diaphragms were bonded and sealed with a vacuum inside each unit to a metallized conductive ring on opposite sides of a ceramic substrate. The substrate served as the fixed plates of two separate capacitors. Manifold vacuum was applied to one chamber and the second served as a reference for compensating and signal conditioning that minimized common mode errors due to vibration and shock.

2.3.4 Capacitive Pressure Sensors

Capacitive pressure sensors have one plate that is connected to a force collector (usually a diaphragm), and the distance between the plates varies as a result of the applied pressure. The nominal capacitance is $C = Ae/d$, where A is the area of the plate, e is the permittivity, and d is the distance between the plates. Two common capacitive pressure sensors used in automotive applications are based on silicon and ceramic capacitors.

Silicon. A silicon capacitive absolute pressure (SCAP)-sensing element is shown in Fig. 2.7. The micromachined silicon diaphragm with controlled cavity depth is anodically bonded to a Pyrex® glass substrate. Feedthrough holes are drilled in the glass to provide a precise connection to the capacitor plates inside the unit. The glass substrate is metallized using thin-film deposition techniques. Photolithography is used to define the electrode configuration. After attaching the top silicon wafer to the glass substrate, the drilled holes are solder-sealed under vacuum to provide a capacitive-sensing element with an internal vacuum reference and solder bumps for direct mounting to a circuit board or ceramic substrate. The value of the capacitor changes linearly from approximately 32 to 39 pF with applied pressure from 17 to 105 kPa. The capacitive element is 6.7 mm × 6.7 mm and has a low-temperature coefficient of capacitance (−30 to 80 ppm/°C), good linearity (≈1.4 percent), fast response time (≈1 ms), and no

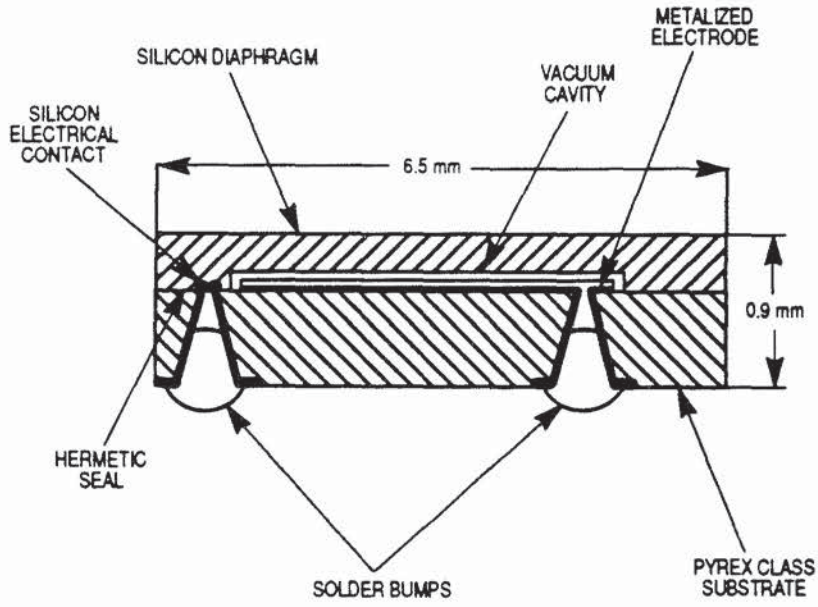


FIGURE 2.7 SCAP sensor.

exposed bond wires. The output of the sensor is typically signal-conditioned to provide a frequency variation with applied pressure for easy interface to microcontrollers.

Surface micromachining and bulk micromachined silicon-on-silicon techniques (see Sec. 2.3.5) have also been used to build silicon capacitive pressure sensors. These approaches also address the addition of signal-conditioning electronics on the same silicon structure.

Ceramic. The ability to make thin diaphragms from ceramic material combined with thin-film deposition to provide metal plates and connections has been used to manufacture ceramic pressure sensors. Their operation and signal-conditioning requirements are similar to the silicon capacitive sensor described in Sec. 2.3.4.

2.3.5 Piezoresistive Strain Gage

Strain-gage pressure sensors convert the change in resistance in four (sometimes in two) arms of a Wheatstone bridge. The change in the resistance of a material when it is mechanically stressed is called piezoresistivity.

The open-circuit voltage of an unbalanced Wheatstone bridge is given by $V_O = \mathbf{E}[(R1*R3 - R2*R4)/(R1 + R2)(R3 + R4)]$, where V_O is the output voltage, **E** is the applied voltage, and $R1$ through $R4$ are the resistive elements of the bridge. Additional variable resistive elements are typically added to adjust for zero-offset and sensitivity, and to provide temperature compensation.

Different approaches to piezoresistive strain gages range from traditional bonded and unbonded to the newest integrated silicon pressure sensors.

Bonded and Unbonded Strain Gages. The bonded resistance strain-gage pressure sensor consists of a filament-wire or foil, metallic or semiconductor, bulk material or deposited film bonded to the surface of a diaphragm. Pressure applied to the diaphragm produces a change in the dimension of the diaphragm and, consequently, in the length of the gage, and, therefore, a change in its resistance ($R = \rho\ L/A$). The change per unit length is called strain. The sensitivity of a strain gage is indicated by gage factor

$$GF = \Delta R/R \div \Delta L/L = 1 + 2\mu + \Delta\rho/\rho \div \Delta L/L \tag{2.1}$$

Foil strain gages have a negligible piezoresistive effect and their gage factor is usually between 2 and 3.

When a pressure sensor is used for measuring an applied force it is called a load cell.

Integrated Silicon Pressure Sensors. The GF for a strain gage is improved considerably (to about 150) by using a silicon strain gage. In addition to the conventional Wheatstone bridge, silicon processing techniques, and the relative size of piezoresistive elements in silicon enable the design of a unique piezoresistive sensor. The sensor signal can be provided from a single piezoresistive element located at a 45° angle in the center of the edge of a square diaphragm which provides an extremely linear measurement. The offset voltage and full scale span of the basic sensing element vary with temperature, but in a highly predictable manner. In addition to the basic sensing element, an interactively laser-trimmed four-stage network has also been integrated into a single monolithic structure (Fig. 2.8). The size of the silicon die, including the diaphragm, sensing element, and signal-conditioning electronics, is only 0.135 in by 0.145 in. The die is attached to the six-terminal package through the use of a stress-isolating layer of RTV (room temperature vulcanizing) silicone. This approach allows a minimum of external components for amplification to provide a usable output signal.

Two silicon wafers are used to produce the absolute piezoresistive silicon pressure sensor (Fig. 2.9). The top wafer is etched until a thin, square diaphragm, approximately one mil in thickness, is achieved. The square area is extremely reproducible as is the 54.7° angle of the cavity wall based on the characteristics of bulk micromachined silicon. The top wafer is

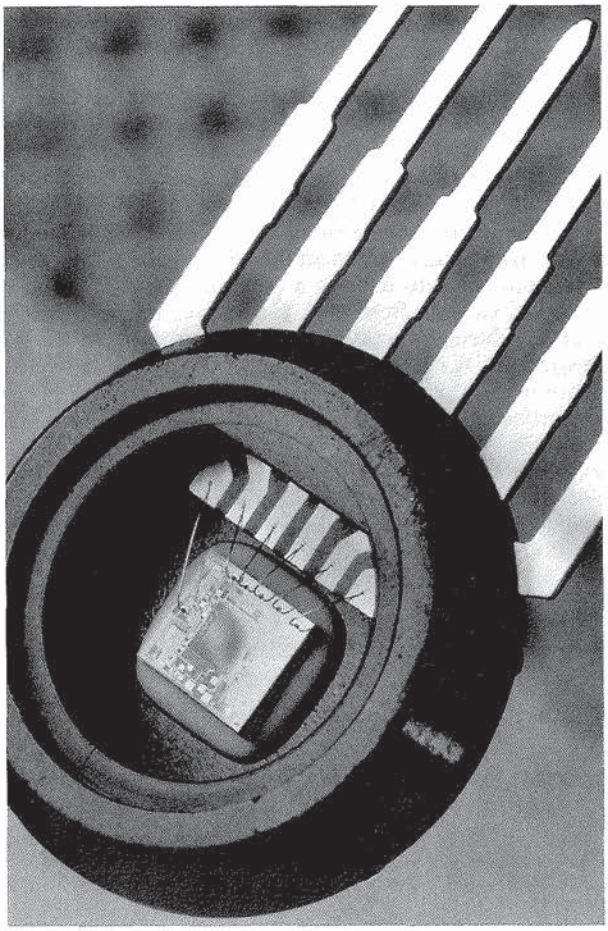


FIGURE 2.8 Silicon pressure sensor with integrated electronics. *(Courtesy Motorola, Inc.)*

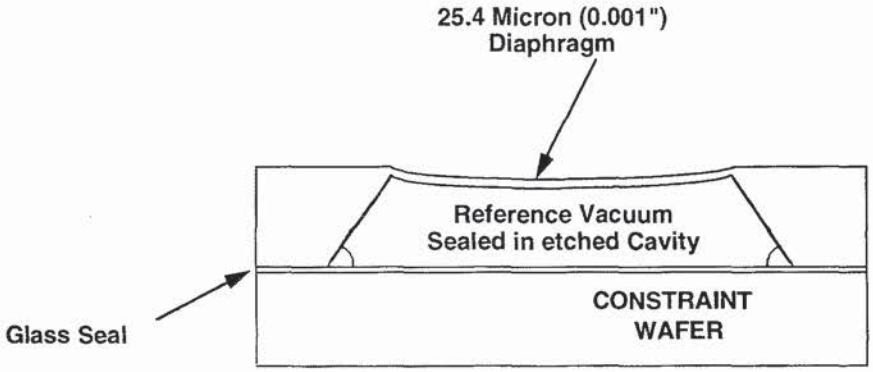


FIGURE 2.9 Cross section of piezoresistive silicon pressure sensor for measuring absolute pressure.

attached to a support wafer by a glass frit to provide a structure which is isolated from mounting stresses.

The bulk micromachining process used to form the diaphragm and the etched cavity in the majority of silicon pressure sensors is a chemical etching process that allows a thin (0.001-in) mechanical structure—the diaphragm—to be precisely etched from a silicon wafer that is approximately 0.015 in thick. Hundreds of sensors can be formed simultaneously on a 4-in (100-mm) diameter silicon wafer, and several wafers can be batch-processed to yield thousands of sensors in a single lot. Silicon pressure sensors can also be achieved by using surface micromachining techniques. In these sensors, a layer of sacrificial material (usually an oxide) is grown on top of a silicon wafer, and material such as polysilicon is then deposited on the sacrificial layer and patterned to achieve a particular structure. The sacrificial material is removed by a chemical etchant. Both bulk and surface micromachining techniques can be combined with semiconductor processing techniques to provide additional circuitry on the same monolithic structure. A number of new terms are used relative to silicon pressure sensors that are defined in the glossary to this section.

Both bulk and surface micromachining, discussed previously, are performed at the wafer level. A polysilicon thin-film sensor that consists of a thin film of silicon that is doped with boron and vapor-deposited over a stainless steel diaphragm is shown in Fig. 2.10. A thin deposited layer of silicon dioxide insulates the silicon from the stainless steel diaphragm. Silicon nitride is used to cover the strain-sensitive elements. Silicon-on-insulated-stainless-steel (SOISS) sensors are not formed using silicon wafer techniques, but they use batch-processing techniques and are inherently suited for harsh environments.

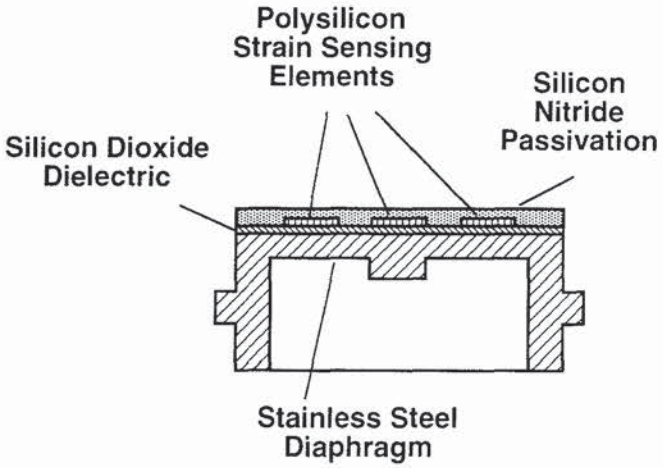


FIGURE 2.10 Polysilicon pressure sensor on stainless steel diaphragm.

2.3.6 Piezoelectric Pressure Sensors

A piezoelectric sensor produces a change in electric charge when a force is applied across the face of a crystal or piezoelectric film. The inherent ability to sense vibration and the necessity for high-impedance circuitry are taken into account in the design of modern piezocrystal sensors. Transducers are constructed with rigid multiple plates and a cultured-quartz sensing element, which contains an integral accelerometer to minimize vibration sensitivity and suppress resonances. A typical unit also contains a built-in microelectronic amplifier to transform the high-impedance electrostatic charge output from the crystals into a low-impedance voltage signal. Units made in stainless steel housings have an invar diaphragm laser welded to seal the

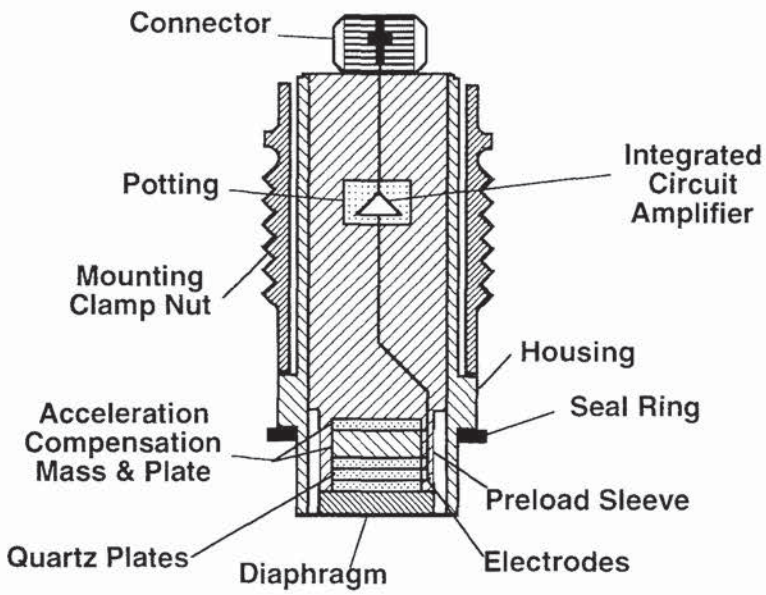


FIGURE 2.11 Piezoelectric pressure sensor.

sensing elements inside the package. Figure 2.11 shows the cross section of a piezoelectric pressure sensor.

More recently, piezo film sensors, which produce an output voltage when they are deflected, provide a very inexpensive method for pressure measurements. One approach has the piezo film cemented to a metallic dimple substrate with the dimple pointed toward the high-pressure source. As the pressure rises, a point is reached when the dimple snaps in the opposite direction and the movement is sufficient for the piezo film to generate a transient voltage.

Surface micromachining techniques have also been combined with piezoelectric thin-film materials, such as zinc oxide, to produce a semiconductor piezoelectric pressure sensor.

2.3.7 Fiber Optic Combustion Pressure Sensor

For extremely high pressure, or pressure measurements at high temperatures, different pressure measurement techniques are used. Figure 2.12 shows an alternative to traditional pressure-sensing techniques that is being developed to sense production vehicles' combustion

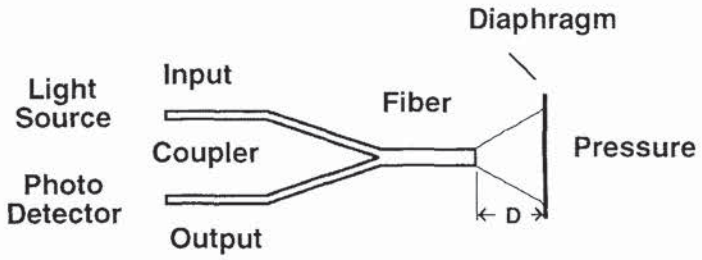


FIGURE 2.12 Fiber optic combustion pressure sensor.

pressure. The fiber optic pressure sensor can withstand temperatures (up to 550 °C) that are well above the normal range for piezoelectric sensors. Furthermore, the design has a normal pressure range of 0 to 1000 psi and overrange capability of 3000 psi.

The sensor's operation is the result of a light source input to an optical fiber coupler and a photodetector at the receiving end of the coupler. Light exits the optical fiber as a diverging cone which illuminates a diaphragm. The maximum angle is determined by an aperture. The amount of light that is returned to the sensor fiber after it is reflected from the diaphragm depends upon the gap D between the fiber and the diaphragm. The diaphragm can be sized to allow the sensor to be integrated into a spark plug for easy access to the combustion pressure of each cylinder. Accuracy within 5 percent has been demonstrated within the 550 °C operating temperature range.

2.3.8 Pressure Switch

A pressure switch can be simpler and more cost effective than a pressure sensor for an application that only requires detecting a single pressure level. Sufficient motion of the force-collecting diaphragm (e.g., elastomer, Kapton®, fluorosilicon) allows a spring to be compressed and a set of contacts to be closed in a traditional mechanical switch. The design of the contacts may allow several amperes to be switched.

Converting the output of any of the electric sensors in Sec. 2.3 to a threshold-sensing level requires additional circuits, including an electronic switch, such as a power FET, to conduct the current. The sensor can have multiple switch points depending on the amount of additional circuitry that is provided.

2.3.9 Pressure Valves/Regulators

Pressure is frequently controlled in automotive applications by pressure regulators or valves. The actuation of these mechanisms can be a result of applied pressure to a mechanical structure such as an integral piston or relief valve held closed by a spring force, or an electric signal generated from a sensor and subsequent activation of a solenoid, which opens a valve or moves a pintle in an orifice to change the pressure. The thermostat in the engine cooling system allows flow based on a minimum temperature being achieved. It must operate independently of the pressure variations in the cooling system. A common solution is an expansion-element thermostat which actuates a valve to redirect the flow of coolant into a radiator bypass line when the control temperature is reached.

Extremely small and precise silicon-based regulators, and even pumps, are possible using micromachining techniques. Figure 2.13 shows a silicon Fluistor™ (or fluidic transistor) microvalve that is approximately 5.5 mm by 6.5 mm by 2 mm. The valve is actually a thermopneumatic actuator which accepts an electric input. The cavity is etched in the middle silicon chip by bulk micromachining described in Sec. 2.3.4 and filled with a control liquid. When the liquid is heated, the silicon diaphragm moves outward over the valve seat. This approach has demonstrated a dynamic range of 100,000:1 controlling gas flows from 4 μlpm to 4 lpm at a pressure of 20 psid.

The microvalve combined with a pressure sensor and electronic feedback loop provides a solid state pressure regulator. It has potential for usage in both gas and liquid flow control applications on future vehicles. The integration of EGR and idle-air control (IAC) has already been accomplished in a somewhat conventional (patented) manner. A feedback-controlled valve with two inlets and a single outlet orifice eliminate stepper motor programming in the engine controller and only one calibration curve is needed for both EGR and IAC functions. The combination of this approach to control systems and newly developed technologies, such as the silicon microvalve, will allow additional advances in vehicle performance, efficiency, and control.

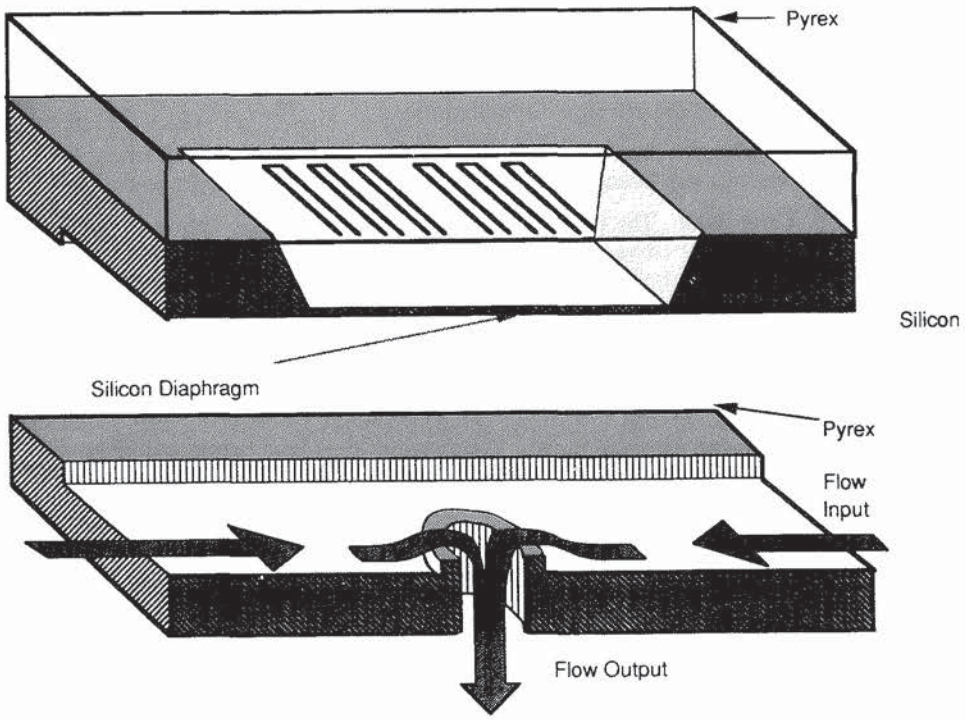


FIGURE 2.13 Silicon microvalve. *(Courtesy of Redwood MicroSystems, Inc.)*

2.4 FUTURE PRESSURE-SENSING DEVELOPMENTS

The different types of pressure measurements, different technologies for measuring pressure, and potential pressure measurements in automotive applications have been explained in this chapter. In addition, alternatives to making pressure measurements, such as indirect sensing and the use of pressure regulators, have been discussed. The use of semiconductor technology applied to sensing applications is producing sensors with inherently more decision and diagnostic capability that can communicate bidirectionally with host microcomputers in complex systems. The desire to directly produce a signal that is compatible with microcomputers rather than requiring analog signals to be converted to digital format through A/D converters is spurring development activity that could affect future automobile systems. Furthermore, recently developed fuzzy logic and neural network approaches to control systems and multiplexing of sensor outputs for use in several systems will make previously cost-prohibitive sensing applications a reality.

Increased and recently initiated sensing activity from industrial organizations such as SAE, the American National Standards Institute (ANSI), and the Institute of Electrical and Electronics Engineers (IEEE) should provide a greater level of understanding, common terminology, and improved specifications and test procedures for the numerous approaches that can be taken to sense pressure in automotive applications.

GLOSSARY

Altimetric pressure transducer A barometric pressure transducer used to determine altitude from the pressure-altitude profile.

Diaphragm The membrane of material that remains after etching a cavity into the silicon sensing chip. Changes in input pressure cause the diaphragm to deflect.

Error band The band of maximum deviations of the output values from a specified reference line or curve due to those causes attributable to the sensor. Usually expressed as "±% of full-scale output." The error band should be specified as applicable over at least two calibration cycles so as to include repeatability and verified accordingly.

Linearity error The maximum deviation of the output from a straight-line relationship with pressure over the operating pressure range. The type of straight-line relationship (end-point, least-square approximation, etc.) should be specified.

Operating pressure range The range of pressures between minimum and maximum pressures at which the output will meet the specified operating characteristics.

Overpressure The maximum specified pressure that may be applied to the sensing element of a sensor without causing a permanent change in the output characteristics.

Piezoresistance A resistive element that changes resistance relative to the applied stress it experiences (e.g., strain gauge).

Pressure error The maximum difference between the true pressure and the pressure inferred from the output for any pressure in the operating pressure range.

Pressure sensor A device that converts an input pressure into an electric output.

Ratiometric (ratiometricity error) At a given supply voltage, sensor output is a proportion of that supply voltage. Ratiometricity error is the change in this proportion resulting from any change to the supply voltage. Usually expressed as a percent of full-scale output.

Response time The time required for the incremental change in the output to go from 10 to 90 percent of its final value when subjected to a specified step change in pressure.

Temperature error The maximum change in output at any pressure in the operating pressure range when the temperature is changed over a specified temperature range.

BIBLIOGRAPHY

"Acustar Electronic Oil Pressure Sensor," *Automotive Industries,* March 1993, pp. 26–29.

Budd, John W., "A Look at Pressure Transducers," *Sensors,* July 1990, pp. 10–15.

Doeblin, Ernest O., *Measurement Systems Application and Design,* McGraw-Hill, New York, 1975.

He, Gang, and Marek T. Wlodarczyk, "Spark Plug-Integrated Fiber Optic Combustion Pressure Sensor," *Proceedings of Sensors Expo,* Chicago, Sept. 29–Oct. 1, 1992, pp. 211–216.

Holt, Daniel J., "ABS Testing," *Automotive Engineering,* March 1993, pp. 26–29.

Keebler, Jack, "Automakers, gas refiners debate vapor control units," *Automotive News,* Oct. 14, 1991, p. 39.

Lynch, Terrence, "Integrated Valve Meters EGR and Idle Air," *Design News,* Feb. 22, 1993, pp. 159–160.

Motorola Pressure Sensor Device Data Book, Q1/93, Phoenix, Ariz.

Norton, Harry N., "Transducers and Sensors," *Electronic Handbook,* McGraw-Hill, New York.

PCB Electronics, Inc., comments in "Piezoelectric Pressure Transducers," *Measurements & Control,* Oct. 1990, pp. 20–222.

Sawyer, Christopher A., "On-Board Diagnostics," *Automotive Industries,* May 1992, p. 57.

Siuru, Jr., William D., "Sensing Tire Pressure on the Move," *Sensors,* July 1990, pp. 16–19.

Tran, Van Truan, "Wind Forces and Moments," *Automotive Engineering,* April 1990, pp. 35–38.

ABOUT THE AUTHOR

Randy Frank is a Technical Marketing Manager for Motorola's Semiconductor Products Sector in Phoenix, Arizona. He has a BSEE, MSEE, and MBA from Wayne State University in Detroit, Michigan, and over 25 years' experience in automotive and control systems engineering. For the past 10 years he has been involved with semiconductor sensors, power transistors, and smart power ICs.

CHAPTER 3
LINEAR AND ANGLE POSITION SENSORS

Paul Nickson
Product Line Manager
Analog Devices, Inc.,
Transport & Industrial Products Division

3.1 INTRODUCTION

Position sensors of one form or another are an integral and necessary part of the modern automobile. Position sensors range in technology from the ubiquitous microswitch warning to the driver of a door ajar to linear variable differential transformers (LVDTs) used in sophisticated active suspension systems. Whether as monitors or as critical parts of safety systems, market and legislative pressures for longer warranties and lower emissions are opening avenues for a wide range of sensing technologies to find a place in the modern automobile.

The automotive systems designer must take into account many factors when choosing the appropriate technology for an application. Each sensor type has its own vocabulary, and it is important when making comparisons to understand how a figure of merit for one sensor relates to that of another. It is equally important to understand how the choice of output signal format, whether digital or analog, can affect the resolution of measurement and subsequently the performance or stability of an automotive system.

The purpose of this chapter is to give an overview of position sensor technologies currently used and available for use in automobiles and to compare their characteristics and suitability for particular applications. Consideration is given to the interfacing requirements of each type of sensor with an emphasis on the advantages and disadvantages of each method as they apply in the automotive environment. Where appropriate, descriptions of applications of the various sensor types in automobile applications are given. Other available technologies and technologies in development which have desirable characteristics for automotive applications are also discussed.

3.2 CLASSIFICATION OF SENSORS

Sensors may be classified in many different ways. From the perspective of a system designer, the basic questions are: What kind of information does the sensor provide and how is the sensor used? For the purposes of this discussion, a position sensor is defined as an electromechanical device that translates position information into electric signals. Sensors can be grouped into two basic categories.

3.2.1 Incremental or Absolute

Position information can be presented in two different ways. Incremental position sensors measure position as the distance from an arbitrary index or zero. Alternatively, position information may be provided that gives an unambiguous or absolute measure of the distance from a well-defined index.

Incremental sensors usually rely on some method of pulse counting. One pulse in the sequence is designed to be wider or of opposite polarity than the others so that it may be used as a nominal zero. A typical optical angle encoder consists of a glass disk marked with a number of equally spaced radial opaque lines and transparent gaps. The disk is illuminated on one side and a light sensor and associated electronics on the other side detect the passage of dark lines and gaps and generate corresponding electric pulses. Dedicated electronics built into the sensor or a remote microcontroller can be used to count the pulses. A zero is established by detecting the wider pulse, known as *North Marker* in optical encoder terminology, and then resetting the pulse counter. The advantage of this data format is that few wires are required to carry the information. Typically, four or five wires would be required depending on the exact details of the format (see Sec. 3.3.2). The biggest disadvantage of incremental sensors is that at power-up the system has no position information and requires a mechanical indexing cycle to find the marker pulse. Another disadvantage is that the system is prone to the effects of noise, which may lead to erroneous counts.

In contrast, absolute position sensors produce an unambiguous output at power-up. Each position or angle has a unique value. The output may be a voltage or frequency or other analog of the input position. Potentiometers are often used in applications requiring this characteristic. Many absolute position sensors have binary digital outputs. The digital formats vary depending on the construction of the sensor. Some optical encoders use Gray code to avoid ambiguities at code transitions. Other sensors, such as resolvers, do not directly produce a digital output but are almost always used with an analog-to-digital converter that may output in parallel or serial form one of the common formats—for example, two's complement or offset binary.

3.2.2 Contact or Proximity

Position sensors are designed to detect the position of components of mechanical systems by either being directly coupled by some shaft or linkage, as in the case of potentiometers or optical encoders, or by some noncontact or proximity means. Environmental issues are often a key influence in the choice of sensor for a given application. High levels of vibration, particularly in small engine applications, may cause rapid wear of the conductive track of a throttle-position-measuring potentiometer. Dirt and dust usually exclude optical sensors from underhood applications due to rapid degradation of the optical path.

The most common form of proximity sensors are based on various methods of magnetic field detection. A device based on magnetic field sensing principles may be more easily isolated from the destructive effects of the harsh environment encountered in many automotive applications.

3.3 POSITION SENSOR TECHNOLOGIES

3.3.1 Microswitches

The simplest form of contact sensor is a switch. Contact position sensors may be as simple as the microswitches that operate anything from brake lights to courtesy lights in the automobile. Many applications of microswitches in position sensing are as limit switches, usually

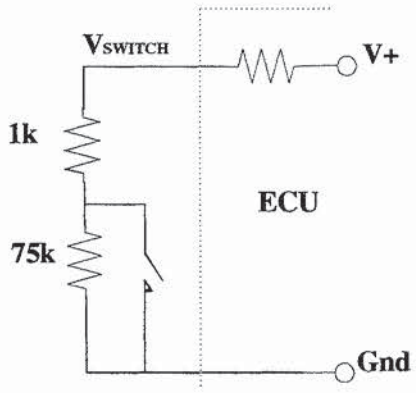


FIGURE 3.1 Diagnosable switch.

wired to limit or warn of the extent of travel of a mechanical component by disconnecting power to an electric motor or by operating an indicator lamp. In some cases, for safety reasons, it is desirable to be able to detect fault conditions that would make the switch inoperable. In some applications, it is possible to connect the switch as shown in Fig. 3.1. In this case, a diagnostic circuit measuring the voltage, V_{SWITCH}, can differentiate between the normal conditions of switch open or closed and can also determine if the switch is disconnected or if V_{SWITCH} is shorted to either power supply.

An undesirable characteristic of switches is that the contacts may bounce on closing. If it is important in the application that the first switching edge is detected, then a simple switch-debouncing logic that rejects noisy signals can be used. If a microcontroller is used to monitor the switch output, then debouncing can be accomplished by software means. This may be a better solution in applications subject to shock or heavy vibration, which may cause occasional false switching. In these cases, a microcontroller can be configured to poll the switch over some period of time and report switch closure only if a number of consecutive readings are the same.

3.3.2 Optical

Optical angle encoders for incremental shaft angular position measurement are constructed of a disk with a series of transparent and opaque equally spaced sectors. The disk can be made of glass for precision applications. Mylar film and metal disks offer high and medium resolutions, respectively, at low cost. The encoder disk is illuminated on one side and light sensors on the other side detect the passage of light and dark sectors as the disk is rotated. (Low-resolution versions such as the Hewlett-Packard HRPG series use an alternative reflective technology.) Most encoders have two sets of light sources and detectors offset by half the width of a sector. Figure 3.2 shows the relationship between the outputs of the light sensors as the encoder is rotated. This format is often referred to as "A quad B," since the signals are in quadrature. The passage of one pair of light and dark sectors over a detector is referred to variously as one cycle, one count, one line or 360 electrical degrees (°e). Encoder resolutions range from around 16 counts per revolution (CPR) for low-cost applications to over 6000 CPR for precision position control systems. Most encoders also include a third signal for use as an index or reference pulse. The index, or North marker as it is sometimes called, occurs once per revolution. The pulse width is usually equal to 90 °e.[1]

Four separate states, each of 90 °e, can be derived from the A and B outputs using integrated circuits available from a number of vendors. This allows a resolution of four times the number of lines on the encoder disk to be achieved. These ICs also determine the direction of rotation of the encoder by observing which output leads the other. By convention for clockwise rotation, the low-to-high transition of A leads the low-to-high transition of B. Control circuitry can be added to improve noise immunity by only allowing valid next states to be counted.

Incremental angle encoder accuracy specifications fall into two categories. The angular position accuracy is the difference between the actual shaft angle and the angle indicated by the encoder. This error is normally expressed in degrees or minutes of arc. The second category includes specifications for the symmetry and repeatability of individual cycles; these are usually

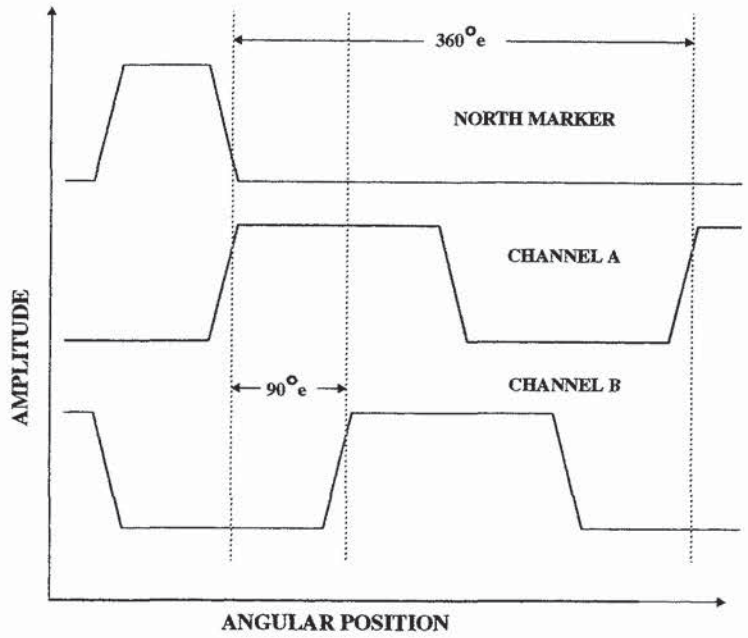


FIGURE 3.2 Encoder outputs.

expressed in electrical degrees. Typical characteristics are detailed in Fig. 3.3 and Table 3.1 Errors are caused by eccentricity and axial play of the code wheel and manufacturing defects in the lithography or etching of the code wheel. Modular encoders consisting of a light source and sensor head are available which use a collimated light source and an array of integrated photodiodes to minimize the effects of these error sources. Differential connection of the photodiodes ensures insensitivity to errors caused by light source variability due to environmental or other factors. A further source of error can occur if the encoder is rotated at high enough speeds that the rise and fall time of the digital outputs significantly affects the pulse width. The light sensor bandwidth usually determines the maximum rotational speed of the sensor. Typical bandwidths are below 100 kHz, which would limit the speed of a 100-CPR encoder to 1000 rpm.

Linear incremental optical encoders are available from many vendors. These allow direct measurement of linear motion. Modular sensor/emitter heads are available that can be used in these applications. The technology is basically the same as incremental angle encoders and the terminology used to describe specifications is the same as for angle encoders. Linear encoders are described in terms of their count density or resolution in counts per mm or mm

TABLE 3.1 Incremental Encoder Specifications

	Minimum	Typical	Maximum	Units
Position error		10	40	min of arc
Cycle error		3	10	°e
Pulse-width error		7	30	°e
Phase error		2	15	°e
State-width error		5	30	°e
Index pulse width	60	90	120	°e

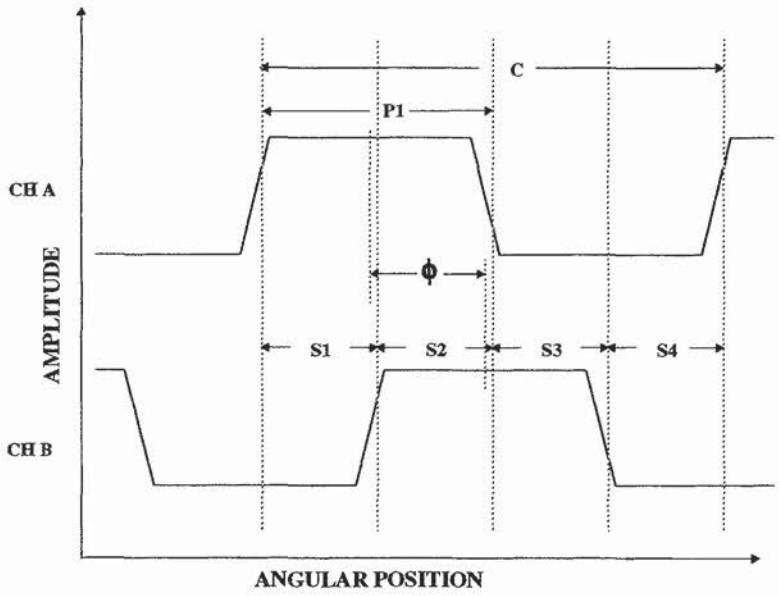


FIGURE 3.3 Encoder definitions.

per count. Line counts range up to approximately 8 per mm, allowing an ultimate linear resolution of around 30 μm.

If it is important to have an unambiguous measure of position as soon as power is applied to a system, then an absolute encoder must be used. Absolute optical encoders are manufactured with resolutions from one part in 2^6 to one part in 2^{16}. The data format can be binary, binary-coded decimal (BCD), or Gray code. An absolute angle encoder is divided into equal sectors which are arranged so that adjacent sectors contain consecutive digital words. The binary bits of each word form N concentric tracks on the encoder disk, where N is the digital word length. N sets of light sources and photodiodes detect the parallel word representing the input shaft angle.

Absolute optical encoders often use Gray code to eliminate code transition errors. In a natural binary sequence between zero and full scale on the disk, all the bits of the digital word change state together. Any misalignment of the code wheel and the light sensors, or sensor-to-sensor misalignment, will cause false codes to be generated. This could be disastrous for a position control system since a misread code could indicate an angle up to 180° away from the correct angle. Gray coding eliminates this problem. Gray code is a unit distance code; consecutive codes differ by only one binary bit. If a code transition is misread, the largest error will be one least significant bit of the digital word.

3.3.3 Potentiometric

Potentiometers are widely used as position sensors in automotive applications such as throttle and accelerator pedal position measurement. The automotive industry increasingly

demands low-cost mechanically and electrically rugged sensors to provide control or measurement of position in the modern automobile. This has resulted in the development of potentiometers that are capable of operational life far in excess of the life of the average car, and in some cases capable of continuous rotational speeds of above 1000 rpm for more than 1000 hours.[2]

Potentiometers can be constructed using a wire-wound track. The resolution of these potentiometers is determined by the number of turns of wire used to wind the track. The resolution of rotary wire-wound potentiometers is often quoted as the number of turns per degree and can be anywhere between 1 (1° per turn) and 7 (8.5 arcmin per turn). The track resistance is proportional to the number of turns used and can be in the range of 10 ohms to 100 kilohms, with a tolerance of approximately 5 percent. Wire-wound potentiometers have advantages where low-value variable resistors are required but do not excel in linearity, resolution, or rotational life which can be as low as 10^5 revolutions. Potentiometers for position-sensing applications are constructed using a resistive track of conductive material, usually a graphite and carbon black doped plastic, and a collector track molded on some supporting substrate. A drive shaft or pushrod draws precious metal multifingered wipers along the tracks. Wiper damping is usually included to make the potentiometer insensitive to vibration. Potentiometers of this type are manufactured with a range of resistance from around 500 ohms to 20 kilohms with a tolerance of 10 to 20 percent (Table 3.2). Potentiometers of this type are capable of excellent linearity and very high resolution.

TABLE 3.2 Potentiometer Specifications

Parameter	Minimum	Maximum
Electrical travel	90°, 10 mm	360, 3000
Nominal resistance	500 ohms	20 kilohms
Resistance tolerance	10%	20%
Resistance temperature coefficient (TC)		500 ppm/°C
TC of V_{out} in voltage divider mode		5 ppm/°C
Linearity error	0.01%	1%

Potentiometric sensors are used as voltage dividers. A reference voltage is applied across the resistive element and the wiper voltage is used as an absolute measure of the position of the actuator. Linear potentiometers are available in a wide range of lengths from 10 mm to 300 cm. Rotary potentiometers are usually restricted to 355° of useful range due to the dead band created by the track-end contacts. Some versions are available with true 360° operation. These use multiple wipers and dedicated electronics to eliminate the dead band.

All potentiometers are ratiometric sensors. That is, the wiper voltage at a given position is some fraction of the reference voltage applied across the resistive track. If the reference voltage is varied, the potentiometer output remains in the same ratio to the reference voltage. Sensor potentiometers, when properly terminated, maintain ratiometric operation over a wide range of temperatures with temperature coefficients typically less than 5 parts per million (ppm) per degree centigrade. Without special signal processing, ratiometricity is compromised at the end of the electrical travel of a potentiometer by the change in resistivity of the track as it joins the end contact and by any parasitic external resistance in series with the track. Figure 3.4 shows the effective limitations on the potentiometer at its endpoints.

Ratiometricity is a very desirable characteristic for a sensor that is used with comparators or analog-to-digital converters. For example, if the same reference voltage that powers the sensor is used as the reference for an analog-to-digital converter, then the measurement system will be insensitive to the absolute value of the reference voltage. A given shaft angle will always be reported as the same digital code. In most automotive control systems, analog-to-digital converters (usually on board a microcontroller) use the regulated 5-V engine controller

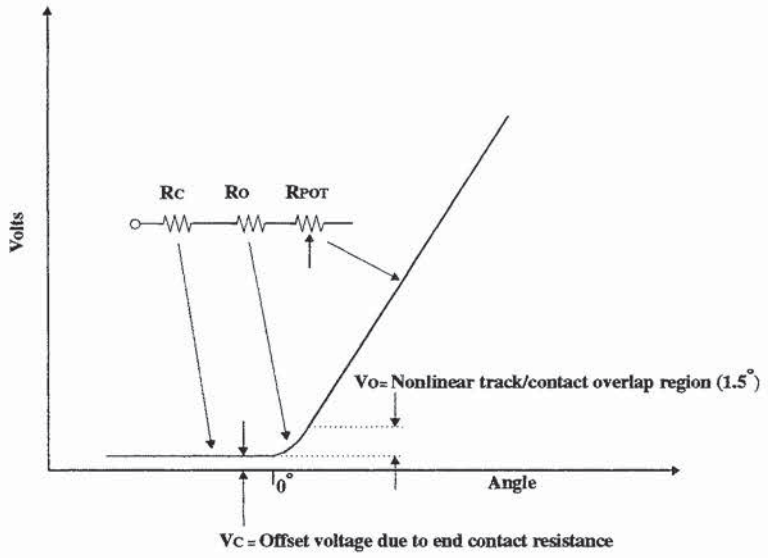


FIGURE 3.4 Potentiometer endpoint nonlinearity.

(ECU) power supply as reference to avoid the additional cost of a separate voltage reference chip. It is desirable that the same voltage, perhaps buffered to isolate the control module from accidental shorts, is used as reference by any ratiometric sensors in the automobile.

Potentiometers are subject to a number of sources of error, of which linearity is the most important. The linearity error is the difference between the actual transfer function of the potentiometer and the ideal transfer function (output voltage change versus mechanical travel) as a percentage of the applied reference voltage. Linearity specifications between 0.03 and 1 percent are available, with sensor cost inversely proportional to linearity. Microlinearity can be an important indicator of the suitability of the potentiometer for use in accurate control systems where large changes in the local gradient of the transfer function may cause instability. Ratiometricity, linearity, and offset errors can be caused by improperly loading the wiper of the potentiometer. The maximum error is at the center of travel of the wiper. If the load is a constant current, such as the input current of a buffer amplifier, a voltage will be developed across R_{EXT} that will cause an offset error at the extremes of travel of the wiper and an additional linearity error at the center of travel. Sensor potentiometers are usually specified with some maximum wiper current (100 nA, typically) to eliminate these errors. Plastic track potentiometers are capable of resolutions better than 0.001 percent of travel. This is primarily limited by the homogeneity of the resistive track material and hysteresis caused by limitations in the mechanical construction of the potentiometer related to bearings, wiper stiffness, and coefficient of friction of the track.

Plastic track sensor potentiometers are capable of operational life in excess of 10^7 revolutions for a rotary sensor or 10^7 strokes for a linear sensor. Unfortunately, no universal standards are available defining test conditions, and these may vary from vendor to vendor. Generally, two types of test are carried out. Dither testing simulates conditions which may exist in control applications or areas with high levels of vibration. The wiper is moved over a small proportion of the travel, say 1 or 2°, at a test frequency of 100 Hz. Information about contact resistance and local gradient changes in the potentiometer transfer function can be gathered rapidly using this technique. A change in gradient (ohms/percent of travel) relative to the mean gradient of the potentiometer can be equated to a change in loop gain in a con-

trol system. Sensitive systems may become unstable at worn points in the potentiometer track due to relative gradient errors.

A second method of potentiometer reliability testing is to repeatedly move the wiper at around 10 Hz over a large proportion of the available range. An excursion from 0 to 50 percent of the available travel will result in the maximum change in linearity error over a large number of cycles, since only one-half of the track is subject to wear. The criteria for failure of a potentiometer will be very application-specific and may, for example, be a doubling in linearity error. It is important to work with the potentiometer vendors to understand how their specifications can be extrapolated to make a prediction of operational life.

3.3.4 Magnetic

By far the largest category of position sensors relies on electromagnetic induction principles. This group of sensors can be broken into subgroups depending on the details of the employment of the phenomenon. Other sensors in this category rely on materials with magnetoresistive or magnetostrictive properties. Electromagnetic sensors have a number of advantages over other technologies. In general, this class of sensors measures or responds to changes in the relative position of components in a magnetic circuit. The components are always separated by an air gap and are not subject to friction wear. In many cases, it is possible to construct rugged sensors that are insensitive to the harshest automotive environments.

Variable Reluctance. The reluctance of a magnetic circuit determines the magnetomotive force (amp turns) required to produce a flux of a given value.[3] Variable reluctance devices operate by sensing changes in the reluctance within a magnetic circuit. In most cases, the reluctance change is caused by a change in the length of an air gap. The change in reluctance causes a change in the magnetic flux which induces a voltage in an output signal coil. The voltage induced is typically a bipolar pulse shape whose amplitude is proportional to the rate of change of flux (Faraday's law).

$$V = \left(\frac{d\Phi}{dt} \right) \tag{3.1}$$

This sensor technology cannot be used at zero speed since, if the rate of change of flux is zero, then the output will be zero.

In automotive applications, variable reluctance sensors are used to detect the position and speed of rotating toothed or slotted wheels in crank-, cam-, and wheel-monitoring applications. An easily magnetized or "soft" magnetic core or bobbin wound with a sense coil is magnetized by a strong permanent magnet such as samarium cobalt (Sm_2Co_{17}). The sense end of the core is placed in close proximity to a toothed gear wheel. The flux change that occurs when a tooth edge passes the sensor causes a voltage to be induced in the coil. Remote signal-conditioning electronics associated with the ECU are used to amplify the signal and produce a signal that a microcontroller can interpret as a position increment. An alternative construction[4] uses a coaxial pole piece to improve the magnetic circuit. This construction is particularly suited to sensing holes or apertures in a sense wheel. Sensors that detect slots and are positioned in close proximity to the target with a small air gap work at low reluctance and are less likely to be disturbed by interfering fields than sensors configured to detect teeth at a low mark space ratio.

Variable reluctance sensors are prone to a number of sources of error. Vibrations or resonance sometimes exacerbated by the attractive forces between the sensor and the target can seriously degrade the signal-to-noise ratio of the device. The sensors' target is usually a rotating ferromagnetic wheel or gear. Eddy currents will be generated by the movement of the wheel in the magnetic field of the sensor. This may lead to false readings from the sensor. In some refinements of variable reluctance sensors, the holes or apertures in the wheel are filled with an electrically conductive nonmagnetic material to homogenize eddy currents.[4]

Significant advantages of variable reluctance sensors in the automotive environment are their simple rugged construction and low cost. Additionally, they have a wide operating temperature range and require only two wires for operation. Variable reluctance sensors can also be used as variable inductance sensors by exciting the sense coil with alternating current and employing inductance-measuring means in the signal-conditioning electronics.

Hall Effect. Electric current is carried through the motion of electric charge. If a conductor is moved through a magnetic field with velocity v, the charges in the conductor will experience a force (Lorentz force) in a direction perpendicular to both the direction of motion and the magnetic field. This gives rise to an electric field of strength:

$$\mathbf{E} = v\mathbf{B} \tag{3.2}$$

The charges will move and a surface charge will develop on the conductor until an electrostatic field forms which counterbalances the electric field due to motion, $v\mathbf{B}$. A voltage due to the movement of charge can be detected with a voltmeter. The voltage is proportional to the field **B** and the velocity and length of the conductor. The result of this effect in thin films of material was first described by Hall over 100 years ago. When he passed current through a rectangle of gold foil in the presence of a magnetic field perpendicular to the plane of the foil, a voltage could be measured across the other axis of the foil.

Devices can be constructed using semiconductor materials which can utilize this effect to detect the strength of magnetic fields. Figure 3.5 illustrates the construction and operation of a silicon Hall effect device. A voltage is amplified across one axis of a thin block of high-resistivity n-type epitaxial material. In the presence of the field **B**, charges move in the direction of the arrow. A voltage directly proportional to **B**, the current density in the silicon and the Hall coefficient (scattering factor) can be measured at the point shown. The sensitivity is low and amplification is required to render a useful signal. For example, with a current of 10 mA flowing in an n-type silicon epitaxial layer 1 μm thick with a doping level of $10^{15}/cm^3$ and a field of 100 mT, a voltage difference of approximately 30 mV will be measured at 25 °C. Offsets caused by resistivity gradients, piezoelectric effects from packaging stress, and contact misalignment can amount to 10 mT or more. Layout techniques, such as cross-coupled structures to minimize the effects of resistivity gradients, can significantly improve offsets. Careful alignment of the Hall cell layout with crystal axes can mitigate piezoelectric effects. Silicon Hall effect devices are insensitive to magnetoresistive effects as the field strengths encountered in most applications have good linearity with errors of <0.1 percent for fields from 0 to >100 mT.[5]

The Hall voltage is strongly temperature dependent and, with constant current bias, is proportional to the magnetic field, the bias current, the carrier concentration, the scattering factor, and a constant G, which is a function of the geometry of the device.

$$VH = GI\mathbf{B}\left(\frac{r}{\text{qnt}}\right) \tag{3.3}$$

A typical silicon Hall device will exhibit a temperature coefficient of the Hall voltage of approximately 1000 ppm/°C under these bias conditions. The temperature dependence can be reduced with a current source, the temperature coefficient of which is designed to compensate for the Hall voltage TC. In this way, the sensitivity of a Hall effect device can be controlled within 1 or 2 percent over the range of temperatures normally encountered in automotive wheel position and speed applications. The temperature coefficient of the compensated device can be matched to the magnetic circuit if necessary; a typical requirement might be to provide a residual TC of 200 ppm/°C to compensate for the TC of a Sm_2Co_{17} magnet.

Hall effect devices can be constructed from semiconductor materials other than silicon—for example, gallium arsenide (GaAs). GaAs offers higher carrier mobility and some promise for higher temperature operation than silicon. Silicon has the advantage of low cost and the availability of integrated circuit processing techniques that can be used to integrate Hall effect devices with sophisticated signal conditioning. Dielectrically isolated silicon processes,

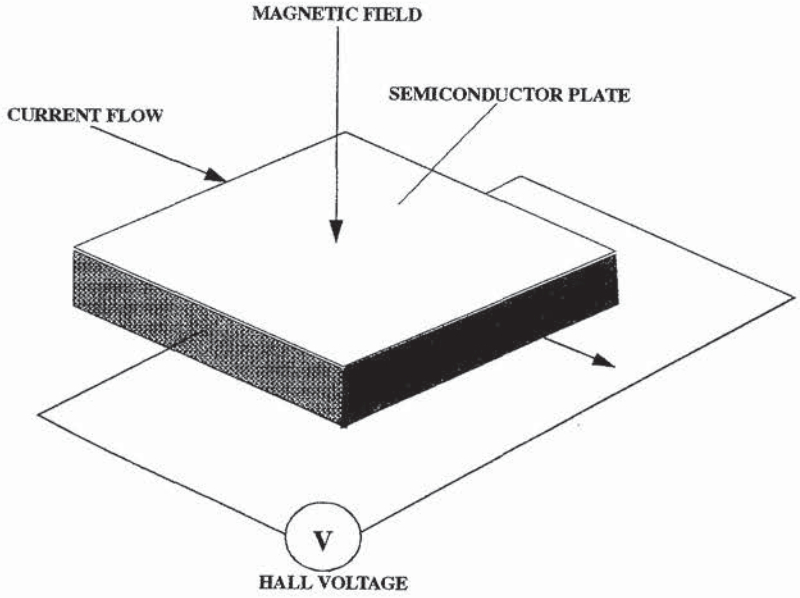


FIGURE 3.5 Hall effect device.

combining integrated circuit techniques with low leakage device isolation, can equal or better the high temperature performance of GaAs.

Hall effect integrated circuits are best categorized in terms of their output characteristics. Analog output devices are usually designed to provide a voltage output that is proportional to the applied magnetic field and also to the power supply voltage. An output ratiometric to the supply allows the device to be easily interfaced to analog-to-digital converters. Analog output devices can be used to construct noncontact absolute position transducers, where the Hall effect device measures a varying field which is designed to be proportional to an angle or linear position. Sensors such as these have no wearing parts other than bearings and can have significant reliability advantages over potentiometers in applications such as throttle-position measurement applications.

Digital output devices are used to construct limit switches or incremental position sensors. The Hall device can be designed to detect homopolar or bipolar fields. Important specifications for digital output devices are operate and release points and the differential between them. The operate point is the maximum field that must be applied to turn the output ON; where ON may be a current-sourcing or current-sinking function. The release point is the minimum field that will guarantee that the sensor is OFF. The differential is the difference between the actual operate and release points. The differential is built in to provide some hysteresis or noise margin to prevent false triggering, particularly at low rates of change of field. The differential may be considerably smaller than the difference between the specified operate and release points. Unipolar devices are specified with operate and release points of the same sign. Bipolar devices are specified with a positive operate point and a negative release point. A caution here is that some devices specified as bipolar do not always require a change of phase of field to operate and release; truly bipolar devices always do.

High-performance Hall effect ICs employ various circuit techniques to improve sensitivity.[6,7] Differential Hall sensors designed for use as gear wheel position sensors use two Hall cells ideally separated by half the gear tooth pitch. This kind of sensor is capable of detecting small changes in unipolar fields. Differential sensors produce an output pulse whose width

depends on the rate at which the gear tooth passes the sensor. At very low and very high speeds a very small mark space ratio output results. At high speeds, the timing of the output will be delayed from the mechanical event by a significant proportion of the tooth pitch. A second method is to use a filter circuit to determine the average value of an alternating field and then detect variations from the average value. This method also eliminates any offset that the sensor may have. This method more accurately tracks the mechanical stimulus. A disadvantage is that the filtering function imposes a lower limit on the speed that can be tracked. The devices detailed in Refs. 6 and 7 have lower limits of around 4 Hz. An additional limitation is that the sensor may fail to operate with a high mark space ratio stimulus since the average value of the input will be close to the value of the longest part of the cycle.

A typical Hall sensor application is shown in Fig. 3.6. The Hall device is assembled into a probe with a biasing magnet. The Hall device orientation will depend on its mode of operation, whether unipolar or bipolar. In all cases, the device is sensitive to fields perpendicular to the plane of the silicon surface. Figure 3.6 also compares output waveforms of the Hall sensor configurations discussed earlier as they would appear in this application.

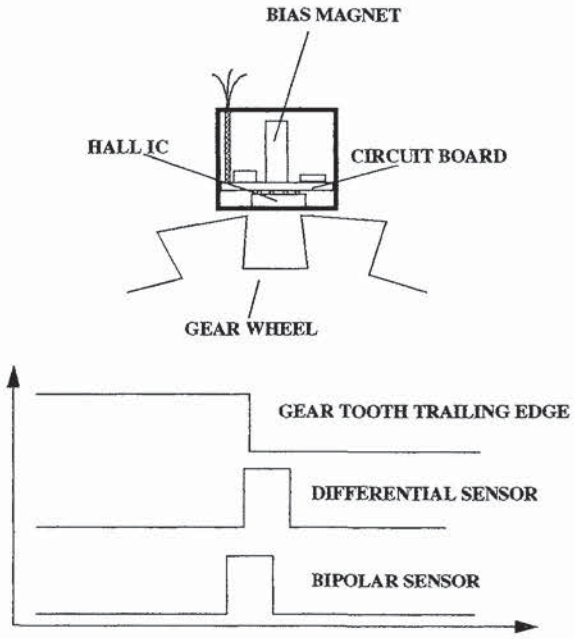


FIGURE 3.6 Hall probe gear position sensor.

Inductive Angle Transducers. Synchro resolvers, or simply resolvers, are absolute angle transducers. Due mainly to their construction, modern brushless resolvers offer the most rugged, reliable, and highest-resolution solution to angle sensing. Resolvers are often considered a high-cost transducer for automotive applications due to the high labor content in the production of most variants. Some designs[8] provide a cost-effective solution by employing sensing and output windings that can be produced on conventional armature-winding machines. Resolvers can be obtained either fully enclosed or as "pancake" devices, with the stator and rotor supplied separated to facilitate over shaft mounting. Resolvers are often referred to by their size. This is the diameter of the case of the device in inches, rounded up to the nearest 10th, and multiplied by 10. For example, a size 11 resolver will be a fraction less than 1.1 in in diameter. Resolver accuracies are specified in arc min. Typical values for

accuracy lie in the range 7 arcmin, with more or less accurate versions available from some vendors.[9]

Resolvers are basically rotating transformers. The construction of a typical device and the output waveforms for a 360° rotation are shown in Fig. 3.7. An alternating voltage connected to the reference input provides primary excitation. The range of frequencies used can be 400 to 20 kHz depending on the construction of the resolver; most resolvers are optimized for the 2 to 5 kHz frequency range. The reference signal is coupled to the rotor via a transformer mounted at one end of the rotor shaft. A second rotor winding couples to two orthogonally oriented stator windings. The stator coils are wound so that as the rotor shaft rotates, the amplitudes of the outputs of the stator windings vary as sine and cosine of the shaft angle relative to some zero.

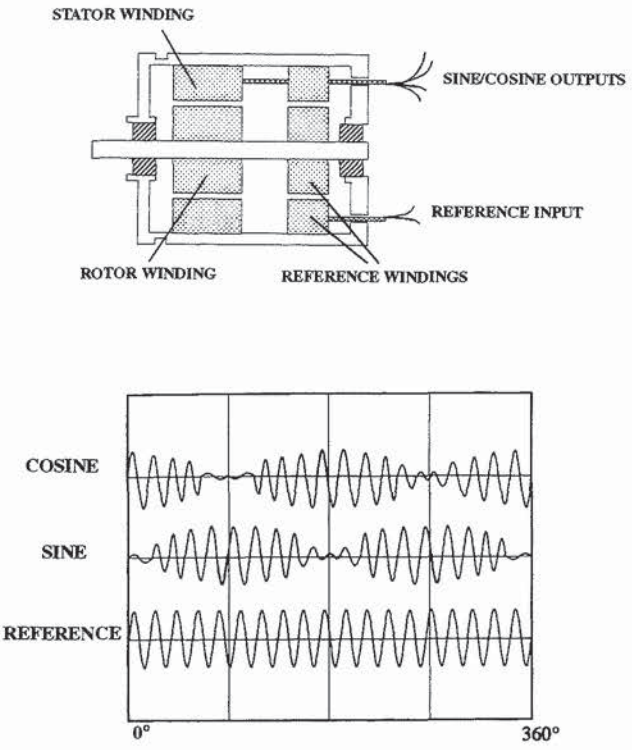


FIGURE 3.7 Resolver construction and output format.

By far, the most economical way of decoding the output of a resolver is to use an IC resolver-to-digital converter (R to D). A functional block diagram of a typical converter is shown in Fig. 3.8. The sine and cosine amplitude-modulated input signals from the resolver representing a shaft angle θ are multiplied by the cosine and sine, respectively, of the current value ϕ of the up/down counter. The resulting signals are subtracted giving:

$$V\mathbf{E} = A \sin(\omega t) \sin(\theta - \phi) \tag{3.4}$$

where $A \sin(\omega t)$ represents the reference carrier.

This signal is synchronously demodulated and an integrator and voltage-controlled oscillator form a closed loop with the counter/multiplier, which seeks to null $\sin(\theta - \phi)$. When the

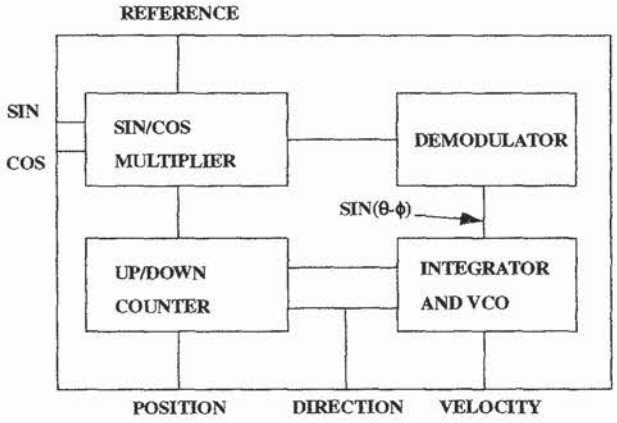


FIGURE 3.8 Tracking resolver-to-digital converter.

null is achieved, the counter value represents the resolver shaft angle within the rated accuracy of the converter. IC R-to-D converters are available that provide parallel or serial digital outputs with resolutions of from 10 to 16 bits and accuracies from 2 to 30 arcmin. Versions are available[10] that emulate standard optical encoder outputs for applications where absolute position measurement is not required but the environment is too harsh for an optical encoder. The system described is a type 2 servo loop which is characterized by zero position and velocity error (not including the limitations of the amplifiers in the IC). R-to-D converters of this type also provide a signal proportional to the angular speed of the resolver from zero to some upper limit, typically 1000 s of rpm, depending on the converter characteristics.

Inductive potential dividers are another class of transducers that are available in many variants.[11,12] A good example is the *rotary variable transformer* or ROVAT.[12] This device comprises a single coil wound on a circular ferromagnetic stator with a number of teeth wound as alternate polarity poles. The stator is excited with an AC signal of around 20 kHz. A rotor with a semicircular conductive screen on its inside surface encircles the stator. The screen reduces the flux linkage between the rotor and stator and the inductance of the screened portion of the stator is reduced, reducing the voltage drop across this portion of the stator. The voltage measured at a central tap on the stator is linearly proportional to the angle of the rotor. Further taps at 90° and 270° from the nominal zero allow a waveform to be measured with amplitude in quadrature with the signal measured at the center point. This allows a 360° absolute angle transducer to be realized, using decoding techniques similar to those which will be described later for the LVDT.

Inductive Linear Displacement Sensors. Shading ring or short-circuit ring sensors are absolute displacement sensors consisting of an E-shaped core with a winding on the central leg of the E. The winding is excited with high-frequency alternating current. An electrically conductive ring of Al or Cu is allowed to slide, maintaining an air gap along the central leg. The ring is attached to the mechanical component, whose position is to be measured. The ring is equivalent to a short-circuited secondary turn in a transformer. The ring has a shading effect, preventing any flux coupling between the legs of the core from its position along the central leg of the core to the open ends of the core (Fig. 3.9).[13] An inductance change can be measured at the terminals of the excitation coil. These sensors are usually used in a potential divider configuration with a reference inductance of similar construction to the main sensor connected in series with the main sensor. A reference alternating voltage is applied across the series-connected reference and sensor inductances, forming an inductive potential divider. The output is then proportional to the ratio of the inductances. This renders the sensor insensitive to tem-

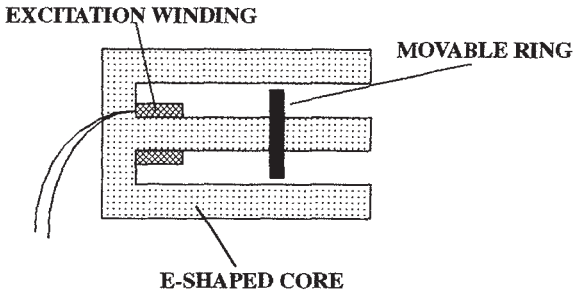


FIGURE 3.9 SCR sensor construction.

perature variations and allows easy adjustment of offsets. Signal-processing electronics can be used to rectify and filter the output and transmit the result to a remote control unit. An alternative construction is to replace the movable ring with an angled channel which can move, relative to the E-shaped core, in a plane perpendicular to the core. Again, the inductance of the sensor is proportional to the linear movement of the angled channel.

Another form of absolute linear displacement sensor is the linear variable differential transformer or LVDT.[14] LVDTs are rugged and reliable and capable of working in harsh environments. Suitable automotive applications include mounting inside hydraulic cylinders in suspension control systems.

LVDTs are constructed from a primary excitation coil positioned centrally on a cylindrical hollow former. Two identical secondary coils are positioned on either side of the primary. The coils have a common core which is free to move within the cylindrical former (Fig. 3.10). The secondaries are normally connected in series, with opposing phases such that with the core centrally positioned and coupling equally to each secondary, the voltage at the node common to both coils will be zero. With this connection, as the core is moved from one extreme of travel through the center to the other extreme, the output signal will vary from a maximum value in phase with the excitation through zero to a maximum value in antiphase with the excitation.

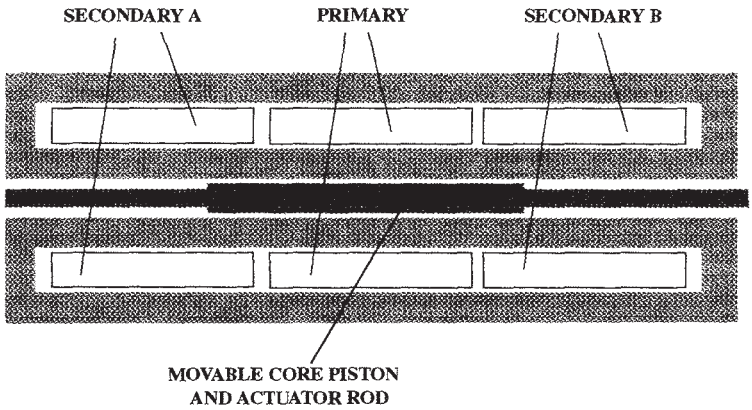


FIGURE 3.10 LVDT construction.

LVDTs are designed to give a linear output within some tolerance, typically ±0.25 percent, over a specified proportion of the available stroke length. The distribution of turns on the secondary coils is carefully arranged to maximize linearity over the widest possible range. LVDTs are available that maintain good linearity with stroke lengths from ±0.05 to ±10 in.

LVDTs operate with effective transformer ratios of between 10:1 to 2:1. The range of primary excitation frequencies can be from 20 Hz to 20 kHz, depending on the construction of the device. Most LVDTs are optimized for the 2- to 5-kHz frequency range. The output signal from an LVDT can be decoded in several different ways and a number of analog and digital integrated circuit solutions exist for this purpose.[15] An example of a typical connection scheme using an LVDT-to-digital converter is shown in Fig. 3.11. In this example, it is assumed that the sum of the voltages across the series-connected secondaries $V_A + V_B$ is a constant over the range of displacements of interest. The majority of LVDTs in production meet this criterion; for those that do not, an additional nonlinearity will result. The IC decodes the function $(V_A - V_B)/(V_A + V_B)$ over the range $[(V_A - V_B)] \leq (V_A + V_B)/2$ into a 13-bit digital word that can be accessed via a three-wire serial interface. Additional bits indicate null and over or under range for signals outside the linear range. The ratiometric decoding scheme described here is insensitive to primary-to-secondary phase shifts, temperature, and any residual null voltage that the transducer may have due to stray capacitive coupling.

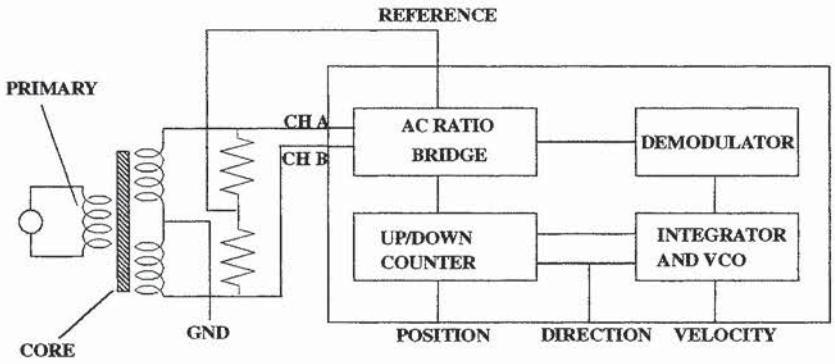


FIGURE 3.11 Tracking LVDT-to-digital converter.

Analog methods of decoding utilize the same basic algorithm as the converter already described. An analog decoder senses the secondary output voltages and evaluates the ratiometric function introduced earlier. The decoder output is filtered and amplified to produce an output voltage proportional to the position of the movable core.

Magnetoresistive. An interesting group of sensors utilize the property of some FeNi alloys such that their resistivity is strongly affected by the presence of a magnetic field. The magnetoresistive effects of one of the useful alloys, Permalloy, which is 81 percent nickel and 19 percent iron, enables sensitive magnetic field sensors with full-scale fields of 5 mT to be built. The variation in resistance is around 2.5 percent for a field of this magnitude. The resistance decreases with increasing field strength; the relationship between field strength and resistance is very nonlinear, approximating a cosine-squared function. Using thin films of Permalloy deposited on a silicon substrate allows signal-conditioning electronics to be integrated with the sensor. Despite the nonlinearity of the phenomenon, accurate linear sensors can be constructed by using the sensor in a bridge arrangement together with flux nulling means, such as a servo-driven coil surrounding the sensor, to effectively operate the sensor at constant resistance. An alternative construction uses opposing "barber pole" thin-film elements.[16] An element is composed of a rectangular thin film of Permalloy overlaid with a series of shorting

stripes of aluminum at 45° to the long axis. Two series-connected elements, which are mirror images of each other (reflected about the long axis), form a potential divider. Magnetoresistive sensors generally exhibit high sensitivity, but this leaves them prone to interference from unwanted fields and, therefore, they are unsuitable for some applications.

Magnetostrictive. Magnetostriction is a property of materials that respond to a change of magnetic flux by developing an elastic deformation of their crystal structure. Magnetostrictive linear displacement sensors utilize this phenomenon by launching a compression wave down a cylindrical waveguide using electromagnetic means, usually a current pulse. The waveguide passes through a movable permanent magnet ring at some distance from a receiver site. The compression wave generated at the magnet position travels to the receiver site at approximately 2800 m/s where it causes a change of flux and generates a voltage pulse in a sense coil. The time of flight of the pulse can be measured to determine the distance of the movable ring magnet down the wire. Transducers with stroke lengths in excess of 7.5 m are available that use this technique.

3.3.5 Other Technologies

A number of other technologies can be applied to position-sensing problems, limited only by engineering ingenuity. Some, like capacitive-sensing techniques, are not tolerant of the automotive environment due to sensitivity to humidity, vibration, or temperature or pressure extremes. Others, like resolvers of traditional construction, are sufficiently rugged but prohibitively expensive unless low-cost manufacturing techniques can be found. Occasionally, a nonobvious method finds a niche. An example is a fuel-level sensor disclosed by workers at Bosch,[17] which, in principle, could be applied as a position sensor. The device operates by exciting a metal rod with acoustic waves such that it resonates. One end of the rod is immersed in the fuel. The resonant frequency is a function of the depth of immersion and the fuel level can be determined by suitable electronics.

A significant influence in the selection of technologies for automotive use is the mandatory inclusion of safety systems. Microwave or laser-ranging techniques can be applied to anticollision systems to anticipate obstacles at nighttime or in poor-visibility driving conditions. These will be a ubiquitous component of automobiles in years to come.

3.4 INTERFACING SENSORS TO CONTROL SYSTEMS

All sensors, whether they have digital or analog outputs, provide measurement of real-world phenomena that are then interpreted by another system to either indicate a value, a warning, or close a control loop. It is vitally important that the integrity of the data is maintained by the proper choice of interface.

Cost and reliability in automotive systems are primary drivers in the choice of interface. For a given performance, the sensor that requires the fewer connections will always be selected. In some cases, such as LVDTs where the basic sensor requires 5 or 6 wires, it may be advantageous to locate the signal-conditioning electronics with the sensor and communicate the processed data to a remote controller via a simple serial interface. Interfacing incremental or serial binary data to a microcontroller is straightforward. In the case of incremental data, typically the mechanical system being monitored is moved until some limit or index is detected. Knowledge of the absolute position of the mechanical component is then known. A counter or register can be set to an initial nonzero value or reset to zero. As the motion is detected, pulses from the incremental sensor can be counted and stored as a measure of position. The indexing cycle must be performed each time power is applied. Binary serial data can be read into a register and used directly with no further processing.

Analog-to-digital converters (ADCs) require some special considerations to optimize performance. Not least of these is grounding. In most applications, chassis ground returns cannot be used. Modern automobiles may have voltage drops of 1 V or more between the chassis at the ECU and the sensor site, due to return currents from electric equipment. This voltage is likely to be noisy with many transients and will certainly upset all but the crudest sensors.

Previous sections have discussed the advantages of ratiometric sensors that can use the same reference as the converter. The advantage is that a least-significant bit of the ADC is always a fixed percentage of the sensor span. This eliminates gain, offset, and temperature errors that may occur if separate references are used. Resolution and gain accuracy of a system can be further optimized by making certain that the span of the sensor output uses all of the available input span of the converter. Many ADCs include on board a microcontroller and use switch capacitor techniques to acquire the analog input values. These present a transient load to the sensor once or twice per conversion cycle. Some sensor outputs, particularly sensors with buffer amplifiers, require isolation from this transient to achieve rated accuracy. A simple technique is to use a simple RC filter on the output of the sensor. This limits the transient current that the sensor output sees and shunts the ADC input with a capacitor.

GLOSSARY

Absolute output sensor The sensor output is an unambiguous measure of position and is valid when power is applied.

Arcminute An angular measure. There are 60 arcminutes in 1 degree of arc.

Incremental sensor The sensor indicates changes in position. An additional position reference, such as a limit switch, is often used with this type of sensor.

Linearity error The amount by which the sensor output differs from an ideal characteristic. Usually expressed in percent.

Ratiometric output sensor An input stimulus causes the output to be a fraction of a reference voltage.

REFERENCES

1. Hewlett-Packard, *Optoelectronics Designers Guide,* San Jose, Calif., 1991–1992.
2. Novatechnik Position Sensor Data, Ostfildern, Germany, 1992.
3. P. Hammond, *Electromagnetism for Engineers,* Pergamon Press Ltd., Oxford, England, 1965.
4. Roland K. Kolter, European Patent Application EP 0 019 530 A1, Applicant Bendix, 1980.
5. Henry P. Baltes, and Popovic, Radivoje S., "Integrated magnetic field sensors," *Proceedings of the IEEE* vol. 74, no. 8, Aug. 1986, pp. 1107–1132.
6. Hartmut Jasberg, "Differential Hall IC for gear-tooth sensing," *Sensors and Actuators,* A21–A23, 1990, pp. 737–742.
7. AD22150 Data Sheet, Analog Devices, Norwood, Mass.
8. Charles S. Smith, U.S. Patent 4 962 331, Assigned to Servo-Tek Products Co. Inc., Hawthorne, N.J., 1990.
9. Clifton Precision, Analog Components Data, Clifton Heights, Pa.
10. AD2S90 Data Sheet, Analog Devices, Norwood, Mass.
11. Novatechnik Angle Sensor Data, Ostfildern, Germany, 1993.

12. Donald L. Hore, and Flowerdew, Peter M., "Developments in inductive analog transducers for 360° rotation or tilt and for linear displacement," *IEE International Conference,* No. 285, 1988.
13. E. Zabler, and Heintz, F., "Shading-ring sensors as versatile position and angle sensors in motor vehicles," *Sensors and Actuators* **3**, 1982/3, pp. 315–326.
14. *Schaevitz Linear and Angular Displacement Transducers Catalog,* Pennsauken, N.J.
15. AD2S93 and AD598 Data Sheets, Analog Devices, Norwood, Mass.
16. F. Heintz and Zabler, E., "Application possibilities and future chances of 'smart' sensors in the motor vehicle," *SAE Technical Paper Series* #890304.
17. E. Zabler, "Universal low-cost fuel-level sensor," Robert Bosch GmbH, Ettingen, Germany.

ABOUT THE AUTHOR

Paul Nickson is product line manager for Analog Devices' Sensor and Automotive Group in Wilmington, Mass. He has BSc Honours in Electronic and Electrical Engineering from the University of Birmingham, England and 16 years' experience in integrated circuit design. For the past five years, he has focused on silicon sensors and sensor signal conditioning.

CHAPTER 4
FLOW SENSORS

Robert E. Bicking
Senior Engineering Fellow
Honeywell, Micro Switch Division

4.1 INTRODUCTION

Measurement of flow rate is important to optimize the performance of several key engine control subsystems. Mass air flow sensors are replacing the indirect calculation of intake mass air flow for improved performance, driveability, and economy. New requirements for on-board diagnostics are opening new applications for flow sensing in the automobile.

If the parameter to be measured is a gaseous mass flow as opposed to a volume flow, this further focuses the sensing technology selection since only a few technologies inherently measure mass flow. For liquid flow, either a mass flow or volumetric flow approach may be used since the density of a liquid changes only a small amount with atmospheric pressure and temperature.

This chapter is intended to give the reader an understanding of where and why flow sensors are being specified for use in engine control systems and an understanding of the trade-offs among alternative technologies in particular applications.

4.2 AUTOMOTIVE APPLICATIONS OF FLOW SENSORS

4.2.1 Intake Air Mass Flow

Electronic fuel injection has almost universally replaced the carburetor in the auto engine. This is because it provides better performance and is the only way to meet government mandated emissions standards. To do fuel injection, the mass flow rate of air going into the engine must be determined. There are two competing approaches to determining the mass air flow rate. The first, *speed density,* calculates the mass flow rate by measuring engine speed (RPM), intake air temperature (T_a) and intake manifold pressure (P). The per-cylinder nominal displacement (V) is known, as is the gas constant of air (R_a). The engine's volumetric efficiency (η) may be modeled as a function of speed. Then, the mass flow rate ($\dot{m}_a$) is calculated as follows:

$$\dot{m}_a = \frac{(\mathrm{RPM}\ V\ \eta\ P)}{(R_a T_a)} \tag{4.1}$$

By using a mass flow sensor, this calculation is eliminated and $\dot{m}_a$ is measured directly. The advantages of using a direct measurement include better accuracy under dynamic conditions because the manifold pressure changes more slowly than the mass flow rate. Also, no assumptions are made about engine displacement or volumetric efficiency, both of which affect the speed density calculation. Volumetric efficiency can change as the intake system becomes contaminated. Mass air flow sensors are presently being used on roughly 20 percent of the cars sold in the United States. They are not more widely used because they cost more than the pressure sensor needed to do the speed-density calculation. The largest drawback to using a mass flow sensor is that a wide dynamic range sensor is needed since the mass flow rate can change over a 100-to-1 range from idle to full throttle, whereas the manifold pressure only changes over a 5-to-1 range. Other important factors in sensor selection include resistance to contamination and particulate damage, accuracy, ability to measure reverse flow, and sensitivity to upstream and downstream ducting.

Fuel is injected to each cylinder on a sequential basis for optimum performance. This requires using the crankshaft sensor and camshaft sensor outputs to time the injection of fuel into the intake manifold at the proper point on the intake stroke. The fuel injector is fed with a constant pressure and is pulse-width modulated to control the amount of fuel injected. The calculation of the mass rate ($\dot{m}_f$) of fuel injected follows:

$$\dot{m}_f = \dot{m}_a \lambda \tag{4.2}$$

where $\dot{m}_a$ = mass air flow rate
λ = stoichiometric air/fuel ratio

4.2.2 Potential or Future Applications of Flow Sensors

Fuel Flow for Gas Mileage Measurement. Driver information systems that predict range and gas mileage need to know the amount of fuel flow. As seen from the preceding discussion of mass air flow measurement, the fuel flow is calculated from the air flow so that this data is already in the engine control computer. By simply summing the injector "on" time over a revolution or multiple revolutions, the fuel flow is known.

Alternatively, the fuel flow can be measured by taking the difference between the fuel coming into the fuel rail and that being returned. This approach could be subject to substantial error, however, due to the fact that these flows are purposely much larger than the net flow into the engine so that the fuel rail will be maintained at a constant pressure. Future fuel-handling systems may simplify the measurement by eliminating the return line and simply modulating the fuel pressure pump to maintain constant pressure. Then, the measured flow would be simply that flow going into the engine.

Exhaust Gas Recirculation Flow. Exhaust gas recirculation (EGR) is performed to reduce the emission of nitrous oxides (NO_x) by cooling the combustion process. If the EGR valve begins to clog or only partially opens, its flow will be reduced and emissions will increase. In the near future, OBDII legislation will require cars to have on-board diagnostics capable of determining when an emissions-related failure (such as that just described) occurs. Measurement of the flow is one way to diagnose a faulty EGR valve. Another way would be the use of an NO_x sensor to measure the emissions. However, no low-cost NO_x sensors have yet been developed.

Secondary Air Pump Flow. The secondary air pump is used to reduce the emissions of carbon monoxide (CO) and hydrocarbons (HC). Measurement of its flow rate is an approach to verifying that it is operating properly and doing its part to reduce emissions. Another way, of course, would be to directly measure the emissions of HC and CO in the exhaust. To date, however, there is no low-cost real-time means to do that. A way to verify air pump operation

without using any additional sensors is to command the air pump full-on when the fuel injection is slightly rich of stoichiometry. The oxygen sensor output should switch from rich to lean if the air pump is operating properly.

Fuel Flow for Fuel-Air Ratio Feedback Control. Present engine control systems treat the fuel flow as a dependent variable, measuring the intake air flow and then operating the fuel injectors to meter the proper amount of fuel into the engine. This tacitly assumes that each of the injectors is precisely calibrated so that a given "on" time provides a given amount of fuel. Fuel injectors typically use a needle valve in an orifice to meter the fuel. The orifice area is proportional to the square of the diameter so that the fuel flow error will be proportional to two times the diameter tolerance. By measuring the fuel flow, the accuracy of fuel injection could be improved. Table 4.1 summarizes the performance requirements by application.

TABLE 4.1

Application	Measurement type	Range, kg/h	Accuracy, %
Intake air	Mass	10–1000	±4
Fuel flow	Mass/vol.	1–66	±10
EGR flow	Mass	30–100	±10
Air pump flow	Vol.	50	±20
Fuel flow	Mass/vol.	1–66	±4

4.3 BASIC CLASSIFICATION OF FLOW SENSORS

4.3.1 Energy Additive or Energy Extractive

Flowing fluid possesses energy, both potential and kinetic. One approach to flow measurement extracts energy from the flow. Alternately, energy may be added to the flow and its effect observed. The energy-additive approach typically is nonintrusive, so the act of measuring doesn't affect the flow. As might be expected, flowmeter selection involves a number of factors, as will be discussed later.

4.3.2 Measurement of Mass Flow or Volume Flow

Intake air measurement requires that the mass of air flowing into the engine be measured. This favors a mass-flow approach, since, otherwise, pressure and air temperature must be measured to calculate the mass flow from the volume flow. The mass flow rate ($\dot{m}_a$) is calculated as follows:

$$\dot{m}_a = \dot{V}_a \rho \tag{4.3}$$

where ρ = air density
$\dot{V}_a$ = volumetric flow rate

The density, in turn, is calculated as follows:

$$\rho = \frac{P}{ZR_aT_a} \tag{4.4}$$

where P = air pressure
Z = compressibility factor
R_a = gas constant for air
T_a = air temperature

4.4 APPLICABLE FLOW MEASUREMENT TECHNOLOGIES

4.4.1 Gaseous Flow

Vane. The vane simply consists of a vane or paddle which is located in the flow duct and is restrained by a spring so that it blocks the duct with no flow. The deflection of the vane is thus proportional to flow. This deflection is read out by a potentiometer. The largest drawback of a vane sensor is that it increases the pressure drop in the intake tract, reducing the volumetric efficiency of the engine. This approach was used in the earliest engine control systems and is being replaced by technologies such as hot-wire air flow sensors which offer much lower pressure drop.

Thermal. This is the favored approach and is used in all engines currently being manufactured that employ direct measurement of the intake air mass. Depending on design details, it provides a nearly direct measurement of the mass flow and thus simplifies the engine control strategy. A variety of designs exists, from the straightforward hot-wire air flow sensor to more complex schemes. The basic idea is to heat a fine wire, and then as the gas flows past the wire, convection removes heat. The amount of heat removed can be measured with an electronic circuit and is proportional to mass air flow rate as the following equation shows:

$$\Delta P \cong \Delta T\,[C_t + (2\pi d C_v \dot{m}_a)^{1/2}]^* \tag{4.5}$$

where ΔP = change in electric power due to a given flow rate
ΔT = temperature difference between air and sensor
C_t = thermal conductivity of air
d = diameter of hot wire
C_v = thermal capacity of air
$\dot{m}_a$ = mass flow rate of air

Note that the first term of the equation isn't proportional to flow rate. Either this needs to be modeled and removed or the change in ambient air temperature needs to be minimized to accurately measure mass flow rate. A hot-wire air flow sensor and its control circuit is shown in Fig. 4.1. Control circuits typically either supply constant power to the heated element or

* From Joseph P. DeCarlo, *Fundamentals of Flow Measurement,* Instrument Society of America, 1984, pp. 173 and 176.

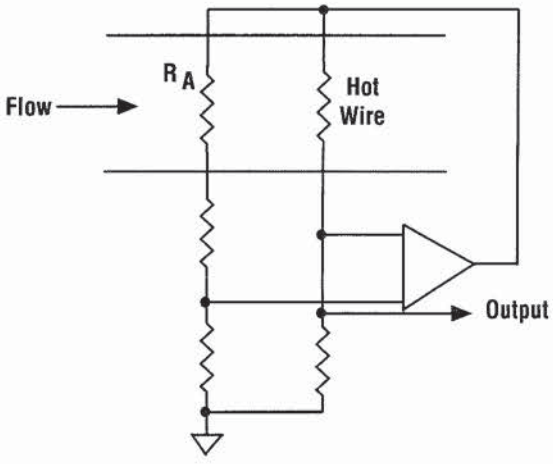


FIGURE 4.1 Hot-wire air flow sensor.

operate it at a constant temperature delta above the ambient temperature. The latter is shown in Fig. 4.2 and is preferred because it simplifies temperature compensation.

One of the problems with the hot-wire sensor is that fine dust particles may pass through the air filter and, under full-throttle conditions, can impact the hot wire with sufficient force to break it. Particulate build-up is also a source of error and has been alleviated by placing the heated element in a bypass channel or by incorporating a burn-off cycle at power-on. Most of the development in hot-wire sensors has been to ruggedize them to withstand the automotive environment and to desensitize them to upstream and downstream flow anomalies. Figure 4.2 is an example of a bypass design for a hot-element mass air flow sensor. It uses a platinum wire wound on a ceramic mandrel and coated with glass as the sensing element. It is located in a bypass channel away from the main flow to reduce the likelihood of particulate contamination. The bypass exits into the main channel through slots which are intended to desensitize it to backflows and backfires.

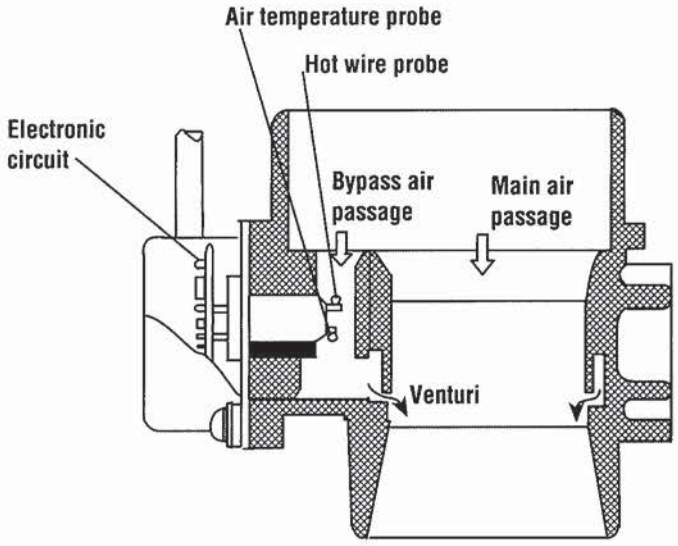


FIGURE 4.2 Bypass hot-wire air flow sensor. *(U.S. Patent 4,264,961)*

Backflow can occur in engines with four or fewer cylinders at low speed. It is most desirable for the sensor to measure the backflow, but most either don't measure it at all or rectify it, which makes the measured flow appear larger than it really is. This error is compensated by measuring the backflow on an engine and then using software to remove the nominal error. One of the issues with bypass designs is that upstream ducting of the air may affect measurement accuracy. A bend just ahead of the sensor will cause the air to move toward the outside of the bend and, depending on where the bypass is located, it may read either too low or too high. This effect is minimized by placing a screen or honeycomb in the sensor to straighten the flow.

Micromachined air flow sensors have been available commercially for several years and efforts are underway to adapt them to automotive applications. Their primary advantages include the low cost of the sensing element due to batch fabrication; excellent performance over temperature due to the close proximity of the heated and reference elements; the ability to measure reverse flows; low operating power due to the small size of the heated element; and fast response, again due to the small size of the heated element. Because of the use of integrated circuit manufacturing techniques, it is no more difficult to include additional resistors on the sensor chip than to simply replicate a hot wire. A calorimetric flow sensor may be constructed by separating the heater and sensor functions as shown in Fig. 4.3. The advantage

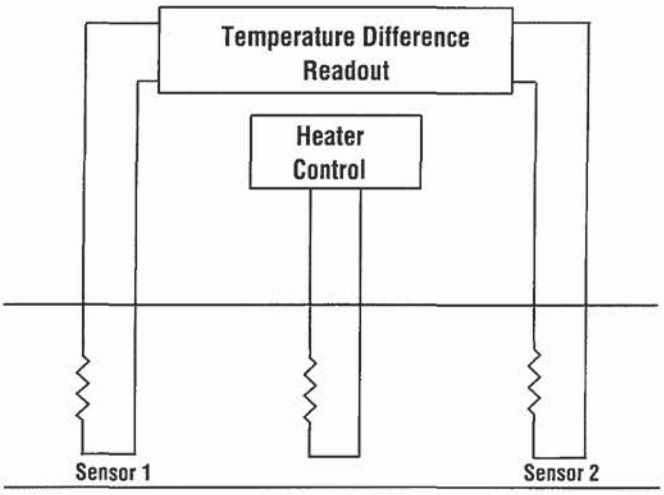


FIGURE 4.3 Calorimetric flow sensor.

of a calorimetric flow sensor is that it does measure the mass flow rate as the following equation shows:

$$\Delta P = \dot{m}_a C_p \Delta T^* \tag{4.6}$$

where ΔP = change in power due to a given mass flow rate
$\dot{m}_a$ = mass air flow rate
C_p = thermal capacity of air
ΔT = temperature difference between air and sensor

One of the issues with microstructure sensors is protection from damage due to particle impact and preventing build-up of contamination. This requires clever packaging.

Table 4.2 compares the main types of thermal mass air flow sensors.

Differential Pressure. A simple way to measure volumetric flow is to place an obstruction in a flow channel and measure the differential pressure drop across it. The flow is proportional to the square root of the differential pressure. This only works well for narrow flow ranges because to operate over a given flow range, the pressure sensor must operate over the square of that range. For the automotive intake flow range of 100 to 1, a pressure sensor with a range of 10,000 to 1 is required, which is not achievable except at high cost. An important advantage of this approach is that it is resistant to contamination, since the pressure sensor merely needs

* From "Fundamentals of Flow Measurement," Joseph P. Decarlo, Instrument Society of America, 1984, pp. 173 and 176.

TABLE 4.2

Type	Effect of contamination	Meas. rev. flow	Affected by ducting	Power used
Hot Wire	High*	Rectifies	Low	Medium
Hot RTD Bypass	Low	Ignores	High	High
Micro RTD	Low	Yes	Low	Low

* Burn-off cycle used to alleviate.

to sense the pressure difference across the flow restriction (in the case of an orifice plate or venturi). The differential pressure approach could be suitable for some low-accuracy or low dynamic range applications such as EGR valve flow. Venturis (Fig. 4.4), flow restrictions (Fig. 4.5), and pitot tubes (Fig. 4.6) all operate on the same principle.

4.4.2 Liquid Flow

Differential Pressure. The preceding discussion on differential pressure sensing applies equally to liquid flows.

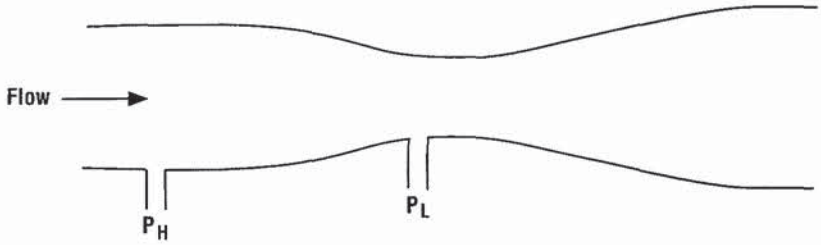


FIGURE 4.4 Venturi flow sensor.

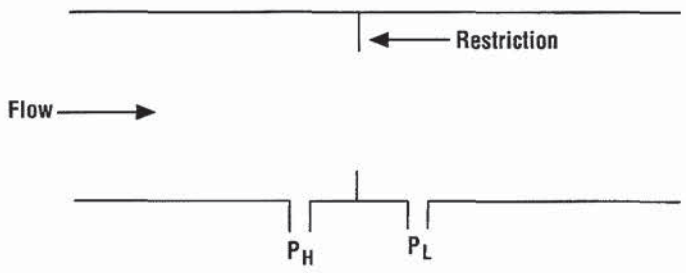


FIGURE 4.5 Flow sensor using a restriction to develop a differential pressure.

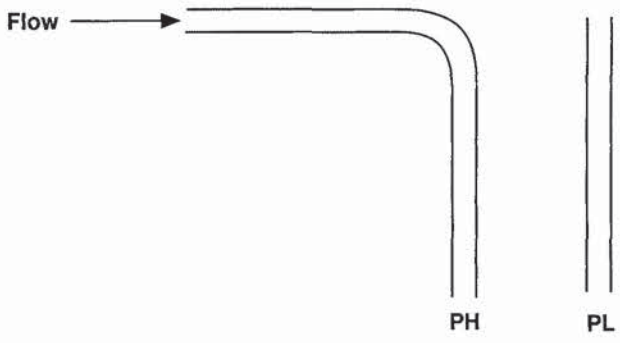


FIGURE 4.6 Pitot tube flow sensor.

Turbine. A turbine blade placed in a flow channel can offer low restriction to flow, and by counting the speed of rotation it can also measure flow. It is analogous to the vane described previously except that it is more subject to wear since it is revolving, not merely deflecting. Use of a noncontacting means of counting revolutions (such as a Hall effect sensor) is recommended. This approach is well-suited to measurement of fuel flow.

Vortex Shedding. Oscillations may be induced in a fluid by placing an obstruction in the flow stream. The oscillations may be measured thermally, by pressure changes or using ultrasonics. This approach doesn't work well at low flow rates due to instability in the vortex shedding mechanism, so it is not suited to wide dynamic range applications such as intake air measurement. An example of a vortex shedding flow sensor is shown in Fig. 4.7.

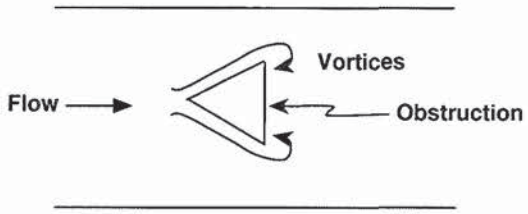


FIGURE 4.7 Vortex shedding flow sensor.

Other Technologies. There are a number of other technologies which are not well-suited to automotive applications, primarily due to cost considerations. A variety of flow sensors has been developed for industrial applications where performance is paramount and cost is secondary. These include the use of ultrasonics, gyroscopic effects, ionization of gases, and so on. These will not be discussed herein. Information on these technologies is given in the books by DeCarlo and Norton. See the list of suggestions for further reading.

GLOSSARY

Backflow The flow of air out of the intake system of an engine. This occurs at low speeds in engines containing only a few cylinders when all intake valves are closed at once.

Exhaust gas recirculation A process in which exhaust gases (containing primarily nitrogen) are returned to the intake system to reduce emissions by lowering the combustion temperature.

Micromachining Formation of very small three-dimensional structures in a semiconductor material.

Speed density A method of calculating the intake mass air flow rate by measuring engine speed and air density.

Stoichiometry The air-to-fuel ratio at which all the oxygen is consumed during combustion.

Volumetric efficiency The volume flow rate of air into the intake system divided by the volume rate of piston displacement.

BIBLIOGRAPHY

Adler, Ulrich, *Automobile Electric/Electronic Systems,* Robert Bosch, Stuttgart, 1988.

DeCarlo, Joseph P., *Fundamentals of Flow Measurement,* Instrument Society of America, Research Triangle Park, 1984.

Heywood, John B., *Internal Combustion Engine Fundamentals,* McGraw-Hill, New York, 1988.

Norton, Harry N., *Handbook of Transducers,* Prentice Hall, Englewood Cliffs, N.J., 1989.

ABOUT THE AUTHOR

Robert E. Bicking received the B.E.E. and M.S.E.E. degrees from the University of Minnesota and has been involved with sensors nearly all of his 32 years in engineering. He is currently involved with the development of new automotive sensors at the Micro Switch division of Honeywell. He has been awarded three U.S. patents and is the author of over 24 technical papers. He is a member of the I.S.A., I.E.E.E., and S.A.E.

CHAPTER 5
TEMPERATURE, HEAT, AND HUMIDITY SENSORS

Randy Frank
Motorola Semiconductor Products

5.1 TEMPERATURE, HEAT, AND HUMIDITY

Temperature sensing and taking into account the effect of temperature on the performance and reliability of automotive components, and therefore the systems that contain these components, is one of the more important aspects of vehicle design. The heat sources that are present on modern automobiles range from the engine itself to catalytic converters, losses in power conversion (e.g. the alternator), and specific heat-generating devices such as heated windshields, seats, and mirrors. Humidity adds to the effect that temperature has on the reliability of components and also impacts vehicle performance and passenger comfort.

5.1.1 Temperature—the Effect of Heat

The temperature of a body or substance is (1) its potential of heat flow, (2) a measure of the mean kinetic energy of its molecules, and (3) its thermal state with reference to its ability to transfer heat to other bodies or substances.

Temperature affects every aspect of automobiles, from the performance of the engine and various vehicle systems to the comfort of the driver and passengers. The wide range of worldwide vehicle operating temperatures (from –60° to +57 °C) and the subsequent localized electronic module temperature ambients underhood (from –40 to +125 °C) and in the passenger compartments (from –40 to +85 °C) affect both the performance and reliability of electronic components. The viscosity of lubricating and cooling fluids is also affected by the wide temperature variations that must be tolerated. Even paint, fabric, plastics, rubber, and other organic and inorganic materials must be designed to survive the environmental extremes of temperature and humidity. Sensing the temperature of these components is essential during vehicle development.

5.1.2 Conduction, Convection, and Radiation

Heat energy is transferred with corresponding temperature changes by conduction, convection, and/or radiation. Conduction occurs by diffusion through solid material or in stationary liquids or gases; convection involves the movement of a liquid or gas between two points, and radiation occurs through electromagnetic waves.

5.1.3 Heat Sources in Vehicles

In addition to the temperature rise that can be generated by sunlight on the metal and glass that form the body of the vehicle, there are several heat-generating devices in a vehicle as shown in Table 5.1. The foremost heat source is the engine in internal combustion engine equipped vehicles. This causes the engine compartment to generally be classified as a +125 °C ambient for electronic components, although considerably higher temperatures are reached within the combustion chamber (≥1000 °C) or on the engine block.

The catalytic converter used to reduce unburnt hydrocarbons and carbon monoxide emissions has a peak catalyst efficiency around 450 °C, minimum operating temperature around 350 °C, and high temperature operation in the area of 1000 °C.

Tire flexure and friction generated between the tires and the road surface are major sources of heat. Moreover, heat is generated by friction between any moving components of a vehicle. Gears and bearings in transmissions, rear axle, and pumps are most notable for generating a significant temperature rise. Brake surfaces also create high temperatures when the brakes are applied.

Even though they do not generate heat, in the process of reducing the temperature of other components, heat exchangers—such as coolant or transmission radiators for the passenger compartment heater and their associated plumbing and electronic heat sinks—have considerably higher than ambient temperatures.

Electric heating elements, including heated back light, windshield, windshield wiper, mirrors, and seats, increase the temperature to improve visibility and driver comfort. Heated windshield wipers can be maintained at 15 °C even when outside temperatures are around −40 °C. In some cases, temperatures in the area of 350 °C are produced by supplemental heating elements for components such as the catalytic converter, the oxygen sensor, and some advanced electric vehicle batteries.

The windings of electric components such as the starter motor, heater motor, alternator, and solenoids are a source of heat and high temperatures, especially in continuously moving, heavily loaded units like the alternator. The resistance in the windings as well as resistance in semiconductor components like alternator rectifiers and power transistors creates a power loss proportional to the square of the current being conducted ($P = I^2R$). In semiconductors, the change in temperature is related to the power dissipated through the thermal resistance.

$$\Delta T = R_\theta P \tag{5.1}$$

where R_θ = thermal resistance in °C/W
ΔT = temperature difference in °C (frequently $T_{junction} - T_{case}$)
P = power in watts

TABLE 5.1 Heat Sources in Vehicles

General category	Example	Max. temperature °C
Engine	Combustion/ignition process	>1000
Catalytic converter	Chemical reaction	>1000
Road/tire friction	Tires	<100
Brakes	Disk/drum	250
Mechanical motion (gears, bearings)	Transmission/rear axle/air pump/power steering pump	200
Heat exchangers	Radiator (coolant, transmission), heater, heatsink	>175
Electric heaters	Windshield, backlight, seats, mirrors	Ta + 25
Electric windings	Motors, alternator, solenoids	<155 (Class F)
Resistors	Ballast resistor	150
Lamps	Headlamps, tail lamps, dash lamps	125
Power transistors	Ignition driver, voltage regulator	Up to 200
Electric vehicle battery	Sodium sulfur	300

Lamps have a significant inrush current (10 · steady state) and a high steady state value which generate high temperatures as well.

Temperature at various locations in the vehicle are shown in Table 5.2.

TABLE 5.2 Temperature at Various Vehicle Locations

Location	Temperature, °C		Slew Rate	Humidity		Frost
	Low	High		High	Low	
Underhood—Engine						
Exhaust manifold	–40	+649	–7 °C/min	95% at 38 °C	0	Yes
Intake manifold	–40	+121	–7 °C/min	95% at 38 °C	0	Yes
Underhood—Dash Panel						
Normal	–40	+121	Open	95% at 38 °C		
Extreme	–40	+141	Open	80% at 66 °C		
Chassis						
Isolated	–40	+85	NA	98% at 38 °C	0	Yes
Near heat source	–40	+121	NA	80% at 66 °C	0	Yes
At drive train temperature	–40	+177	NA	80% at 66 °C	0	Yes

5.1.4 Effect of Humidity

Humidity affects the comfort of the drivers and the passengers as well as the performance of the engine and impacts the reliability of vehicle components. The primary measurement for humidity is relative humidity RH, the ratio of water partial pressure to saturation pressure. Other measurements include dew point (temperature); specific humidity (mixing ratio); the mass of water per unit mass of dry gas; and volume ratio, the parts of water vapor per million parts of air. The dew point is a direct measure of vapor pressure and absolute humidity. Humidity is the water vapor in gas. Water in a solid or liquid is called moisture. High humidity and low dew point can lead to increased moisture content which can affect the operation of vehicle fluids such as fuel and brake fluid and impact solid surfaces like the windshield or brake linings.

5.1.5 Reliability Implications to Electronic Components

The effect of temperature on electronic components was recognized as "one of the most severe environments generally encountered on the automobile." Warm-up time of engine liquids and local ambient temperatures are as high as several tens of minutes, especially in cold weather. A recommended practice, SAE J1211, was published in November 1978 that still serves as a guideline to developing environmental design goals for electronic equipment, especially in the area of temperature variations within various vehicle locations.

Temperature affects the performance and expected life of electronic components. Mechanical stress created by different coefficients of thermal expansion can cause failures in thermal cycling test (air-to-air) or thermal shock (water-to-water) transitions.

The failure of electronic components also occurs because the mechanical durability, solder strength, or glass transition of plastic mold compound is exceeded. Electrical properties including thermal runaway in bipolar transistors and loss of gate control in MOS-gated (Sec. 5.4.6) devices also directly cause failure due to temperature extremes.

The typical failure rate for a semiconductor component can be expressed by the Arhennius equation:

$$\lambda = A \cdot \exp \frac{\theta}{kT} \tag{5.2}$$

where λ = failure rate
A = constant
exp = 2.72
θ = activation energy
k = Boltzmann's constant
$8.62 \cdot 10^{-5}$ eV/°K
T = temperature in °K

The failure rate of semiconductor components typically doubles for every 10 to 15 °C increase in operating temperature. However, increased testing and design improvements have minimized the failures due to specific failure mechanisms. Semiconductor components manufactured in the 90s are orders of magnitude more reliable than those first used in vehicles in the 60s. However, their response time is considerably faster than vehicle liquids and local ambients. For example, the transient thermal response, $r(t)$ $R\theta_{JC}$, which is a reduced level of the thermal resistance (Sec. 5.1.4) based on the transistor operating in a switching mode and being off for a period of time, approaches the dc level within a second. Excessive temperatures can be generated quickly and must be detected within milliseconds to prevent failure.

Accelerated life testing is performed at elevated temperatures to determine the reliability of electronic components and obtain the maximum amount of information in the minimum amount of time. The actual failure mechanisms are a result of the component design and environment in which they operate. Life testing seeks to identify areas that are vulnerable in a particular design and provides minimum acceptable criteria for qualification.

The activation energy is directly proportional to the degree of influence that temperature has on the chemical reaction rate. Temperature acceleration factors for a particular failure mechanism can be expressed as the ratio of the failure rates at two different levels of stress:

$$F_a = \exp\left(\frac{\theta}{k}\right) \cdot \frac{1}{T_r - 1/T_t} \tag{5.3}$$

where F_a = acceleration factor
θ = activation energy
k = Boltzmann's constant, $8.62 \cdot 10^{-5}$ eV/°K
T_r = junction temperature in °K at the rated ambient temperature
T_t = junction temperature in °K at the life test ambient temperature

To determine a component's ability to withstand the adverse effects of temperature and humidity extremes to which the vehicle is exposed, a number of reliability stress tests are included in the qualification procedure. These tests include:

- *High-temperature Storage Life* is performed to accelerate failure mechanisms which are thermally activated through extreme temperatures.

  Typical test conditions: T_a = 70 to 200 °C, no bias, time = 24 to 5000 h
- *Temperature Cycling* (air-to-air) evaluates the ability of a device to withstand both exposure to temperature extremes and transition between temperature extremes. This test also exposes excessive thermal mismatch between materials.

  Typical test conditions: T_a = –65 to 200 °C, cycle = 10 to 4000 (1000 most common)
- *Thermal Shock* (liquid-to-liquid) testing evaluates the ability of the device to withstand both exposure to extreme temperature and sudden transitions between temperature extremes. This testing also exposes extreme thermal mismatch between materials.

  Typical test conditions: T_a = 0 to 100 °C, cycle = 20 to 300 (100 most common)
- *H^3TRB* (high-humidity, high-temperature reverse bias) is an environmental test designed to measure the moisture resistance of plastic encapsulated devices. A bias is applied to accelerate corrosive effects on internal components of semiconductors.

  Typical test conditions: T_a = 85 to 95 °C, RH = 85 to 95%, bias = 80 to 100% of max rating, time = 96 to 1750 h

- *Autoclave* (or pressure cooker) testing is performed to measure device resistance to moisture penetration and resulting galvanic corrosion. This is a highly accelerated and destructive test.
  Typical test conditions: $T_a = 121$ °C, RH = 100%, $p = 1$ atmosphere, time = 24 to 96 h
- *Moisture Resistance* testing evaluates the moisture resistance of components under tropical environments.
  Typical test conditions: $T_a = -10$ to 65 °C, RH = 80 to 98%, time = 24 h/cycle, cycle = 10

In addition, a number of high-temperature tests under different biasing conditions are also performed for semiconductor devices, including high-temperature forward bias, high-temperature reverse bias, and high-temperature gate bias.

5.2 AUTOMOTIVE TEMPERATURE MEASUREMENTS

Numerous temperature measurements are made during the development of the vehicle to ensure the proper operation of systems and components. Some measurements are included in the control system strategy or as diagnostics for production vehicles. A list of the potential temperature and humidity measurements that can be made and may be incorporated in future systems is shown in Table 5.3. The technique used to measure temperature may vary from development to production mode, from one vehicle manufacturer to another and even from system to system within a given vehicle manufacturer. The choice in technology for production sensors depends upon performance, reliability, and cost. For development vehicles, cost is not a major concern, but availability of test probes, compatibility with readout equipment, and ease of use are additional factors that should be taken into consideration.

Common vehicle temperature measurements will be explained in the remainder of this section.

5.2.1 Measuring Liquid Temperatures

A number of liquid temperatures are measured during vehicle development, especially under high vehicle ambient temperatures and severe driving conditions, such as trailer towing up steep grades and stop-and-go Phoenix traffic during summer months. These include coolant, engine oil, transmission oil, fuel, power steering, brake, and battery electrolyte. Monitoring coolant temperature is common in engine control systems in production vehicles. Engine and transmission oil temperature are being evaluated for future monitoring. Racing vehicles frequently monitor oil and fuel temperature in addition to coolant temperature to achieve peak performance. Sensing fuel temperature may be required to meet future emissions standards. A latent heat cooling system has been developed that uses a temperature sensor to control the engine under various operating conditions. The system employs control valves that control the flow of vaporized coolant based on temperature in a manifold connected to the top of the combustion chamber.

Mounting location, fluid contacting the sensor, and packaging are critical concerns for liquid temperature sensors. Upper temperature limits are typically around 150 to 200 °C and operation down to –40 °C is usually required.

Battery temperature measurements in vehicle development use glass thermometers or glass-shielded thermocouple probes to protect the sensor from the corrosive electrolyte.

5.2.2 Battery Temperature

Maintaining the proper state of charge in the vehicle's battery is essential for obtaining adequate cranking rpm during starting and for ensuring optimum battery life. The charge accep-

TABLE 5.3 Sensing Requirements versus Vehicle Systems

System	Parameter
Engine control	Air charge (intake air) temperature Engine coolant temperature Intake air humidity Ambient temperature Fuel temperature
HVAC (climate control)	Humidity in passenger compartment (PC) Temperature (PC) Outside air temperature
Electronic transmission (continuously variable transmission)	Transmission oil temperature
Driver information (body computer inputs)	Coolant temperature Ambient air temperature Temperature (PC) Tire surface temperature Rain sensor Windshield moisture Sun sensor Battery temperature
Multiplex/diagnostics Cruise control Idle speed control Memory seat Navigation	Multiple usage of sensors
Antiskid brakes	Brake fluid temperature
Traction control/ABS	Brake moisture sensor
Air bags	
Suspension	
Electronic power steering	Power steering fluid temperature (also electric assisted)
Four-wheel steering Security and keyless entry	

tance curves of lead-acid batteries require that the charging voltage be modified for temperature. A higher voltage is required at lower temperatures. Cold temperatures also place the most difficult requirements on the battery because the viscosity of engine oil is low and the load on the cranking system is very high.

Compensation circuitry in the voltage regulator in the charging system is designed to provide a voltage that is within an acceptable range over the entire vehicle operating temperature range as shown in Fig. 5.1. Semiconductor sensing techniques (Sec. 5.4.6) are used in this approach. The most desirable location for the voltage regulator is close to the battery. However, some systems integrate the regulator inside the alternator and provide a temperature environment that can be considerably different from the battery. The temperature of the battery can also vary by several degrees from cells closest to the engine to those further away from the engine. The overall mass of the battery prevents these temperatures from changing very quickly. However, a failed voltage regulator that applies full voltage to the alternator can easily cause excessive battery temperatures to be generated.

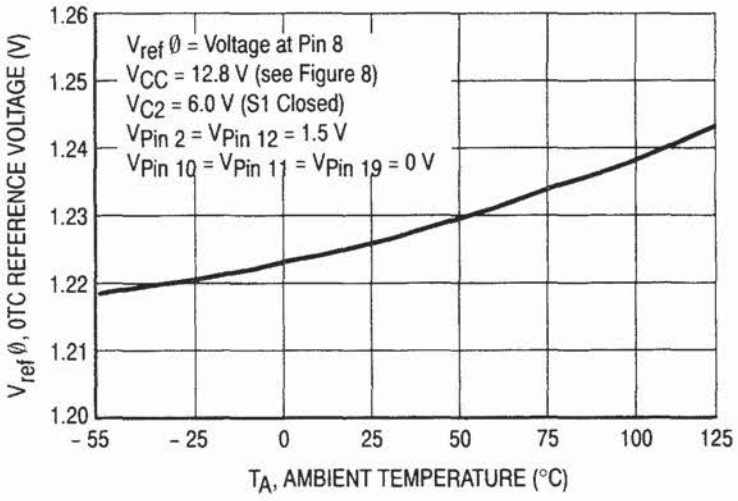


FIGURE 5.1 Temperature compensation in voltage regulator.

Batteries for electric vehicles (EVs) may require maintaining a specified high-temperature operating range. Sodium sulfur cells can store four times the energy of lead acid cells, but the battery's operating temperature must be kept within 300 to 350 °C. Solid oxide fuel cells operate at 1000 °C. If higher-temperature batteries are adopted for use in EV systems, sensing the battery temperature could be part of the diagnostic and fault detection.

5.2.3 Air Temperature Measurements

Air is measured for ambient temperature, passenger compartment temperature, and inlet air on production vehicles. For development purposes, a number of location ambients are always evaluated as potential mounting locations for electronic components. Upper temperature ranges are lower than liquid measurements, with 85 or 125 °C being common upper limits.

5.2.4 Catalyst Temperature

Measuring catalyst temperature during vehicle development has been performed since the earliest implementation of these devices. Concern for excessive temperatures in nearby vehicle locations meant that a number of measurements were made. However, the converter must be at a minimum temperature, usually above 350 °C, to be effective. Tougher exhaust emission standards are spurring manufacturers to utilize the converter sooner after the vehicle is started. Techniques are being pursued that would decrease the warm-up time for the catalyst to increase its effectiveness in controlling emissions. One technique involves heating the catalyst briefly by igniting a measured mixture of fuel and air in an afterburner ahead of the catalyst. Another technique utilizes electric heating to reduce the warm-up time. The power required for this approach would require an excessive power drain from the vehicle's power system. A linear temperature sensor has been developed to measure the temperature across the catalyst. A thermistor (see Sec. 5.4.1) is inserted diagonally inside the converter. A range of resistance has been evaluated and a time constant of about 2 s has been achieved. For fault condition monitoring, the sensor is able to detect a localized hot spot inside the catalyst and may prove to be able to detect misfire as part of future on-board diagnostic equipment required by the California Air Resources Board or the federal Environmental Protection Agency.

The temperature of exhaust gases increases rapidly under severe operating conditions such as continuous high speed or insufficient octane rating. One company has monitored their efforts at controlling the engine's operating temperature by locating a temperature sensor in the exhaust manifold. If the sensor detects an exhaust gas temperature rise above an established limit, a computer sends a command to inject extra fuel into the combustion chamber to cool the engine. Since exhaust gas temperatures can reach 1000 °C, a specially designed magnesium oxide thermocouple in a metal tube is used to sense the temperature (see Sec. 5.4.2). This approach may also be used in the future to reduce catalyst warm-up time.

5.2.5 Oxygen Sensor

The oxygen sensor used as the feedback element in closed-loop engine control systems generates a signal based on the difference in the oxygen concentration from inside the exhaust gas and ambient. As shown in the Nernst equation:

$$V_S = RT/4F \ln (P_a/P_g) \tag{5.4}$$

where V_S = the generated output voltage
P_a = oxygen partial pressure of the air
P_g = oxygen equilibrium pressure
R = gas constant
F = Faraday's constant
T = absolute temperature of the sensor

In addition to being affected by temperature, the oxygen sensor requires a minimum temperature to operate properly and provide a closed-loop control system. Typically, this minimum temperature is ≥450 °C. To reduce the oxygen sensor's warm-up time, heaters are added. A heated oxygen sensor achieves operating temperature much sooner than an unheated unit and allows closed-loop operation to be initiated. Monitoring temperature during engine and vehicle development is part of control strategy and component development.

5.2.6 Tire Temperature Sensing

Sensing the temperature within the tire is part of the pressure measurement and fault detection provided by tire pressure measuring and automatic adjustment systems. In a system like the one developed by Michelin (Fig. 5.2) that is used on Peugot's experimental sports car, the Proxima, both a pressure and temperature sensor are located on each wheel. A circular antenna and a transceiver allow these signals to be sent to an electronic processing module. An air compressor attempts to maintain desired tire pressure under normal conditions. If the pressure-temperature indications go above 85 °C, the system recommends speeds be held below 240 km/h (148 m/h). For temperatures above 90 °C, the limit is 160 km/h and for 95 °C, the limit is 80 km/h. For temperatures in excess of 100 °C, the system recommends stopping the vehicle.

5.2.7 Fault Detection for Electronic Components

As indicated in Sec. 5.1.6, temperature can cause reliability failures from several mechanisms in electronic systems. Therefore, detecting excessive temperature and designing system strategies to account for excessive temperatures can minimize improper operation and catastrophic failures.

Sensing for fault conditions, like a short circuit, is an integral part of power ICs or smart power transistors. The ability to obtain temperature sensors in mixed Bipolar-CMOS-DMOS semiconductor processes provides protection and diagnostics as part of the features of smart power ICs, like Motorola's SMARTMOS™ products. The primary function of the power IC is to provide a microcontroller-to-load interface for solenoids, lamps, and motors. It is a definite

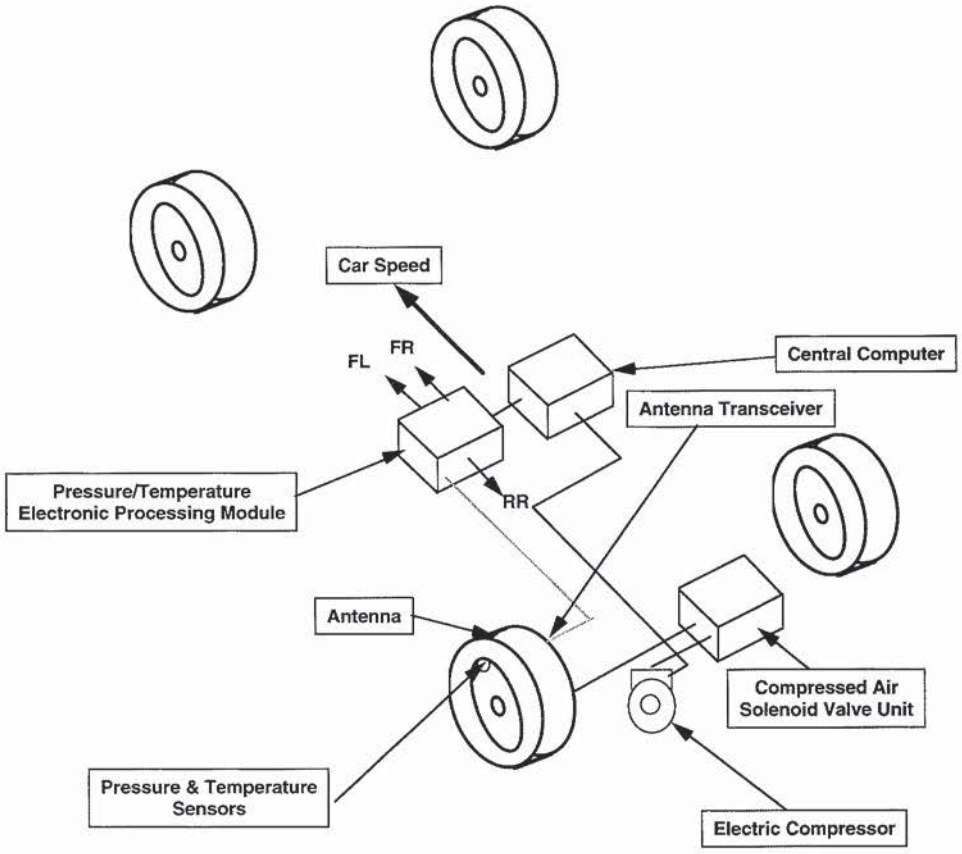


FIGURE 5.2 Temperature sensing in tire pressure control.

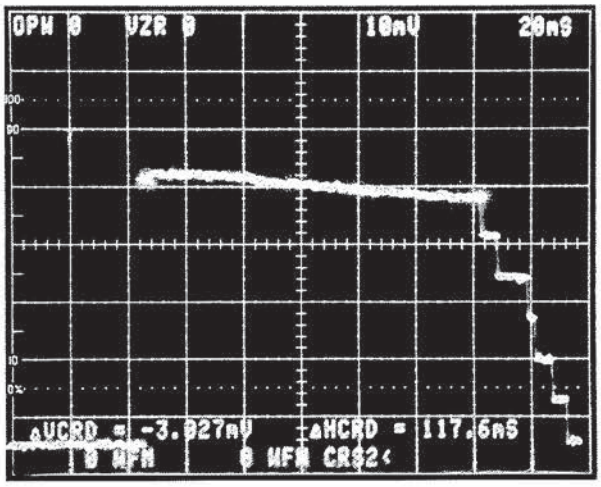


FIGURE 5.3 Thermal shutdown of six outputs of smart power IC.

advantage in multiple-output devices, to sense the junction temperature of each device, provide status input to the microcontroller and, if necessary, shut down a particular unit that has a fault condition. The sequential independent shutdown of six output drivers with a short applied across each unit's load is shown in Fig. 5.3. The action that is taken based on sensing a fault condition can vary depending on the control system. Sensing excessive temperature in a power device may mean that the device can turn itself off to prevent failure in one case and in another situation, a fault signal provides a warning but no action is taken. The remaining portion of the system is allowed to function normally. With the fault conditions supplied to the MCU, an orderly system shutdown can be implemented.

5.2.8 Mass Air Flow Sensor

The generalized model of a transducer is shown in Fig. 5.4. In addition to the desired input, undesired environment effects, like temperature and, to a lesser extent, humidity, are factors that affect the performance and accuracy of the transducer and must be taken into account during the design of the transducer.

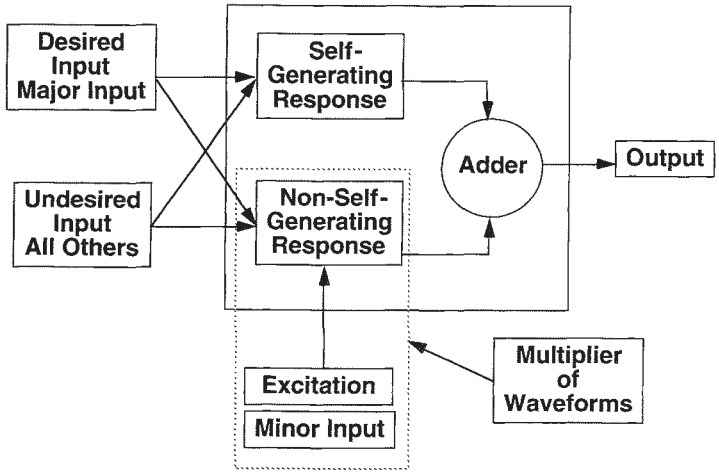


FIGURE 5.4 Generalized model of a transducer.

In the case of mass air flow sensing, temperature effects are used as part of the measurement process. A typical approach for mass air flow involves a hot-wire anemometer with a platinum wire as the sense elements, as shown in Fig. 5.5. The air temperature probe detects the temperature of the air flowing into the combustion process. The hot-wire probe is heated by current supplied from a power transistor. The difference in temperature between the probes is detected as a difference in electrical resistance. The electronic circuitry regulates the heating current to the hot-wire probe to keep the temperature difference constant at any flow rate (Fig. 5.6). The current required is proportional to the air flow rate. One of the main issues in this type of sensor is the response time. Improvements which reduce the response time from 50 to 30 ms have been made in recently developed units.

A microbridge silicon mass air flow sensor has been developed that utilizes temperature-resistive films laminated within a thick film of dielectric material and micromachined cavity as shown in Fig. 5.7. The resistors are suspended over the etched cavity, which also provides thermal isolation for the heater and sensing resistor. Heat is transferred from one resistor to the other as the result of flow. The imbalance in the resistance caused by the heat transfer

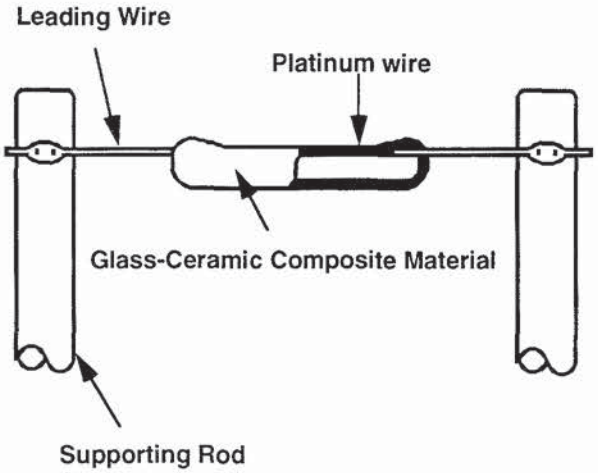


FIGURE 5.5 Probe for hot-wire anemometer mass air flow sensor.

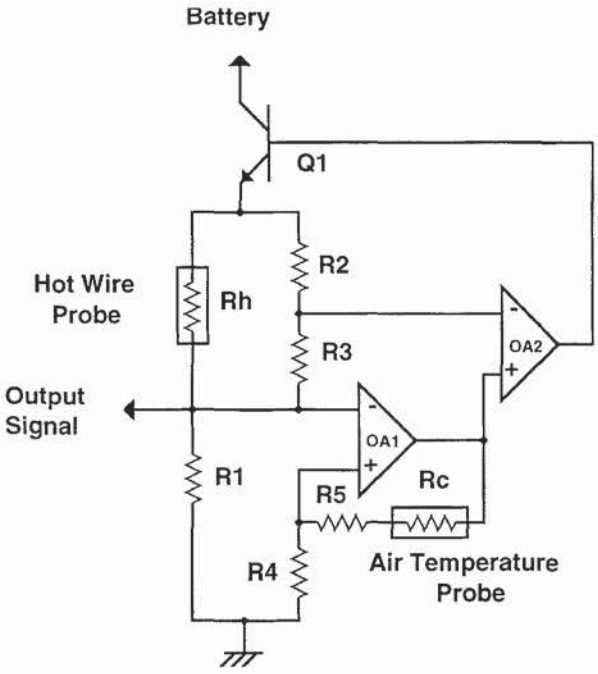


FIGURE 5.6 Circuit for hot-wire anemometer mass air flow sensor.

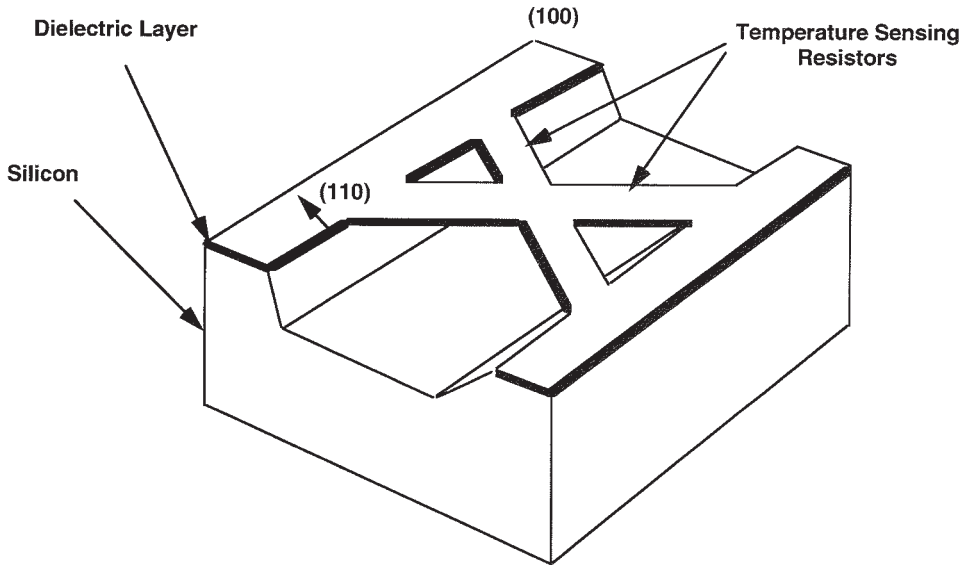


FIGURE 5.7 Temperature sensor in microbridge mass air flow sensor.

is directly proportional to the flow. The small size of the resistors allows fast response time (≤3 ms) and high sensitivity to flow. The device has also been found to be useful as a differential or absolute temperature sensor.

5.2.9 Temperature Measurements in New Systems

As new vehicle systems are developed or existing systems are modified to provide improved performance, or meet legislated or customer demands, new applications for temperature sensors will occur. An example of two systems, air bags and thermo accumulator, is provided.

In inflatable restraint systems, the air bag is inflated based on the ignition of a gas generant. In this system, an igniter receives an electric impulse, passes it through to a heating filament, and actuates a small pyrotechnic charge to ignite the gas generant, which produces nearly pure nitrogen gas to inflate the bag. The initial temperature is high but the nitrogen gas cools rapidly, so a fully deployed bag is only slightly above room temperature. Measuring the temperatures that occur during this event is part of the development activity.

A heat battery, or thermoaccumulator, can store the latent heat from a previously driven vehicle for up to three days at an ambient temperature of –29 °C. The accumulator is made of salt-based crystals which change from solid to liquid form above 78 °C. Heat provided by this device has proven to reduce the emissions which are generated during the first 20 s of the federal emissions test procedure. The unit also has potential application in passenger compartment heating.

5.3 HUMIDITY SENSING AND VEHICLE PERFORMANCE

Correlation of a number of vehicle tests requires that the humidity be monitored. Variations in humidity can explain variations that occur in test results.

Significant, repeatable, and controllable results may lead to humidity sensing being part of future engine control systems and passenger compartment comfort control systems.

Some sensors, such as rain (and/or windshield moisture) or sun sensors, are already being used on some vehicles to control wiper blade and fan controls.

5.3.1 Engine Performance

Injecting water to cool the intake charge and prevent premature detonation is a technique that has been used in high-compression (≥10:1) engines. The cooling effect from evaporation reduces the operating temperature.

An increase in inlet air humidity has been shown to reduce the nitric oxide emissions. Condensation in the gas tank can add a considerable amount of moisture to the fuel. Measuring the actual amount of humidity or moisture present during development tests is useful to identify all variables that can affect performance.

5.3.2 Passenger Compartment Comfort

A traditional passenger temperature control system shown in Fig. 5.8 uses only temperature as the indication for the system to open or close blend doors in the HVAC (heater, ventilating, and air conditioning) system. Laboratory evaluations monitor the HVAC systems' effectiveness under known and controlled humidity conditions. A fuzzy logic system combining temperature and humidity sensing has been investigated by researchers for home use. In this case, additional humidity is available to speed up the effect of adding heat. In a vehicle application, the blower speed could be adjusted based on the presence of humidity or an alternate strategy could be developed which uses less energy from the vehicle to make the driver and passengers comfortable.

The development of new test techniques frequently involves either the development of new sensors or novel use of existing sensors. A sweat impulse test developed at Volkswagen AG allows researchers to evaluate the ability of a seat to direct sweat away from its surface.

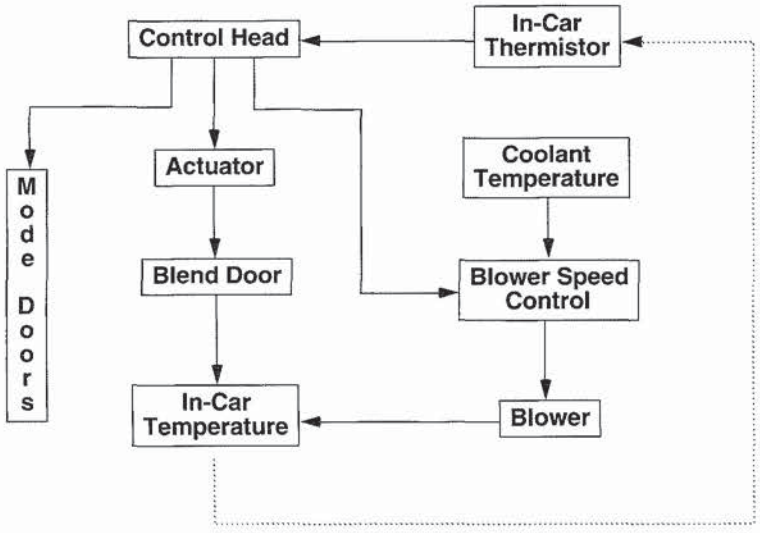


FIGURE 5.8 Automatic temperature control diagram.

As part of the measurement process, a platinum RTD (Sec. 5.4.5) and a capacitive humidity sensor provide data. In excessive sweating situations, the fabric is unable to absorb sufficient moisture, so it is passed to the surrounding environment as increased humidity.

5.3.3 Brake Moisture Sensing

Brake fluids are hygroscopic and, therefore, absorb moisture. Sufficient moisture lowers the boiling point of the brake fluid, and eventually vaporizes the fluid resulting in loss of stopping power. A meter has been developed that can be inserted into the master cylinder reservoir and measure the boiling point of the brake fluid. A heating element boils a sample of the brake fluid. A microcontroller calculates the actual boiling point by the initial temperature reading, temperature drop, and time between cooling and boiling. This approach may be applicable to other vehicle fluids that can be checked during maintenance procedures.

5.4 SENSORS FOR TEMPERATURE

Several different sensing techniques are used in production vehicles and during vehicle development to provide the temperature measurements. Table 5.4 shows a list of these techniques and the temperature range that is typical for this approach. Characteristics of temperature-sensing devices that allow a designer to determine if the approach will meet development or production requirements are explored in this section. In addition to the operating temperature range, linearity, response time, packaging, reliability, and cost are factors that must be considered. Temperature sensors can be as simple as thermal expansion devices, such as mercury or alcohol thermometers, or as complex as infrared sensors used in night vision systems. Common sensors and some recent additions to temperature sensing are covered in this section. Other sensor techniques, including the temperature-generating characteristics of quartz and temperature-sensing reed switches, are also possible but not covered.

5.4.1 Thermistors

Thermistors have found broad application acceptance in several vehicle systems that require temperature sensing from the coolant system to control the engine. The thermistor is a special

TABLE 5.4 Temperature Sensing Techniques versus Operating Range

Sensing technique	Temperature range, °C	Usage: Production (P) Development (D) Future (F)
Thermistor	0 to 500	P
Thermocouple	−200 to +3000	D
Bimetallic switch	−50 to +400 (650)	P
Potentiometer temperature sensor	−40 to +125	P
Platinum wire resistor (RTD)	−200 to +850 (−40 to +200)	D (P)
Semiconductor (junction)	−40 to +200	P
Thermostat (pressure spring)	−50 to +500	P
Fiber optic temperature sensor	+1800	D/F
Temperature indicator	+40 to +1350	D
Infrared	>Ta	F
Liquid thermometer	−200 to +1000	D

class of resistance temperature sensor (Sec. 5.4.5) based on semiconductor type materials that exhibit a wide range of temperature coefficients. The resistivity depends upon the material used to form the thermistor. Both negative (NTC) and positive temperature coefficient (PTC) devices can be produced. Commercially available thermistor forms are beads, probes, disks, and rods. The characteristics of thermistor output are shown in Fig. 5.9. The exact relationship between temperature and resistance depends on the material used for the sensor, but the general form of the output can be expressed by the equation

$$R = R_O \cdot e^{\beta}(1/T - 1/T_O) \tag{5.5}$$

where R = resistance at temperature T in ohms
R_O = resistance at temperature T_O in ohms
e = base of natural log
T, T_O = absolute temperature in K
β = indicator of shape of NTC curve

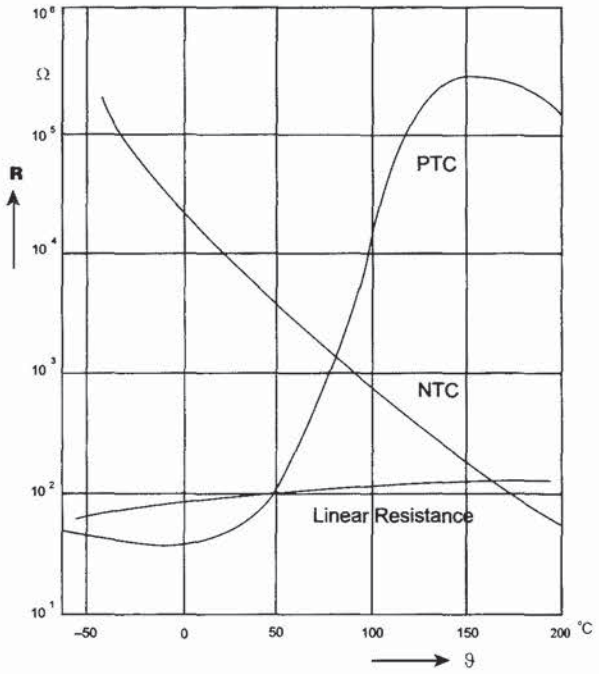


FIGURE 5.9 Thermistor output versus temperature.

This accounts for the high degree of nonlinearity in the output. However, the thermistor's nonlinearity can either be tolerated in many automotive applications, or corrected by an MCU or by circuit techniques. In the case of the water temperature indicator, for example, the thermistor's resistance is matched to a normal operating point on the gage; thus, the nonlinearity during conditions that generate higher temperature is not a problem.

Thermistors have a usable temperature range from –250 to +650 °C. Recent developments in automotive thermistors have focused on techniques to reduce cost and improve the manufacturability and reliability.

5.4.2 Thermocouples

Thermocouples are used frequently during the development phase of the vehicle because of their low cost, the broad range of temperature measurements that are possible from a single thermocouple such as iron-constantan, ease of extending the range through other materials, ease in obtaining a variety of form factors (especially small shapes), ease in forming the thermocouple, and ease in interfacing to temperature-recording equipment. The thermocouple type, wire size, and construction are factors that allow different units to be used from –250 to +1700 °C.

A thermocouple is formed by a pair of dissimilar conductors connected at one end. A voltage output E, proportional to the difference in temperature T_1 and T_2, is the basis of thermoelectric temperature measurements, also known as the Seebeck effect.

Thermocouple junctions can be formed by welding, soldering, or pressing materials together. Measurements can be made very easily; however, the highest accuracy is obtained by taking into account a number of factors, including reference junction considerations and a number (five) of laws that govern thermocouple behavior. Common types and their operating characteristics are shown in Table 5.5.

TABLE 5.5 Common Thermocouples and Application Factors

	Conductor		Temperature	Standard error	Seebeck
ISA code	Positive	Negative	Range, °C	Limit, °C	Coefficient, μV/°C
E*	Chromel	Constantan	0 to +316	±2	62
J*	Iron	Constantan	0 to +277	±2	51
T*	Copper	Constantan	–59 to +93	±1	40
K*	Chromel	Alumel	0 to +277	±2	40
N*	Nicrosil	Nisil	0 to +277	±2	38
S*	Platinum†	Platinum	0 to +538	±3	7
R*	Platinum‡	Platinum	0 to +538	±3	7

* Other temperature ranges and error limits are available.
† 10% Rhodium.
‡ 13% Rhodium.

5.4.3 Bimetallic Switch

A bimetallic temperature switch uses two metal strips with different coefficients of linear expansion that are welded together. Increasing temperature causes the strips to warp predictably. A switch point can be established by an initial calibration that indicates that a critical temperature has been reached and provides an input to an electronic control unit or a lamp indication. Switch types can be disk or cantilever construction. The use of switches does not require analog-to-digital conversion in control systems with microcontrollers, but provides limited information (only above or below a desired operating point) about the system. Early engine control systems used a disk switch in the air cleaner to indicate operation below 13 °C and a cantilever design in the engine coolant to indicate temperature above 71 °C.

5.4.4 Potentiometer Temperature Sensor

A bimetallic actuator can be combined with a high-resolution potentiometer to provide a temperature sensor for automotive applications. The linear movement of the bimetallic stack is transmitted by a stainless steel tube and measured as a distance by a conductive plastic film potentiometer providing a linear indication of temperatures up to 650 °C. The time constant for the sensor is approximately 70 s.

5.4.5 RTDs

A resistive temperature detector, or RTD, made of platinum is the highest inherent accuracy temperature sensor. The RTD has an output that varies with temperature based on the following equation:

$$\text{RTD} = R_O \cdot (1 + A \cdot T + B \cdot T^2) \tag{5.6}$$

where RTD = value of RTD at 0 °C in ohms and is 100 Ω for Pt100 and 200 Ω for Pt200
A = detector constant = $3.908 \cdot 10^{-3}$ ($°C^{-1}$) for Pt100
B = detector constant = $-5.802 \cdot 10^{-7}$ ($°C^{-2}$) for Pt100
T = temperature in °C

A constant current is applied to the RTD to obtain a voltage output over the range from –200 to +850 °C. Linearity can be within 3.6 percent for a range of 0 to 850 °C.

Other metals like nickel, nickel-iron alloy, copper, and aluminum can also be used for the RTDs. In addition to wire-wound RTDs, thin-film techniques also allow an RTD to be applied to a substrate, such as ceramic or silicon.

5.4.6 Semiconductor Techniques

Several semiconductor parameters vary linearly over the operating temperature range. As shown in Fig. 5.10, the gate threshold voltage of a power MOSFET changes from 1.17 to 0.69 times its 25 °C value when the temperature increases from –40 to 150 °C. Also, the breakdown voltage of the power MOSFET varies from 0.9 to 1.18 times its value at 25 °C over the same –40 to 150 °C temperature range (Fig. 5.10). These relationships are frequently used to determine the critical temperature (junction temperature) of semiconductor components in actual circuit operation. External package-level temperature measurements can be several degrees different than the junction temperature, especially during rapid, high-energy switching events. The actual junction temperature and the resulting effect on semiconductor parameters must be taken into account for the proper application of semiconductor devices. Semiconductors can be used for sensing temperature within their operating temperature range.

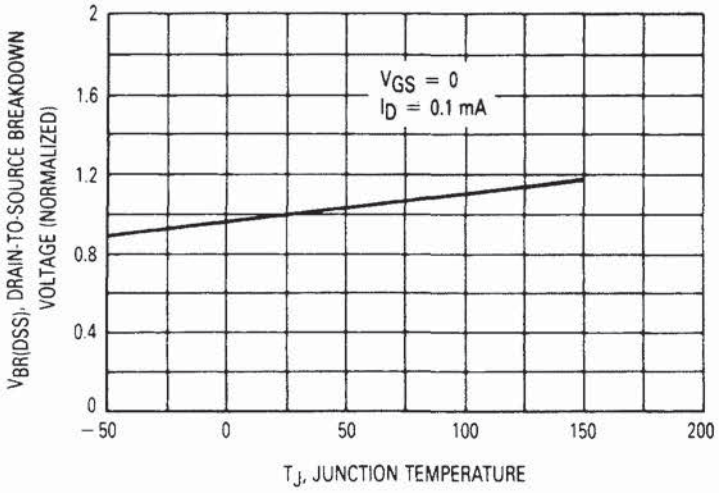


FIGURE 5.10 Breakdown voltage versus temperature.

The most commonly referenced transistor temperature measurement is the V_{BE} relationship to temperature. A silicon temperature sensor that utilizes this effect has a nominal output of 730 mV at –40 °C and an output of 300 mV at 150 °C. The linearity of this device is shown in Fig. 5.11. The total accuracy is within ±3.0 mV including nonlinearity which is typically within ±1 °C in the range of –40 to 150 °C. The readings are made with a constant (collector) current of 0.1 mA passing through the device to minimize the effect of self-heating of the junction.

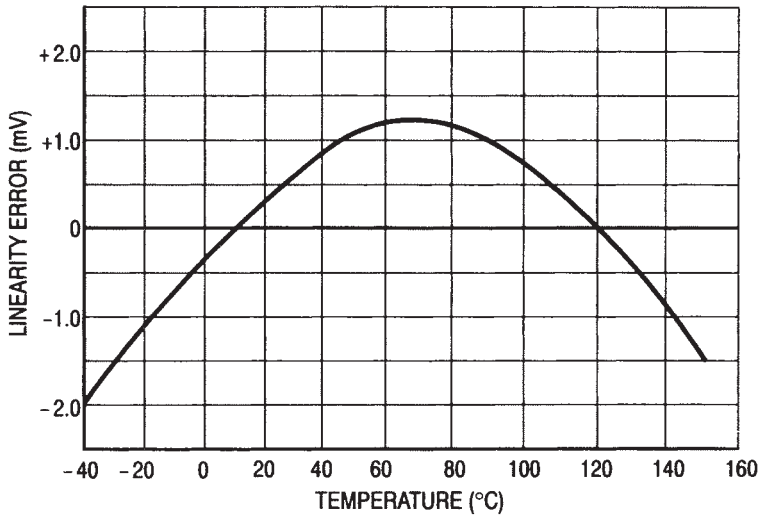


FIGURE 5.11 Linearity of silicon temperature sensor.

The linear relationship of the V_{BE} of a transistor to temperature provides a considerable improvement in accuracy over a thermistor. However, the limited operating range of semiconductor temperature sensors (–40 to 150 °C) has limited their application in vehicles. To be cost-effective in automotive applications, semiconductor sensing is usually incorporated with other sensing or interface devices.

An RTD (Sec. 5.4.5) with an aluminum metal film has been integrated with a semiconductor process to provide a temperature sensor that can directly drive the analog-to-digital input in a microcontroller unit (MCU). The sensor has better than ±2 percent accuracy over a –50 to 150 °C temperature range. On-chip linearization techniques avoid look-up table linearization by the MCU.

Another integrated semiconductor temperature sensor uses the fact that two identical transistors operating at a constant ratio of collector current densities r results in the difference in base-emitter voltages being $(kT) \cdot (\ln r)$. Both Boltzman's constant k and the charge of an electron q are constant, so the resulting voltage is directly proportional to absolute temperature. The voltage is then converted to a current by low-temperature coefficient thin-film resistors.

Polysilicon diodes and resistors can be produced as part of a power MOSFET semiconductor manufacturing process and used as temperature-sensing elements to determine the junction temperature of the device. The thermal sensing that is performed by the polysilicon elements is a significant improvement over power device temperature sensing that is performed by an external temperature-sensing element. An extreme thermal gradient exists between the maximum junction temperature at the top (active area) of the power device and the remote point which is able to be measured by a thermistor or thermocouple. Placing the thermal-sensing element inside the package improves the measurement process considerably.

By sensing with poly resistors or diodes, the sensor can be located close to the center of the power device near the source bond pads where the current density is the highest and, consequently, the highest die temperature occurs. The thermal conductivity of the oxide that separates the polysilicon diodes from the power device is two orders of magnitude less than that of silicon. However, because the layer is thin, the polysilicon element offers an accurate indication of the actual peak junction temperature.

A very low on-resistance power FET that incorporates a temperature sensing diode is shown in Fig. 5.12. The power device achieves an on-resistance of 5 milliohms (including packaging resistance) at 25 °C utilizing a production power FET process. The power FET process was chosen instead of a power IC process because a large die size was required to meet the power dissipation requirements of a very cost sensitive application. A minor modification (the addition of a single masking layer) to a production power FET process allows both sensing and protection features to be integrated. Devices that are produced using the SMARTDISCRETE technology, as this approach is called by Motorola, are one way to provide more cost-effective, smart power semiconductors, especially for high-current or low on-resistance power applications.

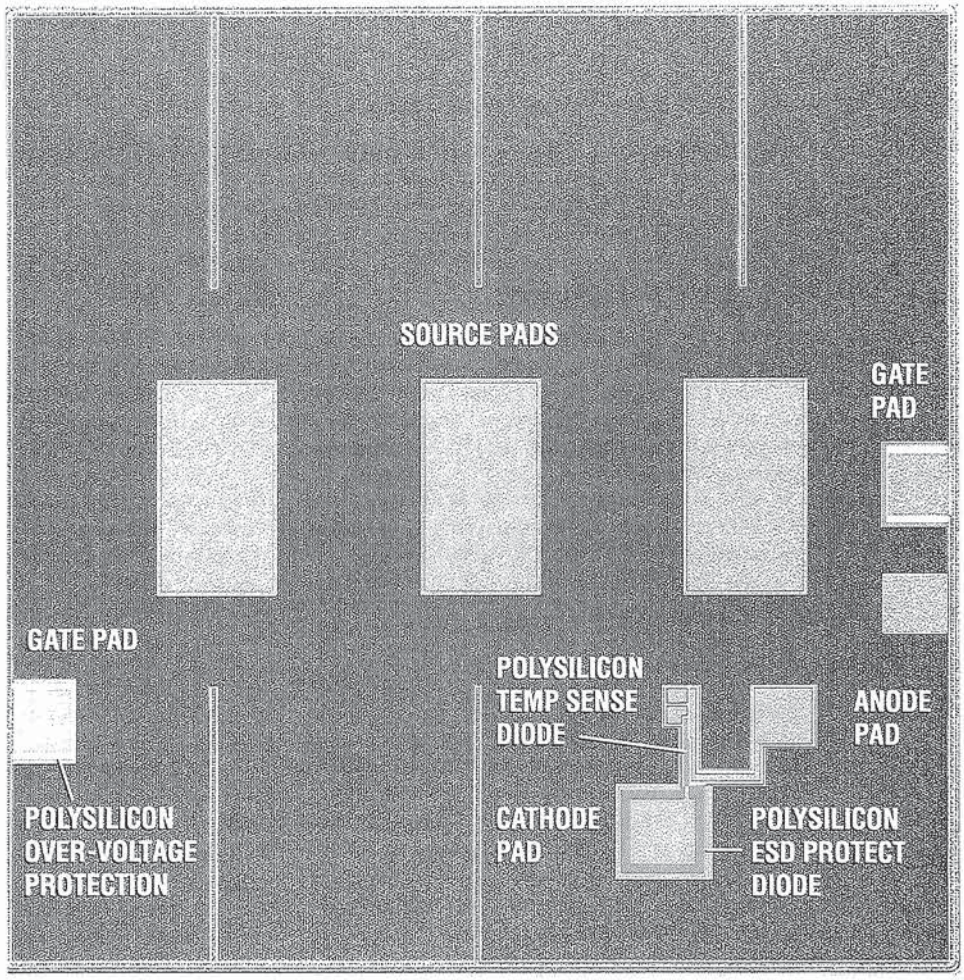


FIGURE 5.12 Temperature sensor integrated in power MOSFET. *(Die photomicrograph Courtesy Motorola Inc.)*

An accurate indication of the maximum die temperature is obtained by monitoring the output voltage when a constant current is passed through the integrated polysilicon diode(s). A number of diodes are actually provided in the design. A single diode has a temperature coefficient of 1.90 mV/°C. Two or more diodes can be placed in series if a larger output is desired. For greater accuracy, the diodes can be trimmed during wafer-level testing by blowing fusible links made from polysilicon. The response time of the diodes is less than 100 μs which has allowed the device to withstand a direct short across an automobile battery with external circuitry providing shutdown prior to device failure. The sensing capability also allows the output device to provide an indication (with additional external circuitry) if the heatsinking is not proper when the unit is installed in a module or if a change occurs in the application which would ultimately cause a failure.

A smart power IC, like Motorola's SMARTMOS™ process, can have multiple power drivers integrated on a single monolithic piece of silicon. Each of these drivers can have a temperature sensor integrated to determine the proper operating status and shut off only a specific driver if a fault occurs. The temperature sensor is basically a band gap reference. In this case, two diodes' junctions are biased with different current densities. Circuitry in the smart power IC process establishes a voltage reference and a trip point.

5.4.7 Thermostat (Mechanical Temperature Sensing)

The earliest method of temperature regulation used in vehicles is still in use today—the thermostat. The expansion element usually activates a valve which redirects the flow of coolant into a radiator bypass line when a specific temperature is reached.

5.4.8 Fiber Optic Temperature Sensors

High-temperature sensors capable of operating up to 1800 °C have been developed using optical fibers. The response time of the sensor is on the order of 500 ms and the accuracy is within ±0.5 percent. The noncontact sensor uses a single optical fiber that has an outside diameter of 0.5 to 1.0 mm. The probe structure is critical to the sensor's sensitivity and the distance coefficient that relates the detection surface area to the distance at which the probe will be placed for measurement.

A sapphire fiber optic sensor has been developed to measure high temperatures such as those found near the tip of a spark plug during combustion. The sensor can withstand temperatures up to 4000 °C and has an accuracy of 0.2 percent at 1000 °C. The sensor has been used to identify knock and to characterize flame front propagation.

5.4.9 Temperature Indicators

Temperature-indicating materials have been developed with melting points that can be calibrated to ±1 °C. When the temperature of the device is exceeded, a phase change occurs and the material changes color. The material is available in paint, patch, pellet, and applicator stick form. Temperature ranges from 38 to 1371 °C are available. This kind of indication is useful during the development phase of automotive components to verify that a critical design temperature has not been exceeded.

5.4.10 Infrared Temperature Sensing

Blackbody emittance is the basis for infrared (IR) thermometry. Objects at temperatures above –273 °C emit radiant energy in an amount proportional to the fourth power of their

temperature. An IR sensor consists of collecting optics, lenses, and/or fiber optics, spectral filtering, and a detector as the front end. The primary characteristic of the optics is the field of view (FOV) which allows a specific target to be measured at a prescribed distance.

Infrared sensors have been used extensively for noncontact temperature measurements in process control. In automobiles, infrared sensors use temperature variations to produce a monochrome image in night vision systems. Board-level temperature testing also utilizes IR-sensing techniques to identify hot spots and potential failure points.

5.4.11 Thermal Actuators/Thermal Cooling

Chemical sensors like the oxygen sensor described in Sec. 5.2.6 require a minimum operating temperature to be effective. Semiconductor technology allows platinum and silicon heaters and temperature sensors to be produced on a thin diaphragm, or window, etched into a silicon substrate. Temperatures can be produced in the window which exceed 1200 °C, while the supporting rim of silicon remains essentially at room temperature. This ability may allow fast-response sensors to be developed for vehicle applications.

Other silicon development activity using micromachining techniques may allow heat pipes and silicon-level heatsinks to remove heat more effectively from rapidly switching semiconductor devices.

5.5 HUMIDITY SENSORS

Humidity sensors are used extensively during vehicle and component development, especially in determining the ability to withstand humidity and temperature cycling tests. Engine performance can be affected by humidity. A sensor that can pass rigorous automotive qualification tests and meet the desired specifications and cost criteria would provide additional input for the engine control system. In a less rugged application, HVAC systems may benefit from the input from humidity sensors. Humidity sensors are not used on production vehicles at the present time. However, some potential techniques that are used in development and nonautomotive applications are described in this section.

Techniques to measure humidity are listed in Table 5.6. Most of these are actually laboratory instruments. The most well known approach to humidity measurements is the use of wet bulb and dry bulb thermomometer readings interpreted through the use of a psychrometric chart which relates all basic humidity readings. The effect of humidity on polymers and the use of semiconductor-processing techniques has created actual sensors that do not require the constant calibration, have less drift, improved reliability, and potentially can achieve automotive cost objectives. Three sensing techniques that have potential for future vehicle use are capacitive, resistive, and oxidized porous silicon.

5.5.1 Capacitive Humidity Sensors

A thin-film polymer capacitive-type relative humidity sensor can be produced by using semiconductor technology. The dielectric constant of the polymer thin film changes linearly with changes in atmospheric relative humidity. A desirable feature, in addition to long-term stability and operation to 180 °C, is fast wet-up/dry-down performance.

A temperature sensor can easily be added to the RH sensor, which increases the operating range and the functionality.

TABLE 5.6 Techniques for Measuring Humidity/Moisture

Principle	Type measurement
Gravimetric hygrometer	Instrument
Pressure-humidity generator	Instrument
Wet bulb/dry bulb (psychrometer)	Instrument
Hair element	Instrument
Electric conductivity	Instrument
Dew cell	Instrument
Chilled mirror	Instrument
Karl Fisher titration	Instrument
Electrolytic	Instrument
Lithium chloride	Instrument
Capacitance hygropolymer	Production sensor
Bulk polymer (resistance)	Production sensor
Thin-film polymer (capacitance)	Production sensor
Gold/aluminum oxide	Production sensor
Oxidized porous silicon (OPS)	Experimental sensor

5.5.2 Resistive Humidity Sensors

Both surface and bulk polymer resistance sensors have been developed for measuring relative humidity. In bulk polymer film units, mobile ions are released from the molecular structure with increased humidity levels resulting in orders-of-magitude change in resistance over the operating range. Diode temperature compensation is used to improve accuracy over the temperature range. With an ac excitation at 1 kHz, a change from 10 to 100 percent RH produces a corresponding resistance variation from $2 \cdot 10^7$ to $2 \cdot 10^3$ ohms.

5.5.3 Oxidized Porous Silicon

A capacitive-type sensor has been developed that uses oxidized porous silicon (OPS) as a moisture-absorbing dielectric between the electrodes of the capacitor. An electrolysis process is used to create a thin porous layer on the top of a silicon wafer. The silicon is converted to OPS by a high-temperature treatment in either oxygen or steam. Metal electrodes are deposited on the top and back of the OPS to complete the capacitive structure.

When water vapor contacts the sensor, it permeates through the porous structure between the electrodes. The response of a typical sensor increases 800 percent when exposed to an RH change from 1 to 40 percent. Since semiconductor techniques are used in the manufacturing process, the potential to integrate signal conditioning and other sensors may provide an attractive solution for future vehicle sensing requirements.

5.6 CONCLUSIONS

Several factors are influencing the temperature-sensing requirements for future vehicles. The ultra-efficient vehicle will operate at higher temperatures and utilize brake energy-recovery. Both of these aspects will increase the temperature environment for other vehicle components.

The requirements for sensors in future vehicle systems include improved performance (sensitivity) and reliability in spite of operating at higher temperatures, lower cost, space and weight reduction, greater functionality including diagnostics and self-test, and ease of inter-

facing with an MCU for adaptive control and interchangeability. The capability of semiconductor technology to provide sensing information for failure detection provides a means for increased reliability in several vehicle subsystems.

GLOSSARY

Accuracy A comparison of the actual output signal of a device to the true value of the input. The various errors (such as linearity, hysteresis, repeatability, and temperature shift) attributing to the accuracy of a device are usually expressed as a percent of full-scale output (FSO).

Maximum operating temperature The maximum body temperature at which the thermistor will operate for an extended period of time with acceptable stability of its characteristics. This temperature is the result of the internal or external heating, or both, and should not exceed the maximum value specified.

Negative temperature coefficient (NTC) An NTC thermistor is one in which the zero-power resistance decreases with an increase in temperature.

Operating temperature range The range of temperature between minimum and maximum temperature at which the output will meet the specified operating characteristics.

Self-generating Providing an output signal without applied excitation, such as a thermoelectric transducer.

Self-heating Internal heating resulting from electric energy dissipated within the unit.

Storage temperature range The range of temperature between minimum and maximum which can be applied without causing the sensor (unit) to fail to meet the specified operating characteristics.

Temperature coefficient of full-scale span The percent change in the full-scale span per unit change in temperature relative to the full-scale span at a specified temperature.

Temperature hysteresis The difference in output at any temperature in the operating temperature range when the temperature is approached from the minimum operating temperature and when approached from the maximum operating temperature with zero input applied.

Thermal time constant The thermal time constant is the time required for a thermistor to change to 63.2 percent of the total difference between its initial and final body temperature when subjected to a step function change in temperature under zero-power conditions.

Transfer function A mathematical, graphical, or tabular statement of the influence which a system or element has on the output compared at the input and output terminals.

BIBLIOGRAPHY

Automotive Electronics Reliability Handbook, SAE, Warrendale, Pa., 1987.
Automotive Handbook, Robert Bosch GmbH, 1986.

BR923/D Discrete and Special Technologies Group Reliability Audit Report 1Q90, Motorola Semiconductor Products Sector, Phoenix, Ariz.

Collings, N., W. Cai, T. Ma, and D. Ball, "A linear catalyst temperature sensor for exhaust gas ignition (EGI) and on board diagnostics of misfire and catalyst efficiency," SAE SP-957, *U.S. and European Automotive Technology Emissions Technology,* International Congress and Exposition, Detroit, March 1–5, 1993.

De Vera, D., T. Glennon, A. A. Kenny, M. A. H. Khan, and M. Mayer, "An automotive case study," *Quality Progress,* June 1988, pp. 35–38.

Dosher, J., "Smart and active sensors," *Proceedings of Sensors Expo West,* 1991, pp. 103C-1–6.

Hao, T., E. F. Malinowski, and C-C. Liu, "High temperature fiber optic sensor for industrial applications," *Proceedings of Sensors Expo West,* San Jose, Calif., March 2–4, 1993, pp. 227–228.

Norton, Harry N., "Section 10—Transducers and Sensors," in D. G. Fink and D. Christiansen (eds.), *Electronics Engineers' Handbook* (3d ed.), McGraw-Hill, New York, 1989.

Pressure Sensor Device Data, Motorola Semiconductor Products Sector, Phoenix, Ariz., 1993.

Recommended Environmental Practices for Electronic Equipment Design, SAE J1211, Warrendale, Pa., Nov. 1978.

Temming, J., "Measuring seat comfort," *Automotive Engineering,* July 1993, pp. 25–30.

CHAPTER 6
EXHAUST GAS SENSORS

H.-M. Wiedenmann, G. Hötzel,
H. Neumann, J. Riegel, and H. Weyl
Robert Bosch GmbH
Lambda Oxygen Sensor Development

6.1 BASIC CONCEPTS

6.1.1 Combustion

The sole products of complete fuel combustion are the nontoxic substances carbon dioxide and water:

$$C_mH_n + \left(m + \frac{n}{4}\right) 0_2 \rightarrow m\ CO_2 + \frac{n}{2}\ H_2O \tag{6.1}$$

The theoretical air requirement for this process is 14.7 kg air for each kilogram of fuel; this corresponds to approximately 10 m^3 air per liter of fuel. The air/fuel ratio is defined as stoichiometric when the engine is supplied with the exact quantity of air required for complete combustion.

6.1.2 Definition of Normalized air/fuel Ratio Lambda

The mixture ratio is defined by the normalized air/fuel ratio lambda (λ):

$$\lambda = \frac{\text{current air-fuel ratio}}{\text{stoichiometric air-fuel ratio}} \tag{6.2}$$

Because the conditions within the engine do not correspond to the absolute ideal required for perfect combustion, a number of products of incomplete combustion occur, even if a stoichiometric air/fuel ratio ($\lambda = 1$) is maintained. Thus, the CO_2 and H_2O are joined by CO, H_2, and HC (hydrocarbons C_xH_y) along with corresponding amounts of not reacted (free) oxygen. The water-gas equilibrium defines the ratio of CO to H_2. At high combustion temperatures, the N_2 and O_2 in the air supply form nitrous oxides such as NO, NO_2, N_2O (generic designation: NO_x).

6.1.3 Composition of Untreated Exhaust Gas

The composition of the exhaust gases entering the catalytic converter (untreated or raw emissions) varies according to the quality of combustion and, even more importantly, as a function

of the normalized air/fuel ratio lambda (λ) (Fig. 6.1). Rich mixtures (lambda < 1, excess fuel) produce high concentrations of CO, H_2, and HC, while lean mixtures (lambda > 1, excess oxygen) generate higher levels of NO_x and free oxygen. The lower combustion-chamber temperatures associated with mixture ratios of lambda > 1.2 result in reductions in NO_x concentrations accompanied by increased HC concentrations. Maximum emissions of CO_2 (greenhouse gas) occur at a slightly lean mixture (lambda ≈ 1.1).

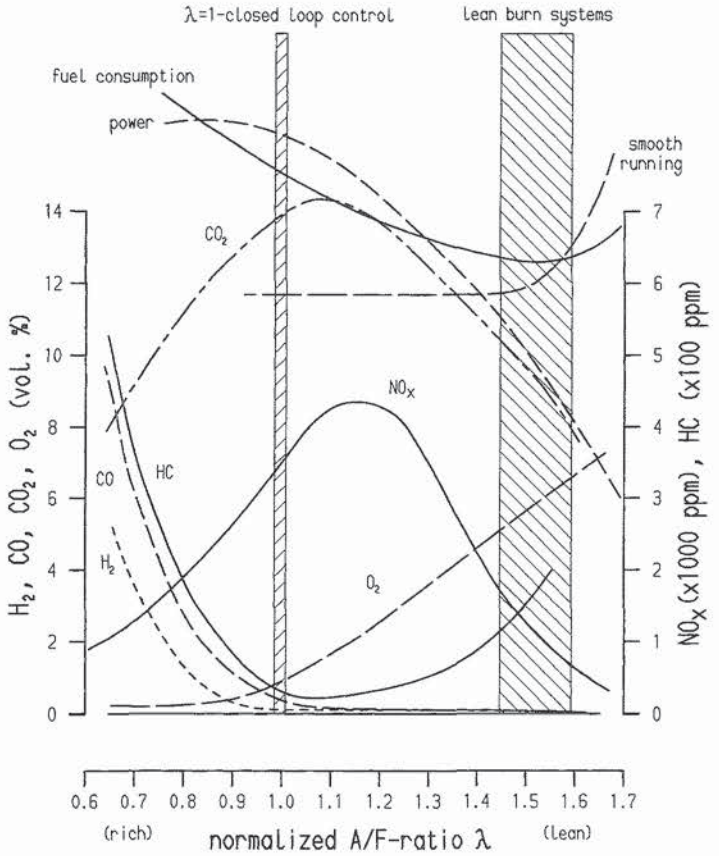


FIGURE 6.1 Typical curves for fuel consumption, power, smoothness, and composition of raw exhaust gases relative to lambda in a spark-ignition engine.

6.1.4 Lambda Closed-Loop Control: Design Concepts

Lambda Control Ranges. The main elements considered in defining the range of the lambda closed-loop control system (Table 6.1) are engine design, emission limits, fuel consumption, and the requirements for performance and smooth running.

Treatment of Exhaust Gases. Catalytic treatment of exhaust gases is essential for achieving compliance with the stringent current United States emission standards (Table 6.2). In the catalytic process, CO, H_2, and HC are oxidized to form CO_2 and H_2O, while the constituents of

TABLE 6.1 Lambda Control Ranges

	Lambda = 1—spark ignition engine	Lean-burn concept	Lean-mix concept	Diesel engine
Lambda-control range	Lambda = 1 $\pm < 0.005$	1.4–1.7	1.4–1.7 and lambda = 1	1.1–7
Target	Opt. catalytic aftertreatment	Fuel economy low CO, HC, (NO_x)*	Improved fuel economy within allowable emission	Reduced emissions particles and NO_x
Lambda sensors	Lambda = 1 sensor	Lean A-F sensor	Wide-range A-F sensor	Lean A-F sensor
Exhaust gas aftertreatment	Three-way-catalyst (TWC)	Oxi-cat NO_x-cat*	TWC NO_x-cat*	Particle filter* NO_x-cat*
Control concepts	Conventional or continuous	Continuous	Lean: continuous lambda = 1: conventional	Full-load and EGR-control
Series production	Discont. control	—	Yes	—

* Not available yet.
TWC: three-way-catalyst; Oxi-cat: oxidation-catalyst; NO_x-cat: NO_x-reduction-catalyst.

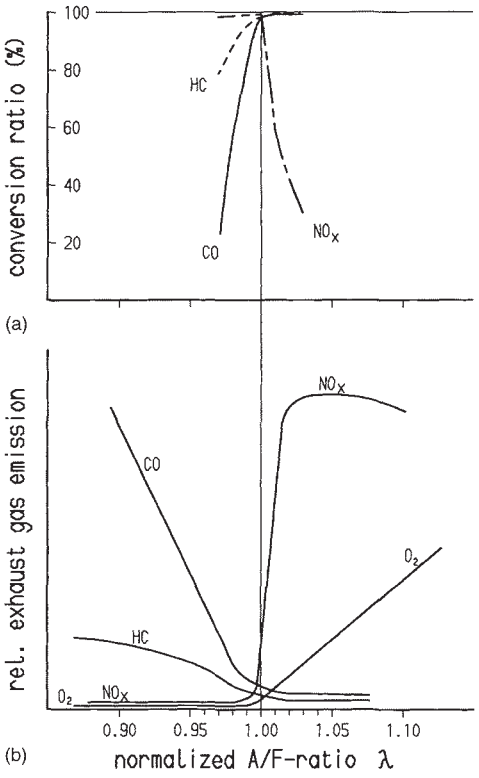


FIGURE 6.2 (*a*) Conversion rates for CO, NO_x, and HC in a new three-way catalytic converter; (*b*) typical curve for exhaust-gas composition downstream from a new three-way catalytic converter.

NO_x are reduced to N_2 and O_2. A three-way catalytic converter TWC (selective catalytic converter) and closed-loop control system featuring a lambda sensor are essential elements in achieving adequate reductions in all three pollutants (Fig. 6.2). The engine must be operated within a narrow range of $|\Delta\lambda| < 0.005$ at lambda = 1.

Oxidation catalysts are sometimes employed to reduce CO and HC emissions from lean-burn and diesel engines. Meanwhile, NO_x reduction catalysts and particle filters are being developed for future applications (Table 6.1).

Lambda Closed-Loop Control. Lambda closed-loop control is incorporated in the engine's electronic control system. The control system regulates lambda upstream from the catalytic converter with the aid of an oxygen sensor. System lag results in relatively long control-response delays, especially at low rpm. For this reason, the system must incorporate a pilot-control function capable of adjusting the mixture to the desired lambda with the highest possible degree of precision. This arrangement makes it possible to avoid excessive control deviations and harmful emission peaks of the kind that lead to high emission levels and reduced vehicle performance.

Conventional Closed-Loop Lambda Control. The closed-loop control concept pre-

TABLE 6.2 Summary of Emission Levels for Spark-Ignition Engines
(Status: July 1992)

		Exhaust emission limits						
	Start of sale	HC	CO	NO_x	HC + NO_x	Unit	Remarks	
California	MY 1993	0.25	3.4	0.4	—	g/mile		
	MY 1994	0.125	3.4	0.2	—	g/mile	TLEV 10 %	% of fleet,
	MY 1997	0.075	3.4	0.2	—	g/mile	LEV 25 %	stepwise
	MY 1997	0.04	1.7	0.2	—	g/mile	ULEV 2 %	increase
	MY 1998	0.0	0.0	0.0	—	g/mile	ZEV 2 %	
USA (Fed.)	MY 1983	0.41	3.4	1.0	—	g/mile		
	MY 1994	0.25	3.4	0.4	—	g/mile		
	MY 2003	0.125	1.7	0.2	—	g/mile	on need	
Japan	4/1981	0.25	2.1	0.25	—	g/km	10-mode	
EG	1992	—	2.72	—	0.97	g/km	MVEG I	
	10/1996	0.15	2.1	0.3	—	g/km	MVEG II	proposal parliament
		—	2.2	—	0.5	g/km		proposal commission
	1996	—	1.5	—	0.2	g/km	MVEG III	proposal Germany

TLEV: Transitional Low-Emission Vehicles
ULEV: Ultra Low-Emission Vehicles
LEV: Low-Emission Vehicles
ZEV: Zero-Emission Vehicles
MVEG: Motor Vehicle Emission Group

sently in use in spark-ignition engines is based on 2-point lambda = 1 control (Fig. 6.3) with the mixture composition oscillating around the optimum lambda. When the exhaust mixture shifts from rich to lean, the voltage from a lambda = 1 probe drops from approximately 800 mV (lambda < 1) to approximately 100 mV (lambda > 1), with a steep signal change (lambda jump) occurring at lambda = 1 (Sec. 6.2.1). Once the signal from the sensor crosses beyond the specified threshold value (e.g., 450 mV), the system responds by progressively leaning out the mixture until the probe signal drops below this border again. Once this process is completed, the system reverts to graduated mixture enrichment. Depending upon the system lag, the subsequent control-oscillation rate will lie in the 0.5 . . . 5 Hz range, with an amplitude of $\Delta\lambda = \pm 0.01 \ldots 0.05$ relative to the median lambda. During the lean-operation period, the catalytic converter stores oxygen for release during the rich-mixture phases. This arrangement ensures high conversion rates despite the control oscillations. Preferably an algorithm with proportional plus integral characteristics (PI controller) is used. Variations in system lag and integrator pitch affect the amplitude and the frequency of the oscillations. The proportional factors have to be adapted. Because load and rpm have a major effect on system lag, control parameters are defined and stored in a load/rpm control map.

Continuous Lambda = 1 Control. A continuous lambda = 1 control can be employed to achieve substantially smaller control deviations, providing attendant reductions in emissions, especially when used with aged catalytic converters. An oxygen sensor with a roughly linear or linearized lambda characteristic curve is required for this arrangement (Secs. 6.2.5 and 6.2.6).

6.1.5 On-Board Diagnosis (OBD II)

From the 1994 model year onward, on-board devices capable of monitoring the operation of all emission-relevant vehicle components will be required in the United States. Pollutant emission sensors (see Sec. 6.6) to measure levels of CO, NO_x, and HC emissions downstream from the catalytic converter represent an ideal means of monitoring the performance of both converter and oxygen sensor. Unfortunately, there are still no sensors of this type suitable for

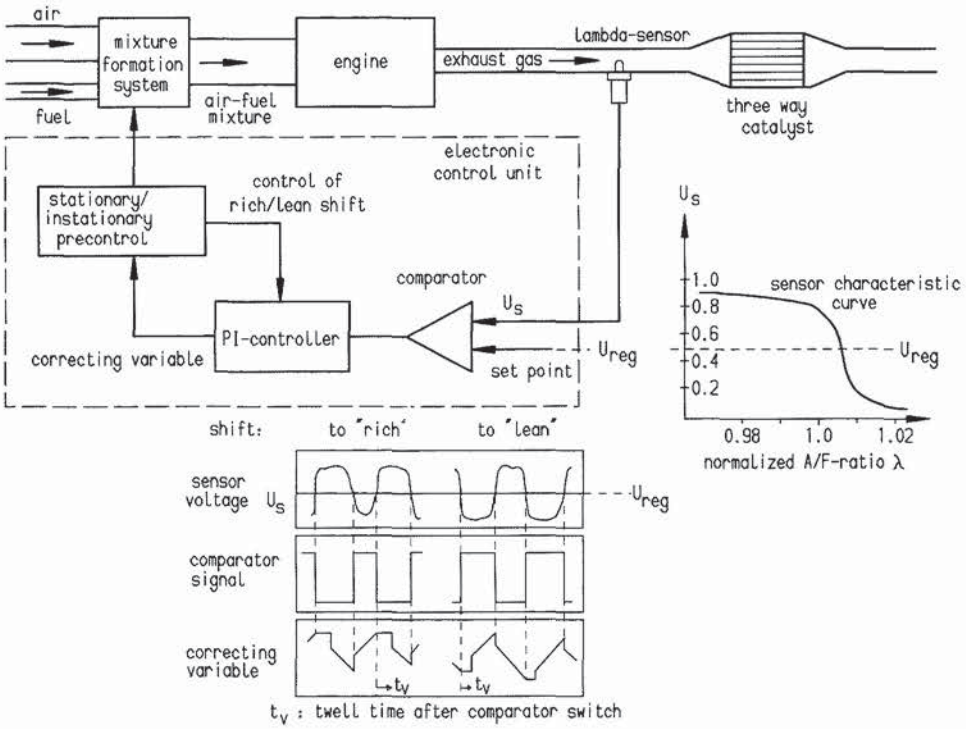


FIGURE 6.3 Operating diagram for lambda control.

use in the engine's exhaust stream. As an alternative, there exist concepts that employ a second oxygen sensor behind the catalytic converter to detect aging in the converter and/or the lambda sensor.

6.2 PRINCIPLES OF EXHAUST GAS SENSORS FOR LAMBDA CONTROL

6.2.1 Lambda = 1 Sensor: Nernst Type (ZrO_2)

Sensing Mechanism. In principle, the lambda sensor operates as a solid-electrolyte galvanic oxygen-concentration cell (Fig. 6.4). A ceramic element consisting of zirconium dioxide and yttrium oxide is employed as a gas-impermeable solid electrolyte. This mixed oxide is an almost perfect conductor of oxygen ions over a wide temperature range (Sec. 6.3). The solid electrolyte is designed to separate the exhaust gas from the reference atmosphere. Both sides feature catalytically active platinum electrodes. At the inner electrode (air; $pO_2'' \sim 0.21$ bar), the electron reaction

$$O_2 + 4e^- \rightarrow 2\,O^{2-} \tag{6.3}$$

incorporates oxygen ions in the electrolyte. These migrate to the outer electrode (exhaust gas; pO_2' variable $< pO_2''$), where the counterreaction occurs at the three-phase boundary (elec-

trolyte-platinum-gas). A counteractive electrical field is created, and an electrical voltage U_s corresponding to the partial-pressure ratio is generated in accordance with the Nernst equation:

$$U_s = \frac{RT}{4F} \ln \frac{pO_2''}{pO_2'} \tag{6.4}$$

where R = general gas constant
F = Faraday's constant
T = absolute temperature
pO_2 = partial pressure of oxygen

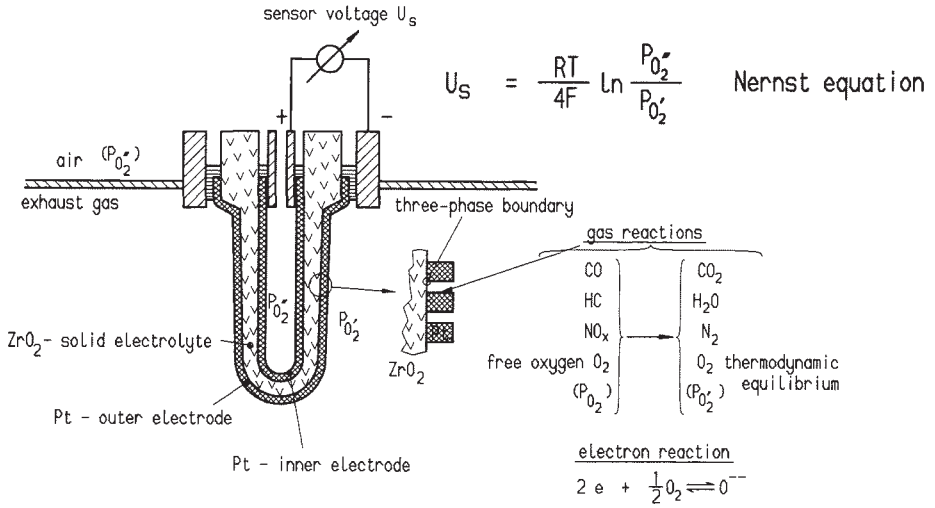


FIGURE 6.4 Diagram illustrating design and operation of a lambda = 1 sensor.

Catalytic Process. Measurements of oxygen content can only serve as the basis for unambiguous conclusions regarding the lambda of the exhaust gas when a state of thermodynamic gas equilibrium is established at the catalytically active electrodes on the oxygen sensor (residual oxygen). The absolute concentrations of the individual components in the engine's exhaust gases fluctuate across a wide range in accordance with the instantaneous operating conditions (warm-up, acceleration, steady-state operation, deceleration). The oxygen sensor must thus be capable of converting the gas mixture that it receives into a state of complete thermodynamic equilibrium. The resulting requirements are a high level of catalytic activity at the electrode and a protective layer capable of limiting the gas quantity. Should it prove impossible to achieve thermodynamic equilibrium at the electrode, the sensor's lambda signal will be erroneous.

Characteristic Curve. The concentration of residual oxygen fluctuates exponentially (by several orders of 10) in the vicinity of a stoichiometric air/fuel mixture (lambda = 1). In accordance with Eq. (6.4), this leads to massive variations in sensor voltage (lambda = 1 jump) with the characteristic lambda curve (Fig. 6.5).

Reference Atmosphere. Ambient air is the most commonly used O_2 reference. Alternatives include a metal-metal oxide mixture or a pumped reference medium. With this method, an O_2 reference pumping current is superimposed on the voltage measurement to generate an approximately constant O_2 partial pressure at the encapsulated reference electrode.

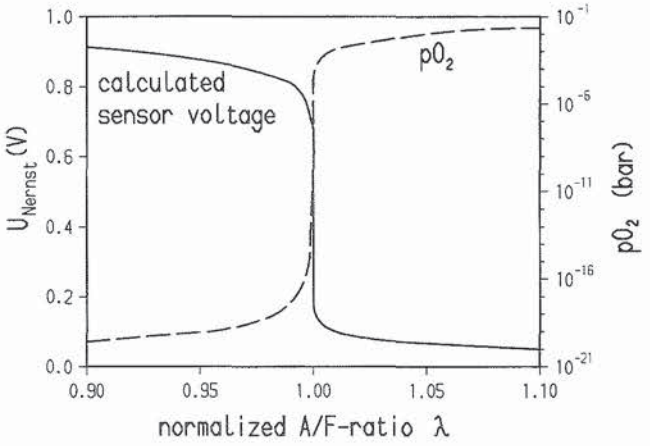


FIGURE 6.5 Equilibrium partial pressure of oxygen pO_2 and resulting curve for Nernst voltage at ≈ 700 °C relative to the normalized air-fuel ratio lambda.

6.2.2 Lambda = 1 Sensor: Semiconductor Type

Sensing Mechanism. Oxidic semiconductors such as TiO_2 and $SrTiO_3$ rapidly achieve equilibrium with the oxygen partial pressure in the surrounding gas phase at relatively low temperatures. Changes in the partial pressure of the adjacent oxygen produce variations in the oxygen vacancy concentration of the material (TiO_{2-x} respectively $SrTiO_{3-x}$), thereby modifying its volume conductivity.

$$R_T = A\ pO_2^{n} \exp\left(\frac{E}{kT}\right) \tag{6.5}$$

where R_T = semiconductor resistance
A = constant
E = activation energy
k = Boltzmann constant
T = absolute temperature
$n \approx 1/4$.

This effect, which is exploited to determine lambda, is also superposed by the conductivity's temperature dependence. Electrical resistance and sensor response times are both inversely proportional to temperature, as equilibrium is achieved more rapidly through diffusion in the oxygen vacancies.

Design. The ability to dispense with an O_2 reference allows an extremely simple design featuring an integrated heater. The porous semiconductor thick-film layer is generally positioned and sintered onto a planar Al_2O_3 substrate between two electrodes. Thin-film layers are also being developed as an alternative.

Characteristic Curve. At lambda = 1, the sensor layer displays an extreme change in conductivity due to the large pO_2 variation (see Sec. 6.2.1 and Fig. 6.5). When new, TiO_2 sensors provide essentially the same response as ZrO_2-based lambda = 1 probes. Variations in the rates of increase for lean and rich resistance and for response time occur over its lifetime, with the emission-control system undergoing a significant shift toward lean.

Principles of Operation. A voltage is applied to the TiO_2 resistor R_T and a serial reference resistor. The voltage drop at the serial resistor depends on R_T, respectively on lambda. In a three-pole version, the applied voltage is taken from the heater voltage; for the four-pole version (insulated ground) a separate supply is employed. Temperature compensation may be required, depending upon the specific application.

6.2.3 Lean A/F Sensor: Nernst Type

Sensing Mechanism. It is always possible to employ a potentiometric lambda = 1 sensor (Sec. 6.2.1) as a lean A/F sensor by using the flat part of the Nernst voltage curve (Sec. 6.4 and Fig. 6.5) to derive the values at lambda > 1. The diminutive characteristic sensor voltages in this lambda range (U_S < 65 mV at lambda > 1.05) mean that stringent requirements for consistency must be met. Minute voltage deviations of just a few mV are sufficient to produce substantial errors in lambda.

Design. For this reason, special sensors must be employed in applications in which quantitative determination of lambda > 1.05 is required. To this end, special electrode manufacturing techniques are used to enhance catalytic activity, while a power heater (18 W) is used in conjunction with a protection tube featuring a lower gas-flow rate to provide a more constant ceramic temperature.

Principles of Operation. A high-output input amplifier is required for measurement of the low voltage levels associated with lean exhaust gases. In vehicle testing, lambda control has proven effective when used together with a classical mixture regulator. Even alternating operation (lambda = 1 and lean) with a sensor in the same vehicle is possible. Depending upon accuracy requirements, lean operation is possible in a range extending to lambda ≈ 1.5.

Nonautomotive Applications. Lambda-controlled operation can be employed to maintain narrow tolerances and to limit drift in oil- and gas-fired units; the benefits include low fuel consumption, enhanced safety, and lower emissions. The use of sensors in stationary power plants—for instance, in central heating plants—is particularly interesting as a potential means of dealing with fluctuations in the calorific content of the fuel.

6.2.4 Lean A/F Sensor: Limiting-Current Type

Sensing Mechanism. An external electrical voltage is applied to two electrodes on a heated ZrO_2 electrolyte to pump O_2 ions from the cathode to the anode (electrochemical O_2 pumping cell) (Fig. 6.6*a*). When a diffusion barrier impedes the flow of O_2 molecules from the exhaust gas to the cathode, the result is current saturation beyond a certain pumping voltage threshold (limiting current condition). Because the resulting limiting current I_l is roughly proportional to the exhaust gas' oxygen concentration c_{O2}

$$I_l = 4FD \frac{Q}{L} c_{O2} \tag{6.6}$$

where F = Faraday's constant
D = diffusion coefficient
Q = effective diffusion cross section
L = effective diffusion length

Design. Figure 6.6*b* provides a schematic illustration of this sensor type featuring planar technology and an integrated electric heater element.

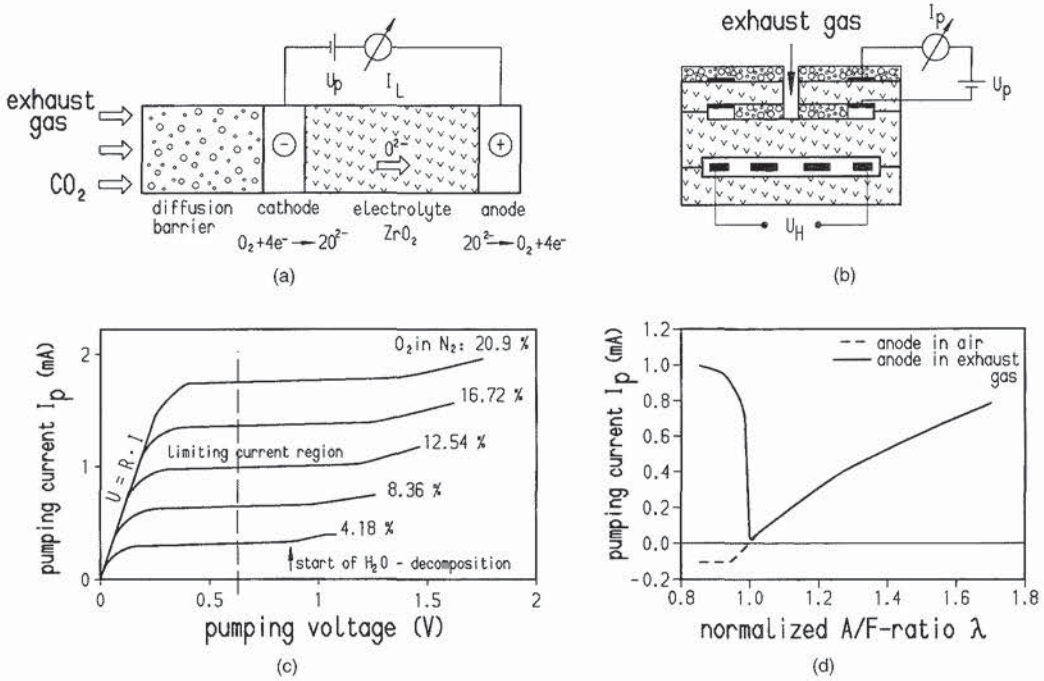


FIGURE 6.6 Limiting current sensor: (*a*) principle of operation; (*b*) schematic diagram; (*c*) current/voltage characteristics; (*d*) response curve as function of lambda.

Characteristic Curve. Figure 6.6*c* shows the current-voltage pattern for a limiting current sensor at various O_2 concentrations. If, at a constant pumping voltage, the pumping current is entered above lambda instead of c_{O2}, the resulting curve will correspond to Fig. 6.6*d*. In lean exhaust gas, the signal drops in response to reductions in lambda as indicated in Eq. (6.7), but below lambda = 1 there is a radical surge. This response pattern is at variance with the basic theory, and can be explained as the result of decomposition effects at the cathode. As a result, this sensor is only suitable for use in lean gases. Depending upon the design of the diffusion barrier (pore radius to free path length for the gas molecules, according to the proportion of gas phases to Knudsen diffusion), various relationships are observed between pumping current (measurement signal) and exhaust pressure and sensor temperature.

Principles of Operation. Depending on the accuracy requirements, a stable operating temperature in the range 600 to 800 °C is needed. Pumping voltages must always be selected to satisfy the conditions for limiting current. In addition, this limiting current has to be calibrated due to the manufacture scattering of the diffusion gap.

6.2.5 Wide-Range A/F Sensor: Single-Cell

Design and Operation. When the anode of a limiting current sensor of the type described in Sec. 6.2.4 is exposed to reference air instead of to the exhaust gas, the total voltage U_S at the probe will be the sum of the effective pumping voltage U_P and a superimposed Nernst voltage U_N.

$$U_S = U_P + U_N \tag{6.7}$$

In operation, holding U_S to, as an example, 500 mV will produce a positive pumping voltage in lean exhaust gases ($U_N < 500$ mV); the diffusion limits the rate at which O_2 is pumped from the cathode to the anode. At lambda = 1 ($U_N \approx 500$ mV), the pumping voltage and, with it, the pumping current drop toward zero. In rich exhaust gas ($U_N > 500$ mV), the effective pumping voltage U_P becomes negative, causing oxygen to be pumped from the reference air electrode to the exhaust gas electrode ($I_P < 0$), where it reacts with the rich gas components CO, H_2, and HC.

Characteristic Curve. The pumping current displays a uniform upward trend between mildly rich and extremely lean lambda values (Fig. 6.6*d*).

Applications. This type of sensor is in series production as a thimble-type design for lean-mix concepts. In this application, operation as a limiting current sensor is restricted to lean conditions; during operation at lambda = 1, the pumping voltage is switched off and the Nernst voltage serves as the control signal.

6.2.6 Wide-Range A/F Sensor: Dual-Cell

Design. Skillful combination of a limiting-current sensor with a Nernst concentration cell on a single substrate will produce a dual-cell wide-range A/F sensor. The pumping and concentration cells are made of ZrO_2. Each cell is coated with two porous platinum electrodes, and they

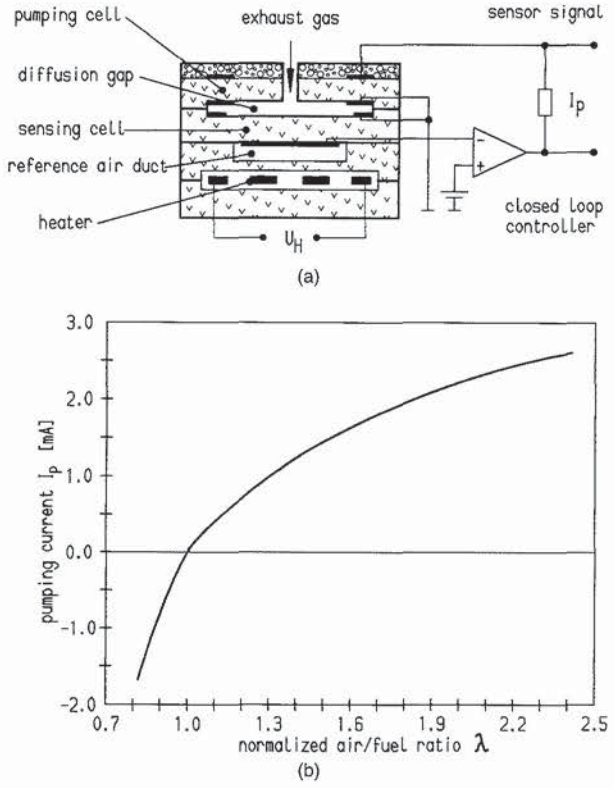


FIGURE 6.7 Dual-cell wide-range A-F sensor: (*a*) schematic diagram; (*b*) response curve as function of lambda ($T_{gas} \approx 400$ °C, $P_H \approx 12$ W).

are arranged with a measurement gap of approximately 10 to 50 μm in height between them. A gas-intake opening in the solid electrolyte connects this measurement gap to the exhaust gas while the gap is serving as a diffusion barrier to control limiting current (Fig. 6.7).

Operation Mode. An electronic circuit regulates the current applied to the pumping cell to maintain a constant gas composition (for instance, lambda = 1) in the measurement gap. This corresponds to a Nernst-cell voltage of $U_N \approx 450$ mV. If the exhaust gas is lean, the pumping cell drives the oxygen outward from the measurement gap. If the exhaust gas is rich, the flow direction is reversed and oxygen from the surrounding exhaust gas (e.g., from decomposition of CO_2 and H_2O) is pumped into the measurement gap.

Characteristic Curve. The pumping current is proportional to the oxygen concentration (or oxygen requirement). With the attendant electronic control circuitry, an unambiguous, linear signal increase over a wide lambda range of 0.7 < lambda < infinite (air) can be obtained if the current is calibrated due to the manufacture scattering (Fig. 6.7). The integrated heating element provides an operating temperature in excess of 600 °C.

6.3 TECHNOLOGY OF CERAMIC EXHAUST GAS SENSORS

6.3.1 ZrO_2-based Sensors

The zirconium dioxide used to manufacture oxygen sensors is not pure; instead, the solid electrolyte consists of a mixed oxide, ZrO_2/Y_2O_3. The Y_2O_3 concentration used for actual applications lies in the 4 to 5 mol% range. This level is commensurate with optimum sensor operation in the important properties of ion conductivity, thermal stability, and mechanical strength.

Ion Conductivity. The substitution of trivalent yttrium ions for quadrivalent zirconium ions leads to the formation of oxygen vacancies through which the oxygen ions can move. Initially, the increases in ion conductivity are roughly proportional to the rise in Y_2O_3 content, as they are accompanied by a simultaneous rise in the number of oxygen vacancies. This pattern is observed with yttrium as well as with other bi- and trivalent alkaline earth metals, at least until a certain level of defect concentration is reached. Beyond this limit, the mutual influence of the vacancies reaches such levels that any further increases in defect concentration are accompanied by reductions in ion conductivity. The maximum is obtained at approximately 9 to 10 mol% Y_2O_3.

Phase Composition and Mechanical Stability. At room temperature, zirconium dioxide is monocline. It becomes tetragonal (metastable) above 1200 °C, and assumes a cubic form beyond 2370 °C. Yttrium doping can be employed to partially or completely stabilize the tetragonal and cubic modification, even at room temperature (PSZ—partially stabilized zirconium; FSZ—fully stabilized zirconium, above approximately 10 mol%, with certain manufacturing processes 7 to 8 mol% Y_2O_3). ZrO_2, with an essentially monocline structure, does not display sufficient resistance to temperature fluctuations for application in automotive exhaust systems—the changes in the geometry of the individual crystallites which accompany the phase transition lead to fatigue symptoms in the ceramic material. At the same time, the cubic phase does not possess the physical strength required to resist the physical shocks and thermal stresses. However, good mechanical characteristics can be obtained by using partially stabilized ceramic materials, with maximum strength being achieved within the range of 2 to 4 mol% Y_2O_3.

Manufacture of Partially Stabilized ZrO_2 Ceramics. The most common procedure consists of mixing and grinding the ZrO_2 and Y_2O_3 together (mixed oxide). The differences in the

length of the diffusion paths—the two oxides are located adjacently as separate powder particles—do not generally produce a completely homogeneous distribution of yttrium in the ZrO_2. Thus, both fully stabilized cubic crystallites and tetragonal crystallites can be formed during the sintering process. During the phase transition to monocline ZrO_2 that takes place during cooling, the latter provoke internal tension, causing the ceramic material to harden. Certain combinations of yttrium concentration and manufacturing techniques can be employed to obtain exclusively tetragonal crystallites, which remain stable even after cooling (TSZ—tetragonal stabilized zirconia). Due to the absence of a phase transition within the temperature range associated with the intended application, stabilized tetragonal zirconium dioxide displays a virtually constant coefficient of thermal expansion. Both of the partially stabilized ZrO_2 ceramics described here are used in the manufacture of oxygen sensors.

6.3.2 Thimble-Type Sensors

Shaping. The ceramic base (Fig. 6.4) is shaped in the usual manner with dry pressing of granulated ZrO_2 and subsequent grinding of the compact material.

Electrodes. The application of the gas-permeable platinum electrodes is one of the most critical processes in the manufacture of oxygen sensors. A basic distinction is made between two different procedures:

Thin-Film Process. In this process, a thin microporous platinum layer (electrode thickness < 1 μm) is applied to the previously sintered base using one of three methods: thermal evaporation, sputtering, or chemical deposition with subsequent thermal treatment.

Thick-Film Process. In this process, a (for example) platinum-cermet layer (mixture of noble metal and ceramic phase) is printed onto the unsintered base (electrode thickness 5 to 10 μm).

As far as actual operation is concerned, the salient difference between the two processes resides in the fact that the three-phase boundary required for the conversion of oxygen (common border of platinum, zirconia, and gas phase) remains restricted to the surface of the base when thin-film electrodes are used. In contrast, the use of thick-film cermet electrodes with a ceramic phase of ZrO_2 results in a substantially larger border area along with enhanced high-temperature adhesion.

Protective Layer. As the outer electrode is exposed to the exhaust gases, a suitable protective layer, combining the requisite porosity with optimal adhesive properties, is essential to ensure the long-term stability of the oxygen sensor. In the conventional plasma or flame-spraying process, a porous layer 50 to 300 μm in thickness (e.g., magnesium spinel) is applied to the sintered and electrodes-equipped base. A cofiring technique with ZrO_2-based protective layers or a combination of both processes is being employed in the production of a new generation of oxygen sensors with enhanced long-term stability.

Well-defined porosity levels must be maintained to ensure satisfactory operation. With cofired protective layers, the required porosity level can be obtained by using organic pore-forming substances, which are burned out during sintering, or by adding a second, sinter-inactive ceramic phase, such as Al_2O_3. Porous protective layers also provide a diffusion barrier for the exhaust-gas molecules before they reach the electrodes. These layers thus exercise a determining influence on the sensor's control characteristics because the components of the exhaust gas which exhibit various diffusion coefficients are not in a state of thermodynamic equilibrium (see Sec. 6.4). This effect can be partially neutralized through defined pigmentation in the protective layers, which not only display catalytic characteristics, but can also inhibit contamination.

6.3.3 Planar Oxygen Sensors

The planar oxygen sensor derives its designation from the fact that, as opposed to the thimble-type sensor, its design arranges all operating layers in plain, consecutive surfaces. The

technology applied in their manufacture is quite similar to the multilayer technology employed to produce ceramic multilayer capacitors and high-density electronic circuit boards (e.g., MCM—multichip modules).

Manufacture of Ceramic Green Sheet. First, an organic binder phase is added to the ZrO_2 powder in a tape-casting process to produce ceramic substrate layers.

Printing on the Ceramic Green Sheet. A screen-printing process (thick-film process) is employed to attach the individual function groups, e.g., galvanic cell, pumping cell, reference-air duct and heater (see also Sec. 6.2) to the green tape in the desired layer sequence. Several printed layers are generally required for each side of the tape; these layers can also consist of various inorganic materials.

Lamination. A number of screen-printed substrates can be stacked and laminated (bonded in a process combining temperature and pressure) to form composite structures representing virtually any desired level of complexity. A simplified version of a lambda = 1 sensor in planar technology is illustrated in Fig. 6.8.

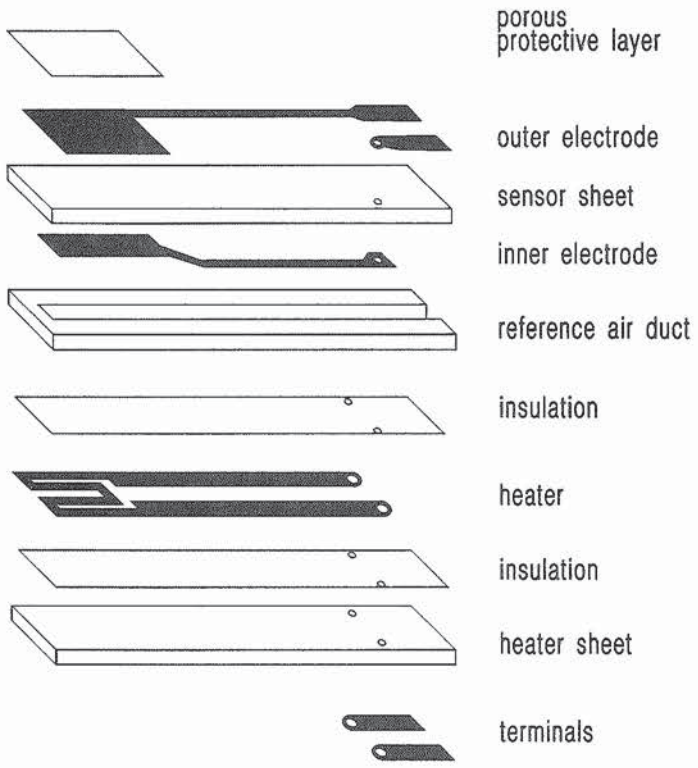


FIGURE 6.8 Layer structure of planar lambda = 1 sensor.

Advantages of Planar Technology. The application of surface-design technology also provides an additional advantage by making it possible to combine a number of individual elements on a single substrate for subsequent separation (multiple application) and sintering to form monolithic sensors with integrated heating elements. The coefficients of thermal expan-

sion of the ZrO_2, cermet, and insulation layers must be precisely matched to ensure that the resulting ceramic elements display adequate thermal and mechanical stability.

Sensor elements manufactured with this new technology offer a number of advantages: the integration of heater and sensor provides a single unit with a lower total thermal mass, allowing substantial reductions in the continuous heating requirement (30 to 40 percent of that required with a thimble-type sensor), shorter delay between activation and readiness for active control, and smaller sensor dimensions. These characteristics make the unit ideal for meeting the demands of the modern motor vehicle (reductions in energy requirements, weight and space savings, and more stringent emission requirements). In addition, planar technology is an essential factor in obtaining complicated operating abilities such as those found in the dual-cell wide-range A/F sensor (see Sec. 6.2.6).

6.3.4 Al_2O_3-based Sensors

Sensors for oxygen and other exhaust gases which employ variations in the conductivity of the sensor material (TiO_2, SnO_2, etc.) to detect changes in the gas concentration are usually designed as planar sensors. The preferred substrate material is Al_2O_3. As the combinations of functions (e.g., heater, electrodes, etc.) are essentially those required with planar ZrO_2 sensors, the manufacturing techniques are also quite similar. Al_2O_3 provides advantages in the form of lower material costs and a simplified internal structure, made possible by the fact that Al_2O_3 maintains its good insulation properties at high temperatures. A disadvantage, particularly in terms of long-term stability, is the difficulty with sintering and sensor-material adhesion, as cofiring technology cannot be employed for this procedure.

6.3.5 Construction

Figure 6.9 shows completely assembled a heated thimble sensor and a planar lambda sensor. Although similar in external appearance—the planar sensor is somewhat shorter and narrower—the two units differ substantially in their respective internal structures. The differences are largely a result of the fact that heater and sensor in the heated thimble-type sensor are two separate components requiring assembly and electrical contact. Various ceramic support components are incorporated to enhance structural integrity and to ensure an impermeable seal between the reference air and the exhaust gases while also providing a degree of impact protection for the sensor unit. On the exhaust-gas side, the sensor elements receive both physical and thermal protection from a protection tube. Due to its basic design configuration, the planar sensor element does not share the rotational symmetry of its thimble-type counterpart. Thus, it must be installed in a specific position; the protection tube features a partial double wall to facilitate orientation. The offset inlet openings provide a uniform flow of exhaust gases past the measurement electrode, regardless of its position relative to the exhaust-gas stream. Thermal insulation from the exhaust gases is also improved.

Although unheated thimble-type sensors are still on the market, they are becoming less important due to the increasing demands for precision of emission control systems.

6.4 FACTORS AFFECTING THE CONTROL CHARACTERISTICS OF LAMBDA = 1 SENSORS

6.4.1 Factors Affecting the Static Curve

Due to the porous protective layer around the electrode, the steep voltage change in the sensor's characteristic curve does not occur at precisely lambda = 1; instead, it is displaced a small

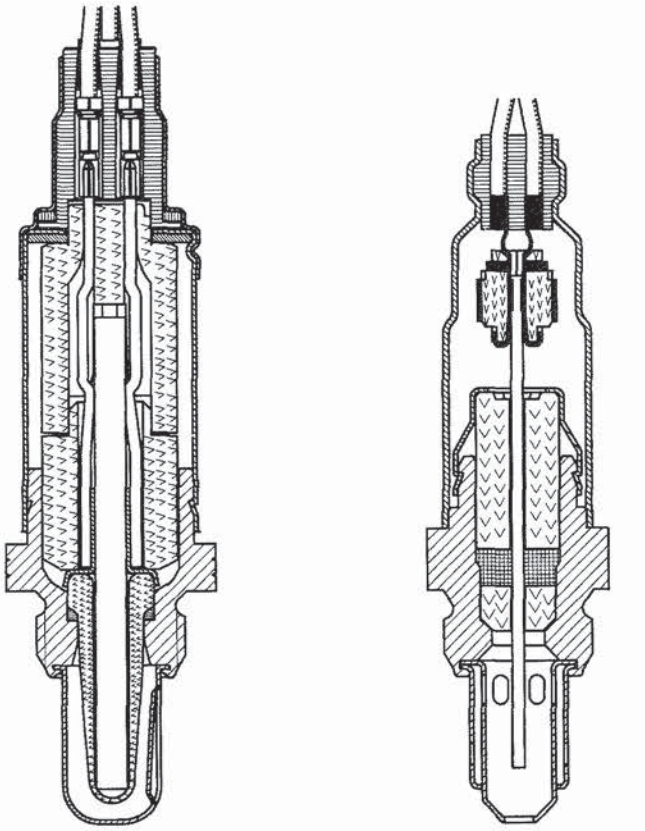


FIGURE 6.9 Construction of thimble-type and planar-type oxygen sensors.

amount into the lean range. The lower the porosity of the protective layer, the greater the diffusion of H_2 relative to O_2 toward the electrode. This phenomenon induces changes in the lambda value at the electrode relative to the lambda of the exhaust gas, the electrode detects richer exhaust gas and the lambda characteristic curve is displaced into the lean range.

Lean Shift. The lean shift is caused by exhaust-gas residue, such as oil ash and SiO_2, which partially plug the pores of the protective layer, and by reductions in sensor temperature.

Rich Shift. Fractures and flake-off of the protective layer lead to reductions in response times and shift the static characteristic curve toward rich. Electrode deactivation, especially due to lead contamination, will produce a flatter static response curve, as free, unconverted oxygen will also be detected. This effect also displaces the switchpoint threshold toward a lower lambda. A similar effect occurs when the reference atmosphere is contaminated (Chemical Shift Down, or CSD), for instance, with exhaust gases or water. Figure 6.10 shows typical characteristic curves for these conditions.

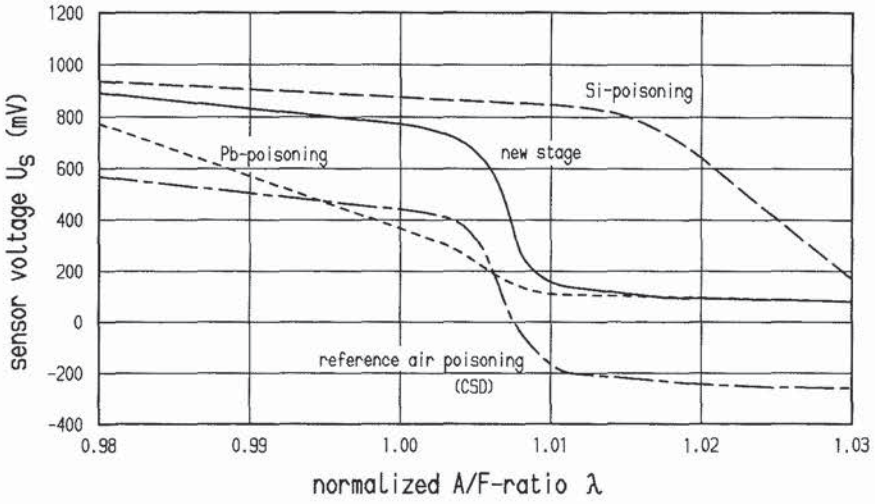


FIGURE 6.10 Effects of contamination on the static response curve of a lambda = 1 sensor as measured at laboratory test stand at $T_{gas} \approx 350$ °C.

6.4.2 Factors Affecting Sensor Dynamics and Control State

Factors. The control state (adjusted lambda) depends upon a number of factors. These include static sensor characteristics, sensor dynamics and asymmetry in sensor response times (switching times), t_{rs} (rich to switchpoint) during the surge from rich to lean, and t_{ls} (lean to switchpoint) during the jump from lean to rich. The sensor switching times t_{rs} and t_{ls} are strongly influenced by the configuration of the protection tube, the protective layer, and the sensor's temperature. Without the protection tube, switching times of less than 10 ms have been measured for ZrO_2 sensors with $T_{ceramic} \approx 900$ °C. However, such rapid response times are neither necessary nor desirable with many of the systems presently in use.

Response Times. An increase in response time that is not accompanied by a variation in the differential between t_{rs} and t_{ls} results in worse emissions with a constant control state due to the fact that the control amplitude increases while the conversion rate of the catalytic converter is reduced (see sec. 6.1.4 and Fig. 6.2).

Asymmetry. An asymmetrical increase in the response times t_{rs} and t_{ls} will also affect the control state due to the lambda shift.

Lean Shift. A pronounced increase in t_{rs} leads to a displacement toward lean. This can occur for any of several reasons. The dwell time in the rich exhaust gas is determined by control frequency and amplitude, and affects t_{rs} through the adsorption of CO and HC. Should the protective layer be obstructed (for instance, by oil ash), or should the sensor temperature drop, t_{rs} will generally rise more than t_{ls}.

Rich Shift. In contrast, a more pronounced rise in t_{ls} will lead to a displacement toward the rich range. The conversion rate of the electrode and the protective layer exercise a major influence on t_{ls}. Lead deposits diminish catalytic activity, causing t_{ls} to rise without exercising any substantive effect on t_{rs}.

Temperature. As a general rule, the susceptibility to sensor contamination is inversely proportional to electrode temperature. The factors determining the temperature of the ceramic material are the heater, the gas temperature, the gas flow rate (rpm, load), and the protection

tube. Depending upon the installation location, the ceramic temperature generally lies between 350 and 1000 °C for exhaust temperatures ranging between 150 and 900 °C. The flow velocity of the gas varies between 2 m/s at idle, 40 m/s at moderate rpm and load, and up to 80 m/s under full load. The influence exercised by gas temperature and the protection tube can be seen in Fig. 6.11. The ceramic temperature has a major effect on the controlled lambda. A protection tube providing a lower gas flow (see Fig. 6.11: 1 hole/4 flaps) and a higher heater

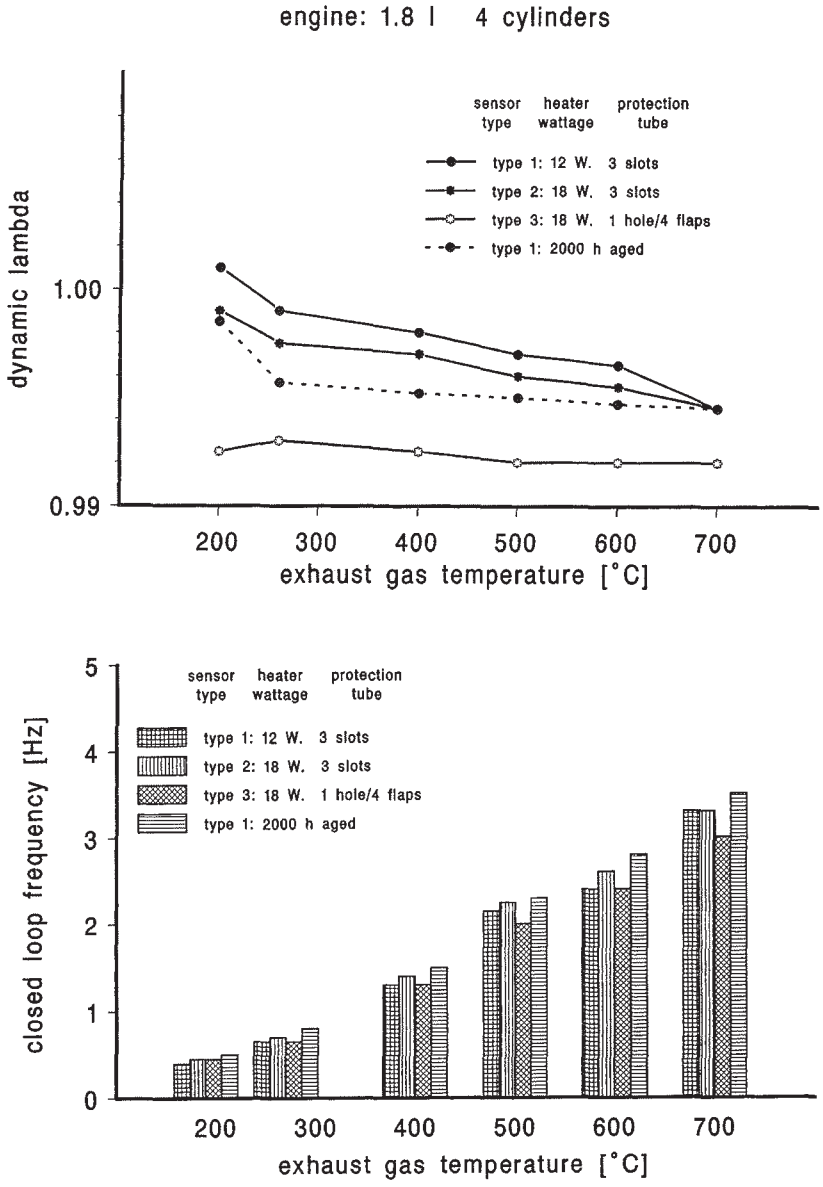


FIGURE 6.11 Effects of exhaust-gas temperature (rpm) on dynamic lambda and control frequency.

power (18 W) will lead to higher ceramic temperatures with enhanced temperature consistency relative to exhaust temperature. This measure is employed to obtain a more stable controlled lambda. The lambda value displays a moderate rich shift compared to that obtained with a more open tube (see Fig. 6.11: 3 slots) and 12 W heater power.

System Lag. System lag is dependent on the sensor's installation position (0.2 to 2 m downstream from the exhaust manifold, with approx. 1 meter downstream from the exhaust valve being typical) and the gas velocity as determined by load, rpm, and the exhaust pipe diameter. It can be up to 500 ms at idle, and drops down to delays as low as 20 ms at full load. The sensor (protection tube, heater power) is of secondary significance (Fig. 6.11). The influence of sensor response times (t_{rs}, t_{ls}) on the control state increases as a function of their share of the total system lag.

tv Shift. The cumulative effect of all extraneous influences is a slight lean control state of the lambda = 1 sensor. Electronic compensation for the lean shift is available in the form of an asymmetrical control-oscillation pattern (see Sec. 6.1.4). This is provided either by means of a switching delay after the sensor surge (tv shift, Fig. 6.3), through variations in the controller's surge amplitude, and/or via asymmetrical ramp slope. Those parameters for the electronic lambda displacement which vary according to load and rpm can be stored as a program map in the electronic control unit (ECU). This method can be employed for lambda corrections up to $|\Delta\lambda| \leq 0.015$.

6.5 APPLICATIONS

6.5.1 Operating Conditions

Environmental Factors. Table 6.3 provides a summary of the types and magnitudes of the major stress factors affecting the sensor. Beyond the actual function of the sensor itself, optimal selection and precise mixture of all applied materials (ceramics, metals, synthetics) are essential factors in obtaining the thermal, mechanical, and chemical characteristics needed to achieve the obligatory extended service life of over 100,000 miles.

Sensor Installation Location. Because it is of determining significance for both external influences and control function, correct sensor location is a critical factor in calibrating the performance of the total lambda-control system. The sensor must be far enough from the exhaust valves to ensure that it monitors representative exhaust gases for all cylinders, but near enough to hold system lag to acceptable levels. Both factors affect emissions.

Thermal Stress. Thermal stresses at the sensor stem from the high-potential exhaust-gas temperatures as well as from the temperature gradients that occur. Exhaust temperature is essentially a function of engine speed, and can reach a maximum of 1000 °C at the sensor on high-performance engines. The most extreme temperature gradients at the sensor occur during hard driving directly in the wake of a warm start at which the exhaust system and the sensor have already cooled off. These conditions can produce exhaust-gas temperature gradients of up to 500 K/s.

Thermal Shock. In the warm-up phase following a cold start, condensation on the pipe walls can cause the ceramic material to crack once the electric heater warms it beyond the critical temperature of approximately 300 °C. Accumulations of condensation can be largely avoided by locating the sensor as close as possible to the engine to provide rapid heating of the upstream exhaust pipe. Depending upon the specific vehicle, a time delay for switching on the sensor heater may be required.

TABLE 6.3 Overview of Stress Factors for Exhaust Gas Sensors

Mechanical stress	Stress caused by environmental conditions	Thermal stress	Stress caused by exposition to exhaust gas
Vibrational stress (<1300 m/s²) by engine vibration (<5 kHz) by pulsation of the exhaust gas (< ±300 mbar) by ambient wind (cable, <10 Hz)	Splash water, gush of water (in case containing NaCl, $CaCl_2$, $MgCl_2$)	Exhaust gas temperature (150 °C ... 1000 °C)	Exhaust gas composition ch between lambda = 0.85 (fuel load enrichment) and lambda infinite (overrun fuel)
Stone impact (<1.5 Nm)	Dust (organic, inorganic origin)	Exhaust gas temperature gradients at cold start (<500 K/s) at overrun fuel cutoff (<200 K/s)	Catalyst poisoning from fuel additives: permissible lead content: <0.013 g/l sulfur content: <0.1 % Br-, Cl-compounds
Installation/removal (torque about 50 Nm when assembled)	Oil, dirt undercarriage protection	Temperature increase after engine stalling (<300 °C at cable outlet)	Oil ash (Ca-, P-, S-, Zn-compounds up to 1 kg within ~60000)
Wire pull test (>70 N)	Fluctuation of the electrical supply system (9 ... 15 V)	Radiant heat (may require corresponding protective measures)	Miscellaneous corrosion products (e.g., Fe-oxides)
Handling (shock load up to 1000 g)	EMC (<200 V/m)		Condensation water
	Variation of climatic conditions (–40 °C ... +50 °C, 10 ... 100% rel. humidity)		

Mechanical Stresses. The main sources of mechanical stress on the sensor are vibration, exhaust-gas pulsation, stone impact, and cable tension (Table 6.3).

Sealing. The cable outlet of the sensor must be sealed to prevent water splash from penetrating to its interior and distorting the reference air (see Sec. 6.4.1 and Fig. 6.10). At the same time, the sealing element must be flexible enough to absorb the accelerative forces associated with vibration as well as any tensile forces transmitted through the cable.

6.5.2 Emission Certification

It would be impossible for a vehicle equipped only with an unregulated catalytic converter and lacking an oxygen sensor to achieve compliance with the American emission standards. When a regulated catalytic converter is used together with an oxygen sensor, emissions remain well below the limits. Even after vehicle mileage corresponding to more than 100,000 miles, the sensor limits emission increases to a minimum (Fig. 6.12). The introduction of a new system must be accompanied by precise calibration of the engine-control system to match the production tolerances and aging characteristics which affect the sensor's control response. The considerable effects that temperature and aging exercise on the catalytic converter must also be considered.

Durability Tests

Life Cycle. To determine aging behavior, the sensors undergo a standard durability test in the exhaust-gas stream of a test engine. The test cycle is designed to simulate combined urban and rural operation, and has been adopted by many automobile manufacturers as a standard program. The mean average speed is approximately 100 km/h for a midrange car. All newly developed products must withstand 2000 h (more than 100,000 miles) operation in this program without failure.

Hot Durability Testing. In a substantially more demanding test, essentially consisting of simulated full-load operation, the sensors must survive at least 1000 h (roughly 100,000 miles). The main stress factors for the sensors in these tests are thermal exposure (exhaust-gas temperatures reaching 950 °C, housing temperatures up to 650 °C), exhaust gas loads, and vibration.

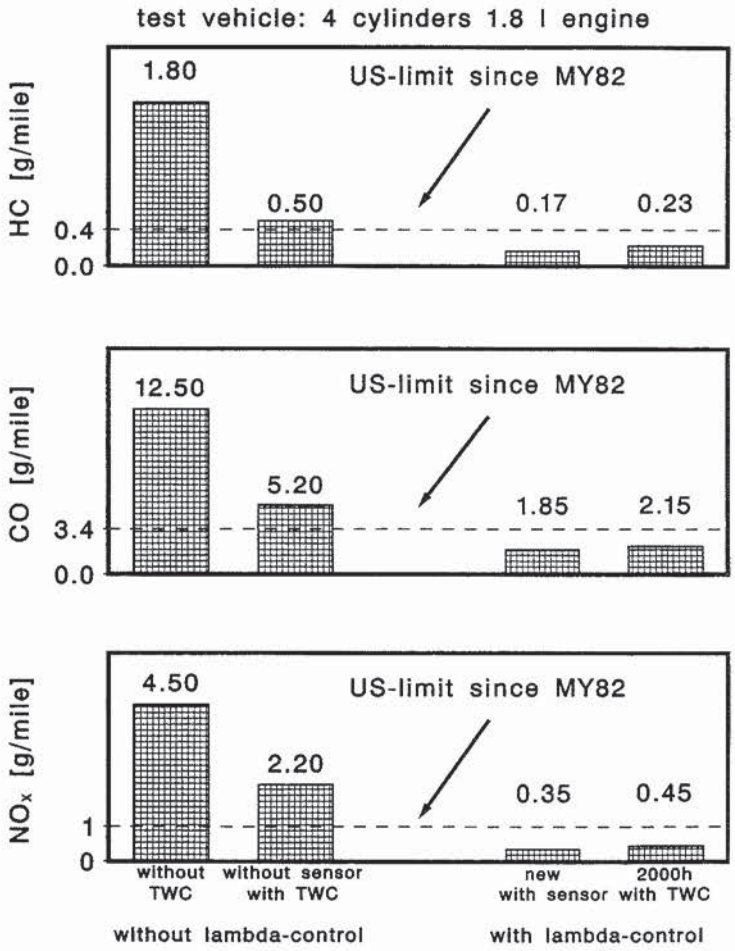


FIGURE 6.12 Comparison of emissions levels without catalytic converter, with unregulated converter, with controlled converter (new and old sensors, converter approx. 10,000 km).

Durability Testing of Contamination Resistance. Short durability tests featuring higher levels of Pb, Si, and oil are employed to determine the sensor's resistance to contamination damage.

6.6 SENSOR PRINCIPLES FOR OTHER EXHAUST GAS COMPONENTS

Sensors capable of monitoring the levels of the regulated toxic exhaust substances—CO, NO_x, and HC—would be desirable as elements of the On-Board Diagnosis systems (OBD I and OBD II) specified by the California Air Resources Board. Due to the extreme operating con-

ditions, only a few of the many tested configurations have a chance to be used in engine-exhaust gas.

6.6.1 Mixed-Potential Sensors

Sensing Mechanism. If reduced catalytic activity prevents gas equilibrium from being achieved at the electrode of a galvanic ZrO_2 cell, competing reactions can occur. These, in turn, prevent a state of reduction/oxidation equilibrium in the oxygen, and lead to the formation of a mixed potential. This potential is determined by electrode activity, temperature, and gas composition. It is difficult to design manufacturing techniques that can consistently produce electrodes capable of maintaining specific, defined rates of catalytic activity over extended periods of time. Every change in electrode activity (for instance, due to aging) leads to a variation in mixed potential. Pt electrodes give rise to a mixed potential at very low temperatures, with extended response delays relative to lambda sensors. Other electrode materials with lower rates of catalytic activity continue to provide a mixed potential at higher temperatures, making it possible to achieve response times < 1 s.

Selectivity. Careful selection of electrode materials, operating temperature, and selective precatalyzation layers offers opportunities for improving selectivity.

Operation Mode. Because the effect is temperature-sensitive, a constant sensor temperature must be regulated in the range of 300 to 600 °C, depending upon the individual application.

6.6.2 Semiconductor Gas Sensors

Sensing Mechanism

Ceramic and Thick-Film Layers. On the surface of nonstoichiometric metal oxides such as SnO_2, TiO_2, In_2O_3, and Fe_2O_3 (*n*-type semiconductors), oxygen is absorbed and dissociated in air at high temperatures and is bonded to the crystal lattice. This leads to the formation of a thin depletion layer at the crystallite surface with a resultant arch in the potential curve. This phenomenon produces reductions in surface conductivity and higher intercrystalline resistance at the boundaries between two crystallites; this is the major factor determining the total resistance of the polycrystalline metal oxide. Oxidation gases such as CO, H_2, and C_xH_y, which react with surface oxygen, increase the density of charge carriers in the boundary layer and reduce the potential barrier. Reduction gases, such as NO_x and SO_x raise the potential barrier and, with it, the surface/intercrystalline resistance.

Thin-Film Layers. Compared to porous polycrystalline materials, thin-film layers display only limited numbers of crystalline boundaries at the layer surface for reaction with the exhaust gases. The depletion-layer barrier makes up a substantial proportion of the layer thickness on thin-film devices, meaning that changes in charge-carrier density within the barrier layer due to adsorbed gases will also result in substantial variations in the total resistance.

Selectivity. Selected doping materials and temperatures can be employed to achieve selectivity for CO, HC, or NO_x. The resistance of metal-oxide semiconductors is always a function of the partial pressure of the O_2.

Specific Influences of Exhaust Gas. Engine exhaust gas with minimal, highly variable O_2 partial pressures is characterized by a high level of O_2 cross-axis sensitivity. Within the lambda = 1 to lambda < 1 operating range, an irreversible long-term change in sensor resistance leading to disintegration of the metal oxide is also possible. High operating temperatures are conducive to the diffusion of oxygen vacancies and doping materials. When amplified by sintering effects, these lead to resistance drift and attenuated sensor response.

Standard operating temperatures for metal-oxide gas sensors range from 100 to 500 °C, with up to 600 °C being seen in some restricted applications. With SnO_2, a high level of selectivity for, for example, CO can be achieved at temperatures as low as the 100 to 200 °C range. However, the response times can be measured in minutes. At higher temperatures, in the 300 to 550 °C range, response times of less than 1 s can be obtained, but with reduced selectivity.

6.6.3 Catalytic Gas Sensors

Sensing Mechanism. The catalytic gas sensor is essentially a temperature sensor featuring a catalytically active surface. An exothermic reaction (basically an oxidation reaction in air) at the catalytically active surface causes the temperature of the sensor to rise. The increase in temperature is proportional to the concentration of an oxidation gas in an excess-oxygen atmosphere. A temperature sensor of the same basic design—but without the catalytic response—is employed to enhance sensitivity and to provide temperature compensation by means of differential-signal evaluation. Materials employed for the temperature sensor include coiled Pt wire, Pt thin-film and thick-film layers, transistors, and NTC and PTC thermistors.

Operating Conditions. Catalytic gas sensors are relatively insensitive and can be employed in a range >1000 ppm. Excess oxygen is required to monitor concentrations. Due to the sensitivity to flow rate, these units are generally operated in flowheads with diffusion limiters. Measurement and reference sensors must be exposed to the same flow conditions, with no mutual thermal influences. Symmetrical configurations in pellet form are thus preferred over planar shapes. Operating temperatures for commercial sensors are in the 500 to 600 °C range, but extension to higher temperatures is possible. The requirement for long-term stability in the catalytic converter is the limiting factor governing the operating temperature. Catalytic gas sensors are not selective, instead providing a summing signal for all combustion gases. Thus catalytic sensors for application in exhaust gases are restricted to monitoring the state of the catalytic converter.

BIBLIOGRAPHY

Arndt, J., "Ceramics and Oxides," in W. Göpel, J. Hesse, and J. N. Zemel (eds.), *Sensors—Fundamentals and General Aspects,* vol. 1, S.252 ff, VCH Verlagsgesellschaft, Weinheim, 1989.

Barnes, G. J., R. L. Klimisch, and B. B. Krieger, "Equilibrium Considerations in Catalytic Emission Control," SAE Paper 730200, *SAE National Automobile Engineering Meeting,* Detroit, Jan. 1973.

Bosch, *Automotive Electric/Electronic Systems,* VDI, Düsseldorf, May 1987.

Bosch, *Automotive Handbook,* VDI-Verlag, Düsseldorf, 1986.

Duecker, H., K.-H. Friese, and W.-D. Haecker, "Ceramic Aspects of the Bosch Lambda-Sensor," SAE Paper 750223, *SAE National Automobile Engineering Meeting,* Detroit, 1975.

Gruber, H. U., and H. M. Wiedenmann, "Three Years Field Experience with the Lambda-Sensor in Automotive Control Systems," SAE Paper 800017, *SAE National Automobile Engineering Meeting,* Detroit, Feb. 1980.

Moseley, P. T., and B. C. Tofield (eds.), *Solid State Gas Sensors,* Adam Hilger, Bristol and Philadelphia, 1987.

Saji, K., "Characteristics of Limiting Current-Type Oxygen Sensor," *J. Electrochem. Soc.: Electrochem. Science and Technology,* vol. 134, no. 10, 1987, pp. 534–542.

Saji, K., H. Kondo, T. Takeuchi, and I. Igarashi, "EMF Characteristics of Zirconia Oxygen Sensor in Non-equilibrium Gas Mixtures Containing Combustible Gas and Oxygen," *Proc. 1st Sensor Symposium,* Tsukuba, Japan, June 1981, pp. 103–107.

Soejima, S., and S. Mase, "Multi-Layered Zirconia Oxygen Sensor for Lean Burn Engine Application," SAE Paper 850378, *SAE National Automobile Engineering Meeting,* Detroit, 1985.

Takami, A., "Development of Titania Heated Exhaust-Gas Oxygen Sensor," *Ceramic Bulletin,* vol. 67, no. 12, 1988, pp. 1956–1960.

Wiedenmann, H. M., "Characteristics of Oxygen Sensors for Lean Exhaust Gas," *VDI Berichte* 578, S.129–151, VDI-Verlag, Düsseldorf, 1985.

ABOUT THE AUTHORS

HANS-MARTIN WIEDENMANN studied physics and meterology in Tübingen and Munich, where he received his Ph.D. in Physics in 1969. He joined Bosch in 1970 and is now senior manager for the development of lambda oxygen sensors.

GERHARD HÖTZEL studied chemistry in Konstanz and received his Ph.D. in 1986 at the University of Stuttgart. At Bosch, he is responsible for physical bases and measurement techniques for lambda oxygen sensors.

HARALD NEUMANN studied solid-state electronics in Aachen where he received a Ph.D. in 1988. Since that time, at Bosch he has been responsible for the technology of planar-type ceramic sensors.

JOHANN RIEGEL studied electrical engineering in Karlsruhe and received a Ph.D. in 1989. At Bosch he is responsible for the investigation of planar-type ceramic sensors.

HELMUT WEYL studied mechanical engineering in Darmstadt and was graduated from there in 1966. He has been with Bosch since 1968 and is now responsible for proving lambda oxygen sensors.

CHAPTER 7
SPEED AND ACCELERATION SENSORS

William C. Dunn
Motorola Inc.
Semiconductor Products Sector

7.1 INTRODUCTION

In the automotive arena, speed and acceleration sensors are used in a wide variety of applications, from improving engine performance through safety to helping to provide creature comforts.

Speed sensing can be divided into rotational and linear applications. Rotational speed sensing has two major application areas: engine speed monitoring to enhance engine control, and performance and antilock breaking and traction control systems for improved road handling and safety. Linear sensing can be used for ground-speed monitoring for vehicle control, obstacle detection, and crash avoidance. Acceleration sensors are used in air bag deployment, ride control, antilock brake, traction, and inertial navigation systems.

In most cases, there are a number of different sensor types available for a specific monitoring function. However, the choice of sensor for a specific application can be difficult to make. The selection may be determined by the familiarity of the system's designer with the sensor. On the other hand, the output from one sensor can be used for several applications, and the individual requirements of each application may eventually determine the sensor to be used.

Electronics and electronic sensors are making rapid inroads into the automotive market. In order to analyze the large amounts of sensor data needed for low emissions and efficient engine control, it is necessary to process the information using microcontrollers (MCUs), which can operate at high speeds and in real time. Sensors that can convert information directly into a digital format for MCU compatibility have a distinct advantage over an analog output format. Digital signals are also supply-line voltage-insensitive, virtually unaffected by noise, and have better resolution than can be obtained with analog signals. If the addition of an analog-to-digital converter (on or off the MCU) is required for compatibility, the system cost is increased. The accuracy of the control system is only as good as the integrity of the sensor data supplied to the MCU. Hence, the importance of the performance of the sensor.

This chapter concentrates on speed and acceleration, and therefore does not go into all the other different types of applications for which many of these sensors can be used in the automobile.

7.2 SPEED-SENSING DEVICES

In automotive applications, the environment must be taken into consideration. The sensing must be accurate, the devices must be rugged and reliable, and they must function in the presence of oil, grease, dirt, and inclement weather conditions. These requirements have severely limited the use of a number of otherwise practical alternatives, such as optical sensors and contact sensing.

In the area of rotational speed monitoring, the most practical devices use magnetic field sensing. These sensors are Hall effect devices, variable reluctance (VR), and magnetoresistive or magnetic resistance element (MRE) devices. Both the Hall effect and VR devices have been widely used and have a proven track record. The MRE device has only recently come into its own with improved technology and provides a viable alternative to the Hall effect device. The MRE device has a higher sensitivity and a wider operating temperature range than the Hall effect device.

For the measurement of ground speed and object detection, optical, radar, laser, infrared, and ultrasonics have been explored. Linear sensing devices typically use the Doppler effect for speed sensing and pulse modulation for distance measuring. These devices are used for object detection in blind spots when reversing or changing lanes, and in such applications as collision avoidance systems.

7.2.1 Variable Reluctance Devices

Variable reluctance devices are in effect small ac generators with an output voltage proportional to speed, but they are limited in applications where zero speed sensing is required. The operating frequency range of the VR device is from about 10 Hz to 50 kHz. It is insensitive to mechanical stress and has a wide temperature operating range from –40 to 190 °C. The supply voltage and offset drift will depend upon the control electronics. The VR device, originally designed around existing automotive electromechanical systems, was adapted for electronic control. The ferrous metal in the VR system is designed for maximum output voltage at low rpm (revolutions per minute), to get as close as possible to zero speed sensing without generating excessive voltages at maximum rpm (up to 150 V). This device gives a linear output voltage with frequency. Most systems use MCUs for data processing, so the VR device needs an analog-to-digital (A/D) converter to generate a digital signal for compatibility with the MCU. Although the VR device itself is inexpensive, the extra costs for data conversion may eventually lead to its demise in many automotive applications.

7.2.2 Hall Effect Devices

The Hall effect exists when a current flowing in a carrier experiencing a magnetic field perpendicular to the direction of current flow results in the current being deflected perpendicular to the field and to the direction of the current. The Hall effect is shown in Fig. 7.1. A current I_c flowing through the device between terminals 1 and 2 will produce a potential VH between terminals 3 and 4, when a magnetic field **B** is applied perpendicular to the device. The potential VH is determined by the strength of the magnetic field and the current flowing. Hall effect devices can be manufactured with indium, gallium arsenide, or silicon. A comparison of their properties is given in Table 7.1.

As can be seen, silicon is the most sensitive material. It is compatible with ICs and has a wide operating temperature range. The Hall effect device is well known both in industry and in the automotive arena for rotational and position-sensing applications. However, recent developments in Hall sensing devices, such as differential sensing and integration, have given improved sensor characteristics, which may result in greater potential in automotive applications.

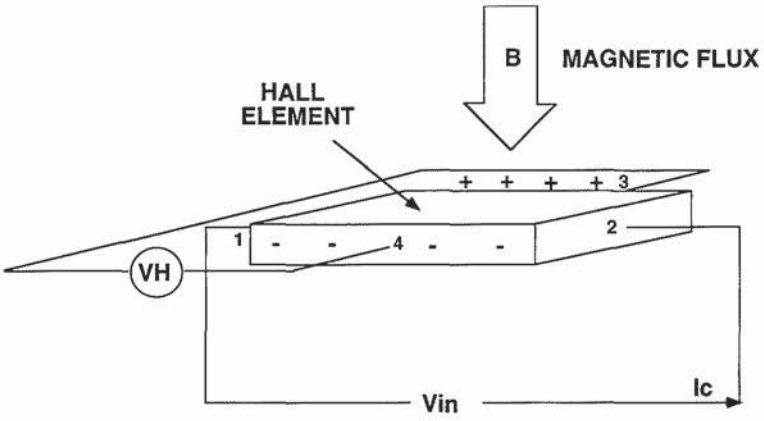


FIGURE 7.1 Hall effect.

The Hall effect device is very versatile, flexible in use, easy to package, and can be used for zero speed sensing (it can give an output when there is no rotation). Hall devices give a frequency output that is proportional to speed, making them compatible with MCUs. The Hall device is normally configured as a bridge to minimize temperature effects and to increase the sensitivity of the sensor.

A typical Hall effect sensor configuration with waveforms is shown in Fig. 7.2. The teeth of the ferrous wheel concentrate the magnetic flux when the teeth come into close proximity to the Hall sensor and magnet. The output from the sensor is a sinusoidal waveform, whose frequency is the rpm of the ferrous wheel multiplied by the number of teeth on the wheel. The resolution of the system depends on the number of teeth in the wheel (typically 20).

7.2.3 Magnetoresistive Devices

The magnetoresistive effect is the property of a current-carrying ferromagnetic material to change its resistivity in the presence of an external magnetic field. For example, a ferromagnetic (permalloy) element (an alloy containing 20 percent iron and 80 percent nickel) will change its resistivity by 2 to 3 percent when it experiences a 90° rotation in a magnetic field.[1] The resistivity value rises to a maximum when the direction of current and magnetic field are coincident, and is a minimum when the fields are perpendicular to each other. This relationship is shown in Fig. 7.3. This attribute is known as the Anisotropic Magneto Resistive Effect. The resistance R of an element is related to the angle q between the current and the magnetic field directions by the expression:

$$R = R_{\|} \cos^2 q + R_l \sin^2 q \tag{7.1}$$

TABLE 7.1 Comparison of Properties of Hall Effect Materials

Material	Operating temp. °C	Supply voltage	Sensitivity @ 1 kA/m	Frequency range
Indium	–40 to 100	1V	7 mV	0 to 1 MHz
GaAs	–40 to 150	5 V	1.2 mV	0 to 1 MHz
Si (with conditioning)	–40 to 150	12 V	94 mV	0 to 100 kHz

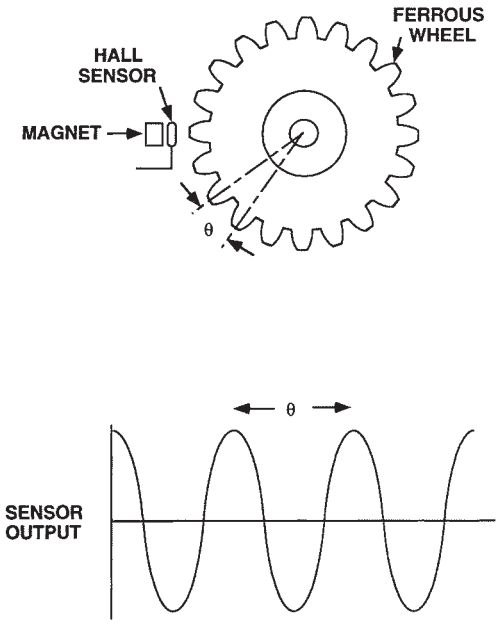


FIGURE 7.2 Hall sensor and output waveforms.

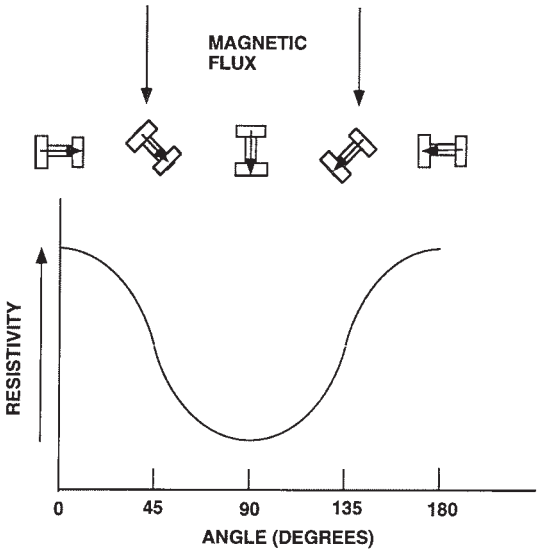


FIGURE 7.3 Relationship between magnetic field and resistivity change in MRE devices.

where $R_{\parallel}$ is the resistance when the current and the magnetic field directions are parallel and $R_{\perp}$ is the resistance when the current and the magnetic field directions are perpendicular.

MRE devices give an output when stationary, which make them suitable for zero speed sensing. MRE devices also give an output frequency that is proportional to speed, making for ease of interfacing with an MCU. For good sensitivity and to minimize temperature effects, a bridge configuration is normally used. In an MRE sensor, aluminum strips can be put across the permalloy element to linearize the device. This configuration is shown in Fig. 7.4 together with a typical MRE characteristic. The low-resistance aluminum stripes cause the current to flow at 45° in the permalloy element, which biases the element into a linear operating region.

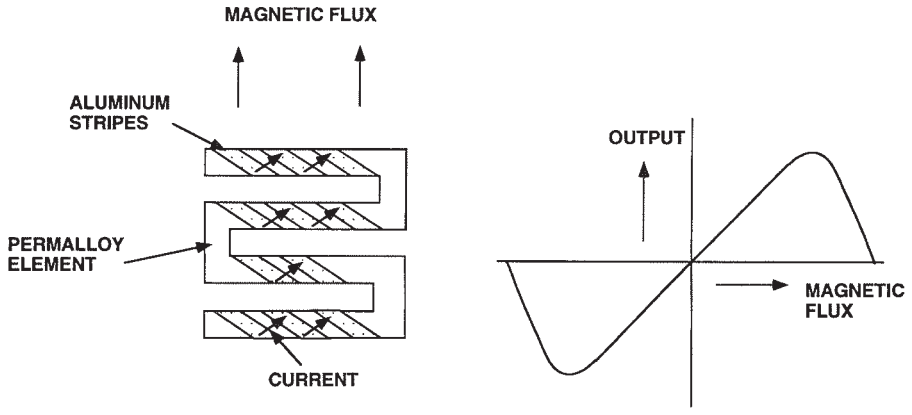


FIGURE 7.4 Use of permalloy strip for linearization in MRE devices.

Integrated MRE devices can typically operate from –40 to 150 °C, over a supply voltage range of 8 to 16 V, and at frequencies from 0 to 1 MHz. In comparing the MRE sensor to a Hall effect device, the MRE has a higher sensitivity, is less prone to mechanical stress, has a wider operating frequency range, has the potential of being more cost effective, gives better linearity, and is more reproducible. However, it is more sensitive to external magnetic fields. Table 7.2 shows a comparison of rotational sensing devices.

TABLE 7.2 Comparison of Sensing Devices

Sensor type	Operating temperature (°C)	Sensitivity 1 kA/m (mV)	Frequency range	Mechanical stress
Hall effect	–40 to 150	90	0 to 100 kHz	High
MRE	–40 to 150	140	0 to 1 MHz	Low
VR	–40 to 190		1 to 50 kHz	None
Magnetic transistor	–40 to 150	250	0 to 500 kHz	Low

It should be noted that Hall effect and MRE devices have many applications in the automobile outside of rotational speed sensing, such as position sensing, fuel-level sensing, and active suspension. The magnetic transistor is showing potential in rotational speed sensing and position-sensing applications, and may eventually be another viable contender to the Hall effect device in the automotive market.

7.2.4 Ultrasonic Devices

Ultrasonic devices can be used to measure distance, ground speed, and as a proximity detector. To give direction and beam shape, the signals are transmitted and received via specially configured horns and orifices. The transmitter and receiver horns are similar in shape, but are normally separate to accommodate different characteristics. The ultrasonic devices are made from PZT crystal-oriented piezoelectric material ($PbZrO_3$—$PbTiO_3$).

For the measurement of distance or object detection, a pulse of ultrasonic energy is transmitted and the time is measured for the reflected pulse to return to the receiver. The frequency of the transmitted ultrasonic waves are typically about 40 kHz and travel with a velocity of 340 m/s at 15 °C. This velocity changes with temperature and pressure. However, these parameters can be measured and corrections made if high accuracy is required. The repetition frequency and power requirements depend on the distance to be measured. To measure speed, the distance variation with time can be measured and the velocity calculated. A more common method is to use the Doppler effect, which is a change in the transmitted frequency as detected by the receiver due to motion of the target (or, in this case, the motion of the transmitter and receiver).

7.2.5 Optical and Radio Frequency Devices

Optical devices are still being used for rotational speed sensing. They are normally light-emitting diodes (LEDs) with optical sensors. Figure 7.5 shows a typical optical sensor system. An optical sensor detects light from an LED through a series of slits cut into the rotating disc, so that the output from the sensor is a pulse train whose frequency is equal to the rpm of the disc multiplied by the number of slits. The higher the number of slits in the disc, the smaller the angle of rotation that can be measured. The optical sensor can be a single photodiode or a photodiode array as shown. This array gives a more accurate determination of the position of the slit, resulting in higher resolution of the position of the disc.

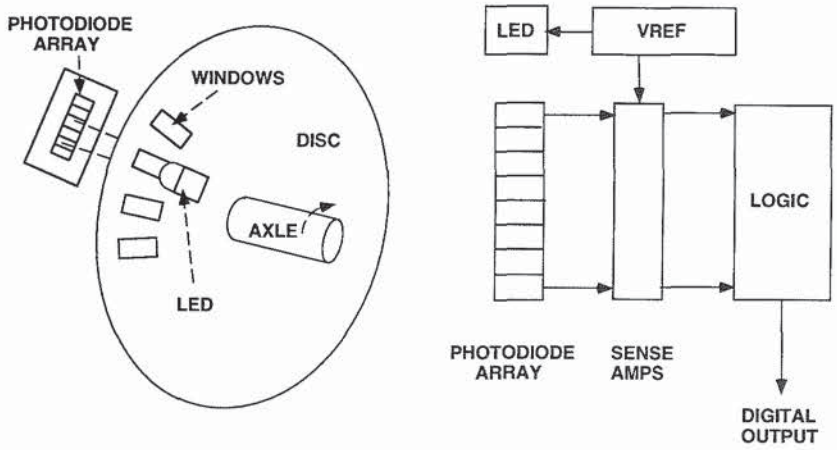


FIGURE 7.5 Optical sensor.

Optical and radio frequency (RF) devices are used for object detection, linear approach speed, and distance measurements in crash avoidance systems where distances greater than about 10 m are involved. These devices use the same principles as the ultrasonic devices. Optical devices normally use lasers or infrared devices for the transmitting source and optical sensors for the receivers. RF devices use gallium arsenide or Gunn devices to obtain the power and high frequency (about 100 GHz) required in the transmitter. The high operating frequency is set to a large extent by the need for a small antenna. These applications are under development and are discussed in Sec. 7.7.2.

7.3 AUTOMOTIVE APPLICATIONS FOR SPEED SENSING

There are several applications for rotational speed sensing. First it is necessary to monitor engine speed. This information is used for transmission control, engine control, cruise control, and possibly for a tachometer. Electronics and electronic sensing in the automobile were brought about by the need for higher-efficiency engines, better fuel economy, increased power and performance, and lower emissions. Second, wheel speed sensing is required for use in transmissions, cruise control, speedometers, antilock brake systems (ABS), traction control (ASR), variable ratio power steering assist, four-wheel steering, and possibly in inertial navigation and air bag deployment applications.

Linear speed sensing can be used to measure the ground speed. This measurement also has the possibly of use in ABS, ASR, and inertial navigation. Similar types of sensors can be used in crash avoidance, proximity, and obstacle detection applications.

7.3.1 Rotational Applications

The high timing accuracy that can be obtained with fuel injection systems and replacement of points by sensors have made cost-effective engine control and low maintenance a reality. Adjustment of the stoichiometric ratio of air to fuel, accurate ignition timing, and oxygen sensors in the exhaust system, have vastly improved engine performance and greatly reduced emissions over widely varying operating conditions. The two important factors in engine con-

trol are the engine speed in rpm and the crank angle. These signals are used by the engine control MCU for determination of fuel injection and ignition timing. The engine rpm measurement range is from 50 to (say) 8000 rpm. A resolution of about 10 rpm is required for an accuracy of about 0.2 percent. For injection and ignition control in a six-cylinder engine, the interval between combustion at maximum rpm is 2.5 ms, so that this time sets the injection period. In practice, a crank angle accuracy of between 1 and 2 degrees per revolution is required. Newer systems with sequential fuel injection, may also require information on TDC (top dead center) for each cylinder to determine the timing. With the low frequencies involved in this application, either Hall effect or MRE devices can be used for monitoring both the engine rpm and crank angle.

Vehicle speed measurements are in the range 0 to 180 km/h (120 mph) and digital displays must have an accuracy of 1 km/h. Some systems have a mechanical pickoff from the drive shaft, which can then use optical sensors for the measurement of road speed. However, newer systems have a pickoff located directly on the drive shaft, which makes optical devices less practical. It is preferred to eliminate the remote sensing via mechanical coupling to save the cost of the associated mechanical components, seals, maintenance, and so on. One method of pickoff is a ring magnet with between 4 and 20 magnetic poles (depending on the required resolution). Figure 7.6 shows such a system using an MRE sensing device. The magnetic flux changes are sensed by an MRE bridge sensor when the magnet disc is rotating. The bridge is supplied from a voltage reference circuit, and its output is amplified and shaped to give a frequency output that is proportional to shaft rotation speed. A ferrous toothed wheel pickoff with magnet and flux concentrator can also be used (see Fig. 7.2). Vehicle speed sensing can be performed with Hall effect, MRE, or VR devices. The number of pulses P per second from the detector are counted to measure speed S, from the following relationship:

$$P = N \times S \times K \tag{7.2}$$

where N = the number of magnetic poles on ring magnet or wheel teeth
K = a constant determined by axle ratio and wheel size

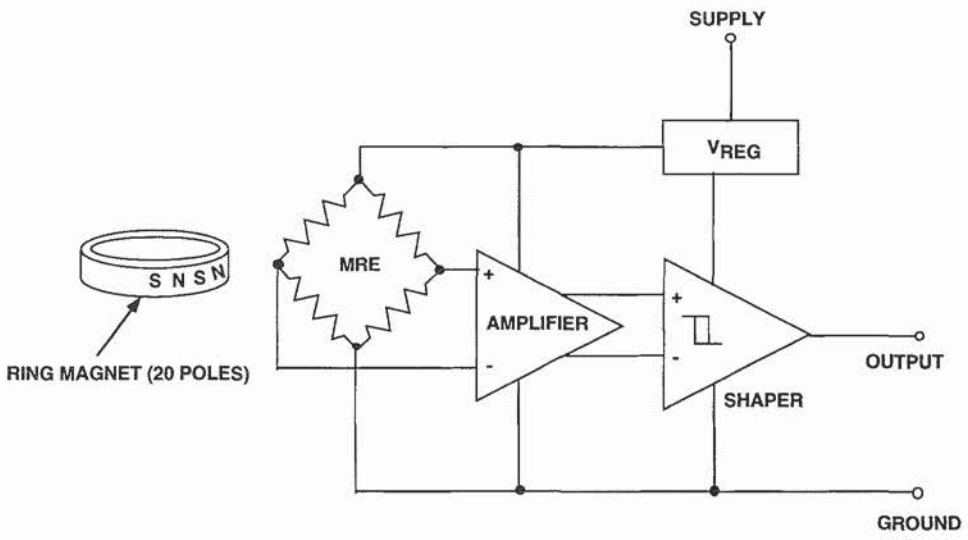


FIGURE 7.6 MRE speed-sensing module.

The resolution in vehicle speed is then:

$$\frac{P}{S} = N \times K \tag{7.3}$$

The typical system requirements are an operating temperature of –40 to 120 °C, rotational speed detection of 5 to 3000 rpm (1000 p/s), and a duty cycle ratio of 50 ± 10%.

In applications such as ABS, ASR, and four-wheel steering, additional speed sensors are attached to all four wheels so that the slip differential between the wheels can be measured. VR devices have been used and are very cost effective in this application. But the cost of other devices is dropping and as they become cost effective, they are being designed into new systems. In electronic transmission applications, information from the road and engine speed sensors, as well as torque data and throttle position are required for the MCU to select the optimum gear ratio. Electronic control can ensure smooth transition between gear ratios. Transmissions using electronic control are also smaller than conventional automatic transmissions, thus enabling more gear ratios for better performance, higher torque, efficiency, and acceleration.

Cruise control systems require information from the road and engine speed sensors to control the throttle position, and possibly the optimum selection of transmission ratios. Variable ratio assisted power steering also requires information from the wheel speed sensors for adjustment of the steering ratios for ease of turning at low speeds and good road control at high speeds. If automatic tire pressure adjustment becomes a reality, this system may also require information from the wheel speed sensors.

Another application for rotational speed sensing is to control the speed of the radiator cooling fan. The speed of the fan is determined by the coolant temperature. Hall effect devices (MRE can also be used) have been used to monitor the position of the armature and speed of the cooling fan motor. The motor controller uses this information to modulate the power to the motor through a three-phase bridge driving circuit for the control of the fan motor speed.

7.3.2 Linear Applications

Under linear applications are the detection of obstacles close to the vehicle, crash avoidance, distance of the chassis relative to the ground for ride control, measurement of ground speed for ABS, ASR, and inertial navigation. Ultrasonic devices are normally used for short distance measurements (<10 m) and RF devices for long distance measurements (see Sec. 7.6.2). For the measurement of objects from 0.5 to 2 m using ultrasonics, a pulse repetition rate of about 15 Hz is used. The reflected pulses take from 3 to 12 ms to return. The return time T is given by

$$L = C \times \frac{T}{2} \text{ (m)} \tag{7.4}$$

where L = distance to target
C = the transmission speed [given by $C = 331 + 0.6$ t (m/sec)]
T = temperature (at 15 °C), the speed of ultrasonic waves is 340 m/s.

In the case of chassis-to-ground measurements for ride control and ground speed measurements, the distance to be measured is from 15 to 50 cm and a higher pulse repetition rate can be used (up to 50 Hz). In this case, the reflected pulse takes from 0.9 to 3 ms to return. For ride control applications, an accelerometer has an advantage over distance measurement, in that it is unaffected by varying distance measurements over rough terrain.

7.4 ACCELERATION SENSING DEVICES

Acceleration sensors vary widely in their construction and operation. In applications such as crash sensors for air bag deployment, mechanical devices (simple mechanical switches) have

been developed and are in use. Mechanical switches are normally located at the point of impact in the crash zone. With the development of micromachined devices, solid state analog accelerometers have been designed for air bag applications. The analog accelerometers are centrally placed on the automotive frame. These devices can be very cost effective when compared to mechanical switches and are rapidly replacing the electromechanical devices. Silicon micromachined sensors provide a higher degree of functionality, can be programmed, have high reliability, have excellent device-to-device uniformity, and can be integrated with memory circuits to create a more accurate sensor. Additional features such as self-test and diagnostics are also available.

7.4.1 Mechanical Sensing Devices

Mechanical switches are simple make-break devices. Figure 7.7 shows the cross section of a Breed type of switch or sensor. The device contains a spring, a metal ball, and electric contacts in a tube. On impact, the inertia in the ball causes it to move against the retaining force of the spring and closes the electric contacts at the end of travel. An alternative to this device is the one shown in Fig. 7.8. It consists of a cylindrical mass wound in a flat spring. The seismic mass

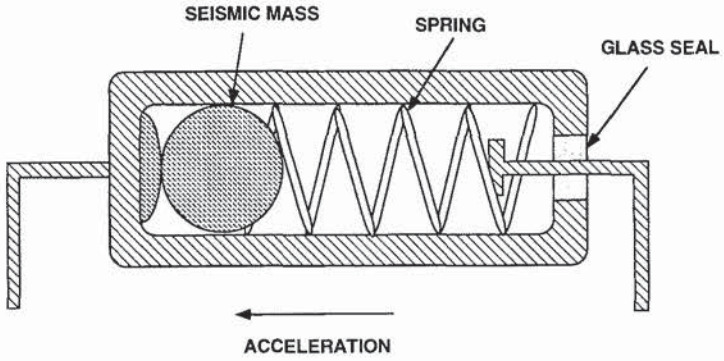


FIGURE 7.7 Cross section of mechanical sensor.

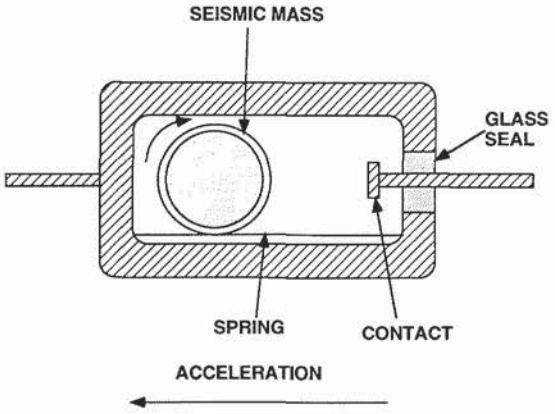


FIGURE 7.8 Cross section of mechanical switch.

rolls on impact against the spring tension, and again makes electrical contact at the end of travel. The machining tolerances on these devices are high and give wide variations in the acceleration trigger point.

7.4.2 Piezoelectric Sensing Devices

Piezoelectric devices consist of a layer of piezoelectric material (such as quartz) sandwiched between a mounting plate and a seismic mass. Electric connections are made to both sides of the piezoelectric material. The cross section of such a device is shown in Fig. 7.9. Piezoelectric material has the unique property that when a force or pressure is applied to opposite faces of the material, an electrical charge is produced. This charge can be amplified to give an output voltage that is proportional to the applied force. Piezoelectric devices can be effective in some applications, but are not suitable for sensing zero- or low-frequency acceleration, that is <5 Hz due to offset and temperature problems (pyroelectric effect). Piezoelectric sensors have a high Q, low damping, and a very high output impedance. Self-test features are also difficult to implement. The main advantages of piezoelectric devices are their wide operating temperature range (up to 300 °C), and high operating frequency (100 kHz).

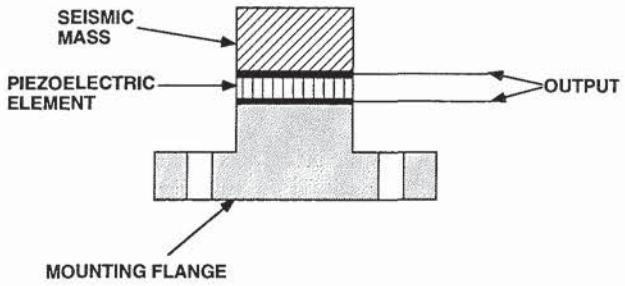


FIGURE 7.9 Cross section of a piezoelectric accelerometer.

Figure 7.10 shows a typical signal-conditioning circuit[2] and the trimming network used with piezoelectric sensors. The output from the sensor is fed to a charge amplifier, which converts the charge generated by the sensor into a voltage proportional to the charge. The circuit

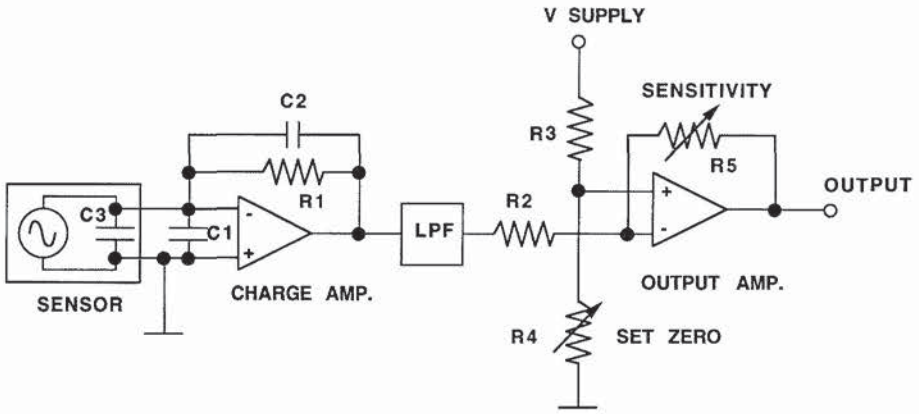


FIGURE 7.10 Piezoelectric signal-conditioning circuit.

is a modified virtual ground voltage amplifier. Feedback via capacitor C2 and resistor R1 is used to maintain the input at a virtual ground potential. This type of circuit minimizes the effect of stray or ground capacitance C1. The output voltage from the amplifier is fed via a low-pass filter (LPF) to an output amplifier, where it is trimmed for offset by R4 and sensitivity by R5. In system development, the sensitivity is set by the piezoelectric material used. Higher-sensitivity materials however, exhibit higher sensitivity to temperature variations.

7.4.3 Piezoresistive Sensing Devices

The property of some materials to change their resistivity when exposed to stress is called the piezoresistive effect. In silicon, the sensing resistors can be either P or N type doped regions, which can be very sensitive to strain. The resistors are also sensitive to temperature, so that the strain gauge is normally designed as a bridge configuration to minimize temperature effects and to obtain higher sensitivities (see also Chap. 2). In order to maintain good linearity, the operating temperature of piezoresistive devices is limited to about 100 °C. The nonlinearity is caused by excessive junction leakage current at high temperatures. Higher operating temperatures have been obtained using oxide-isolated strain gages (up to 175 °C). An uncompensated strain gauge has a typical error of 3 percent over the operating temperature range –20 to 80 °C. This error can be reduced with a compensating resistor, and still further reduced to about 0.5 percent by the use of thermistors, over an improved operating temperature range of –40 to 110 °C.

Piezoresistive sensing can be used with bulk micromachined accelerometers. Such a device is shown in Fig. 7.11. The strain-sensing elements are diffused into the suspension arms. These elements can then detect strain in the arms caused by acceleration forces on the seismic mass.

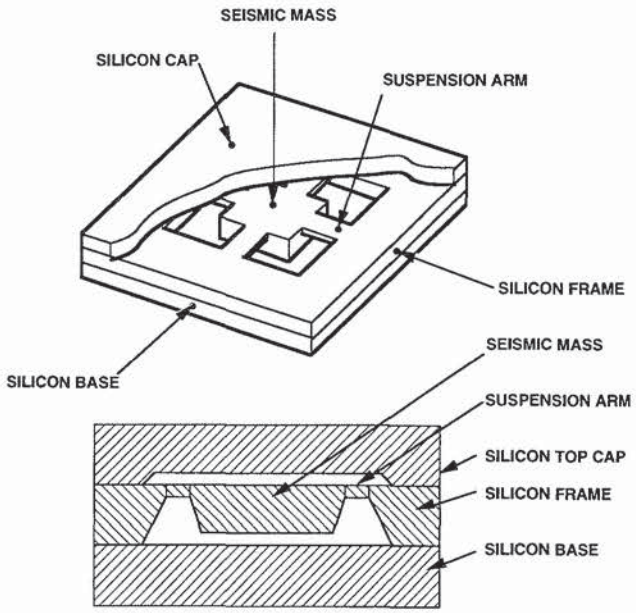


FIGURE 7.11 Bulk micromachined accelerometer.

7.4.4 Capacitive Sensing Devices

When used with micromachined structures as shown in Figs. 7.11 and 7.12, differential capacitive sensing has a number of attractive features when compared to other methods of sensing: easily implemented self-test, temperature insensitivity, and smaller size. In addition, comparing capacitive sensing to piezoelectric sensing reveals that capacitive sensing has the advantages of dc and low-frequency operation and well-controlled damping. When compared to piezoresistive sensing; differential capacitive sensing has the advantage of a wider operating temperature range and requires less complex trimming. Capacitive sensing has one other major advantage over other sensing methods in that it can be used in closed-loop servo systems. In these systems, voltages can be applied to the capacitive plates to produce electrostatic forces, which will balance the forces on the seismic mass due to acceleration. The main advantage of closed-loop operation is to make the sensor to a large extent independent of process variations. Signal-conditioning circuits can be designed to detect changes in capacitance of <0.1 fF, so that plate capacitances in the range of 200 to 400 fF can be used. The small spacing between capacitor plates in micromachining technology (2 μ) enables practical acceleration sensors as small as 500 μ × 500 μ. When designing capacitive sensors, care must be taken to ensure that the sensing voltages are properly balanced to minimize offsets due to electrostatic forces. These forces can also be produced by internal noise sources. The attributes of capacitive sensing are a linear response, operation over wide temperature ranges (–40 to 150 °C), and a frequency response from dc to about 2 kHz.

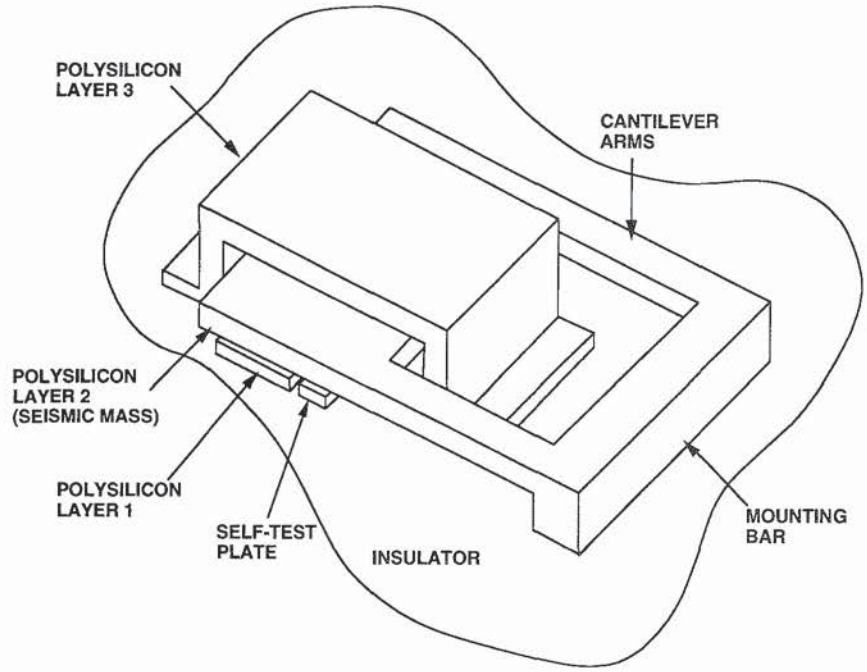


FIGURE 7.12 Surface micromachined accelerometer.

Micromachined Structures. There are a variety of types of micromachined structures that can be used in accelerometers. These structures fall into two technologies: bulk micromachined structures and surface micromachined structures. Bulk micromachined devices are structures

etched out of silicon wafers. Figure 7.11 shows the cross section of a bulk micromachined device consisting of three layers of silicon bonded together. The center layer is shaped to form a seismic mass suspended by four arms[3] (a cantilever structure has also been designed with two suspension arms[4]). When acted upon by acceleration forces, the seismic mass moves between the top and bottom plates. In this case, the movement can be sensed using piezoresistive elements diffused into the suspension arms, or differential capacitive sensing can be used between the seismic mass and the upper and lower silicon plates. The top and bottom plates can also be made of glass with metalized areas to form the top and bottom capacitors. Such devices have been designed to operate from the high g range (>1000 g), down to sensors with resolution in the μg range. Closed-loop control techniques are normally used in these lower g ranges.

The surface micromachined device shown in Fig. 7.12 is built using layers of polysilicon and sacrificial glass, which are alternately deposited and shaped. In this case, three layers of polysilicon and two layers of sacrificial glass were used. After deposition of the third polysilicon layer, the sacrificial glass is etched away leaving the freestanding structure as shown. A number of etch holes are normally placed in the second and third layer of polysilicon to speed up the etch process. These etch holes are also used to control the squeeze film damping and bandwidth of the device. The seismic mass of the second-layer polysilicon in these devices is of the order of 5×10^{-10} kg. A second plate under the middle polysilicon can be used for self-test. This function is achieved by applying a voltage to the self-test plate, which in turn will produce an electrostatic force on the center polysilicon plate causing it to deflect. This deflection will simulate an external acceleration force. An alternative to the polysilicon and glass structure is nickel with sacrificial copper.[5]

Differential capacitive sensing is used with all of these structures. Both bulk and surface micromachined devices have a very rugged construction. These devices use squeeze film damping to control bandwidth and to ensure critical damping of the resonant frequency (about 2 kHz for bulk and 10 kHz for surface micromachined devices). Film damping also ensures high resistance to shock in the sensing direction. In the directions perpendicular to the sensitive axes, the devices are rugged by construction with low cross-axis sensitivity (<3 percent). An accelerometer designed to sense a few g will typically have a shock tolerance of well over 5000 g. As already noted, surface micromachined devices that have been developed for air bag deployment have analog outputs. These devices normally operate from a 5-V supply, have a bandwidth of about 1 kHz and a sensitivity of 40 mV per g, giving a full-scale output with ±50 g input. Both open- and closed-loop techniques have been used for sensing. In comparing bulk and surface micromachined devices, the bulk structure is larger, using crystal-oriented etching with end stops, which require extra diffusions; whereas, the surface micromachined device uses isotropic etching (masking) with different materials acting as end stops. Surface structures have the potential for easier integration and use a simpler less costly process, but do require annealing.

Open-Loop Sensing. Open-loop signal-conditioning circuits amplify and convert the capacitance changes into a voltage. Such a CMOS circuit using switched capacitor techniques is shown in Fig. 7.13. The circuit contains a virtual ground amplifier to minimize the effect of stray capacitance. The positive input of the amplifier is referenced to a voltage of $V_{REF}/2$, when switch $S2$ is closed the amplifier has unity gain, and the voltage on the middle plate of the sensor is set to $V_{REF}/2$. After $S2$ is opened, $S1$ is switched so that any differences between sensor capacitors C_1 and C_2 produces a charge at the negative input of the amplifier. This charge produces a voltage (V_{out}) at the output of the integrating amplifier. The output voltage of the amplifier is given by

$$V_{out} = V_{REF} \frac{C_1 - C_2}{C_3} \tag{7.5}$$

where C_1 and C_2 are the sensor capacitances
C_3 is the integrator capacitance

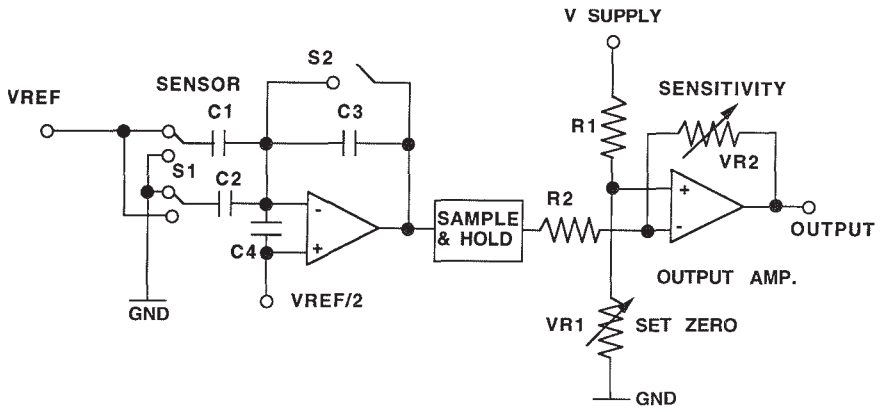


FIGURE 7.13 Capacitive sensing integrator circuit.

If the reference voltage is made proportional to the supply voltage, a ratiometric output is obtained. That is, the system gain is proportional to the supply voltage. This is a requirement in some systems to facilitate the design of the A/D converter.

A block diagram of the system is shown in Fig. 7.14. The system contains an internal oscillator, voltage reference, amplifier, sample and hold, switched capacitor filter, trim network, and output buffer. The output voltage in such a circuit is proportional to the capacitance change. This change is proportional to 1/displacement, or 1/acceleration, giving rise to some nonlinearities. However, the displacement is small compared to the spacing between the plates, so that the output voltage approximates to acceleration, giving less than 3 percent nonlinearity. The filter is used for noise reduction and to set the bandwidth for specific applications. Trimming is used to set the zero operating point and sensitivity of the system.

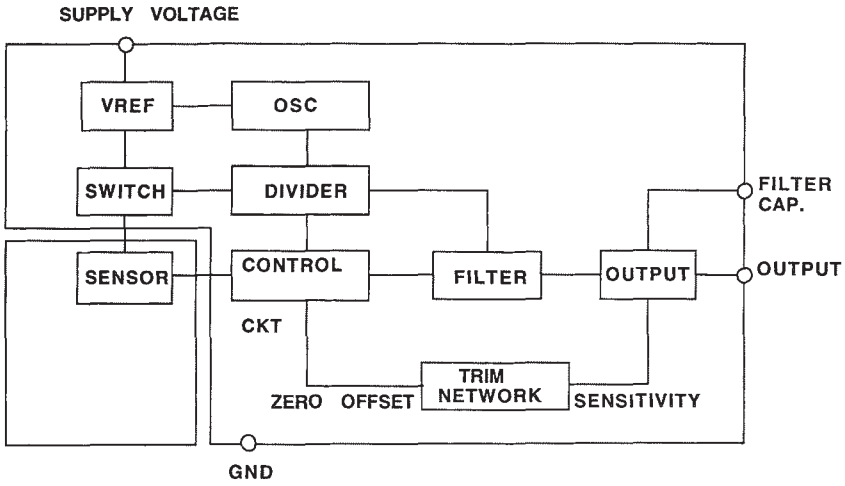


FIGURE 7.14 Signal-conditioning block diagram.

An alternate circuit with improved linearity is shown in Fig. 7.15. In this case, the output voltage is fed back to the input of the integrator, forming a bridge circuit. The feedback also sets the amplitude of the driving voltage across the sensing capacitors, so that it is proportional to their displacement. This also balances the electrostatic forces on the middle plate.

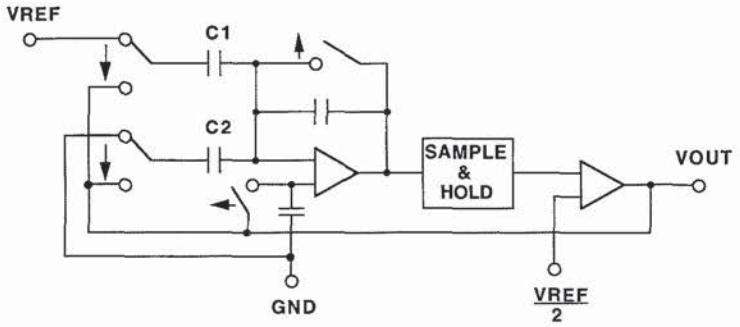


FIGURE 7.15 Linearized circuit schematic.

In this case

$$V_{out} = \frac{(C_1 - C_2)}{(C_1 + C_2)} \frac{V_{REF}}{2} \tag{7.6}$$

where $C_1 \propto 1/d_1$
$C_2 \propto 1/d_2$
$d_1 + d_2 = K$ (constant)

so that

$$V_{out} = V_{in} \frac{(d_1 - d_2)}{K} \tag{7.7}$$

showing that, in this case, V_{out} is proportional to displacement and acceleration giving improved linearity (<1 percent nonlinearity).

Closed-Loop Sensing. An alternative to the open-loop sensing circuit is the closed-loop sensing circuit, which can be configured to give an analog or digital output. Figure 7.16*a* shows the balanced electrostatic forces exerted on a seismic mass by upper and lower capacitor plates, which are at voltages of $+V$ and $-V$. If the seismic mass experiences a force due to acceleration, and a voltage δV is applied to the middle plate to generate enough electrostatic force to counterbalance the acceleration force, then the forces are as shown in Fig. 7.16*b*.

That is

$$m \times a = \frac{C(V + \delta V)^2}{2d} - \frac{C(V - \delta V)^2}{2d} \tag{7.8}$$

from which

$$m \times a = 2\,C \times V \times \frac{\delta V}{d} \tag{7.9}$$

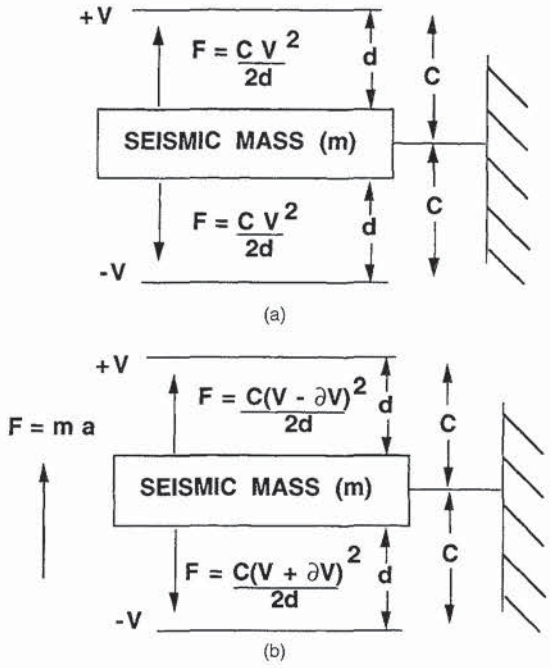


FIGURE 7.16 Electrostatic force diagram.

This shows that an acceleration force can be balanced by a linear voltage applied to the center plate. This voltage can be amplified to give a linear output voltage proportional to acceleration.[6]

In a micromachined structure, the electrostatic force produced by a δV of about 1 V can counterbalance the force produced by an acceleration of 50 g on the seismic mass. Figure 7.17 shows a block diagram of the analog closed-loop system. The top and bottom plates are dc-biased by the resistor divider network consisting of R1, R2, and R3. The ac antiphase signals used for sensing the position of the center plate are fed to the top and bottom sensor plates via the capacitors (C1, C2) from the control logic. The analog output voltage from the filter is feedback to the positive input of the integrator. When the integrator is clocked into the unity gain phase, the feedback voltage is applied to the center plate of the sensor. The electrostatic forces produced on the center plate by the differential voltage between the top and bottom plates, and the feedback voltage on the center plate, will force the center plate back to its normalized position as given in Eq. (7.9). Other methods that have been used for closed-loop operation are pulse width modulation (PWM), and delta sigma modulation (DSM).

The block diagram and waveforms of a PWM system are shown in Fig. 7.18. In this case, the center plate is held at a fixed voltage (0). The output of the integrator is amplified and converted into a PWM signal (VP), this signal and the inverted signal (VPN) are fed to the top and bottom plates of the sensor. The leading edges of the VP and VPN signals are used to sense the position of the middle plate. The electrostatic force generated by the voltage × time differential between the middle plate and the top and bottom plates, will act on the middle plate to counterbalance the forces on the plate due to acceleration. With zero acceleration, the PWM signal has a 50 percent duty cycle, so that the resulting electrostatic forces on the seismic mass are zero. The transfer function of the PWM system is given by

$$\text{Duty cycle} = \frac{\text{Acc} \times g \times d}{C \times V_{\text{REF}}^2} \tag{7.10}$$

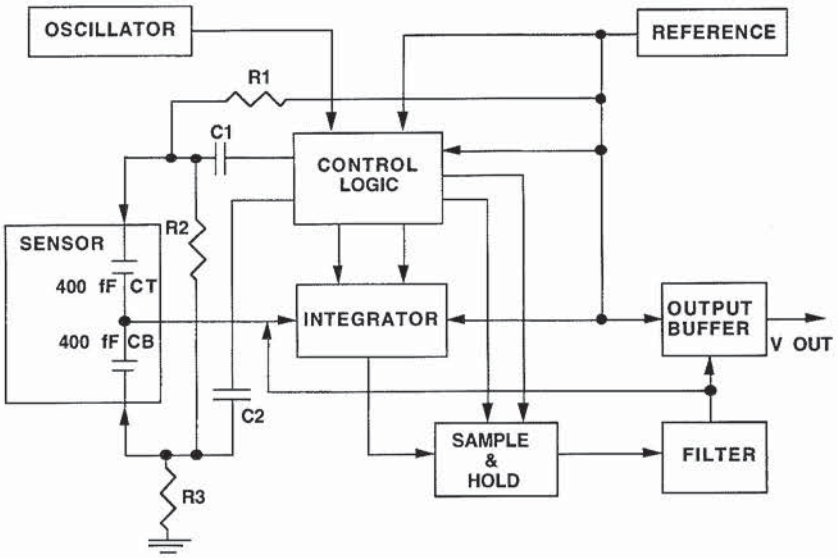


FIGURE 7.17 Block diagram of analog closed-loop system.

where g = gravitational constant
d = plate spacing
C = plate capacitance

The output can either be the PWM digital signal, or an analog output (obtained by feeding the PWM signal through a low pass filter).

Figure 7.19 shows the block diagram and waveforms of a DSM system. In this case, the middle plate is held at VF (volts). As can be seen, transient edges of the plate-drive waveforms are used for sensing the position of the center plate. The output from the integrator is fed to a comparator, whose output is then clocked into a latch where it sets up a "1" or a "0" (high or low) depending on the output from the integrator. The output from the latch is used to apply a voltage V_{REF} to the appropriate top or bottom capacitor plate, so that the electrostatic forces generated by the voltage V_{REF} – VF will maintain the center plate in its no-load position. The one-bit serial data stream from the latch can be fed directly to the MCU, or fed via a decimator circuit (which will convert the data into an 8-bit word) to the MCU. Alternatively, an LPF can be used to convert the data into an analog output. Bipolar or CMOS circuits can be used for signal conditioning. However, CMOS signal conditioning has the following advantages: a very high input impedance, good switching characteristics, low power requirements, small size, compatibility with MCU processes with the prospect of future integration, switched capacitor filters are available for noise reduction and bandwidth control, and EPROM technology is available for trimming (see Fig. 7.14). BiMOS circuits have also been used for signal conditioning.[6] In the BiMOS circuits, thin-film resistors are used to enable laser trimming of the zero offset and voltage reference; external capacitors and resistors are used for filtering and gain control.

Single vs. Multichip Control Circuits. The processing of sensors is not completely compatible with IC fabrication. Consequently, for the integration of sensors and ICs, a number of additional steps are required, which can have a detrimental effect on yields. The question then arises as to which is the most cost effective: a sensor die plus a control die with the additional cost of assembly, or a monolithic approach.

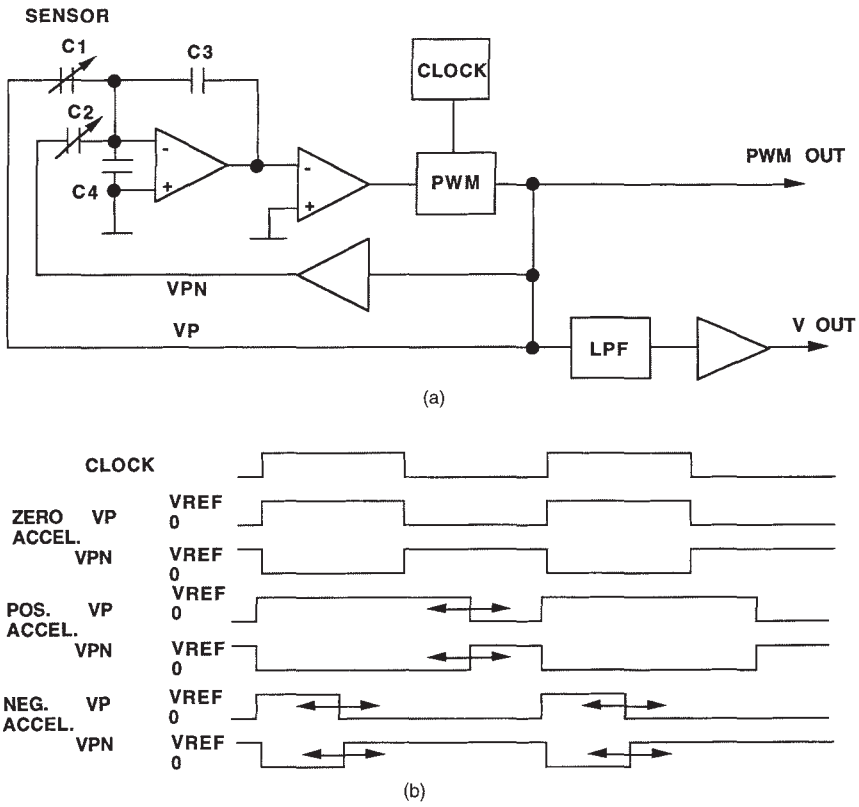


FIGURE 7.18 PWM block diagram.

In the case of micromachined devices, there are a number of advantages to using the dual-chip approach, such as flexibility, in that one type of sensor can be interfaced with a number of different types of control die, or several types of sensors can be interfaced with one type of control die. This provides a variety of input and output options. The control die and sensor can be developed simultaneously, minimizing development time. Problem solving is made easier, and the processing for both die can be optimized for performance. With the two-die approach, the sensor can also be capped and sealed during processing in a clean room atmosphere, thus eliminating contaminants and particles for good longevity. In the monolithic approach, this is not the case. In the monolithic approach, changes required to improve one section can affect the other section, which may then require additional changes in that section. The main disadvantage of the dual-chip approach is the introduction of parasitic capacitances. However, these can be addressed by existing control circuit design techniques.

7.5 AUTOMOTIVE APPLICATIONS FOR ACCELEROMETERS

Accelerometers have a wide variety of uses in the automobile. The initial application is as a crash sensor for air bag deployment. This application is normally associated with head-on collisions, but can also be applied to rear-end impact collisions to prevent rebound impact between the passengers and the windshield. An extension of this application is the use of

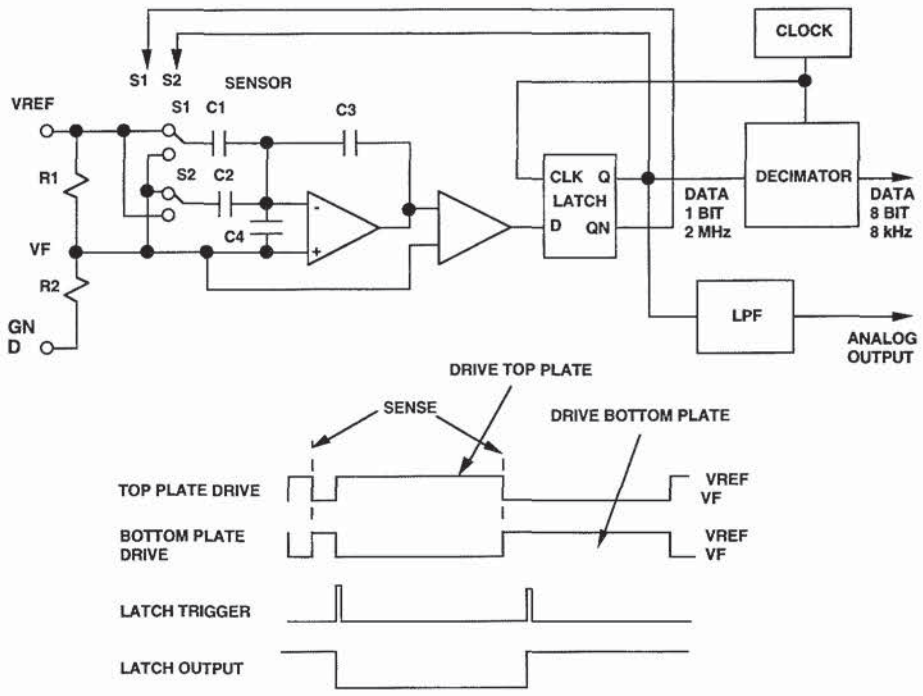


FIGURE 7.19 Block diagram of Delta Sigma Modulator.

accelerometers for the detection of side impact. This application will require additional air bags to the side of the occupants. Other low *g* linear accelerometers are being developed for ride control, ABS, traction, and inertial navigation applications.

Solid state acceleration sensors have special mounting requirements that are different from normal integrated circuits. These are to ensure that acceleration forces are transmitted to the sensor package. An advantage of the solid state device is self-test features for diagnostics. In acceleration applications, the devices are required to operate over the temperature range –40 to 85 °C (125 °C under the hood), and to withstand >2000 *g* shock. Other similar devices that have application in the automotive arena are vibration devices.

7.5.1 Air Bag Deployment Application

Crash sensors that use mechanical switches (sensors) are typically located some 40 cm from the point of impact, which necessitates the use of multiple sensors (normally 3 to 5 sensors are used in multipoint sensing) for crash sensing and air bag deployment. These devices are velocity change detectors, and are calibrated to make contact when the change of velocity in the passenger compartment is at least 20 km/h, this being the velocity change at which the front seat occupants will strike the windshield.

A centrally located analog sensor can be used as a crash sensor (single point). In the case of a centrally located accelerometer, the g level to be sensed is lower than that of a point-of-impact device. However, only one device is required to monitor the crash signature. This signature will vary with different types of chassis and different types of impacts. Consequently, an MCU is used to monitor the output of the accelerometer to determine if a crash has occurred. The typical output of a centrally located accelerometer during a 48 km/h crash is shown in Fig. 7.20. Deceleration of the vehicle and occupant displacement are also shown. At 48 km/h, the sensor has 20 ms to detect the crash and trigger the air bag. This results in infla-

tion of the air bag 50 ms after impact, at which time the occupant has moved about 18 cm or approximately halfway to the windshield and at the contact point with the inflated air bag. During the initial 20 ms, deceleration can reach 20 *g*, but the average is about 5 *g* when the air bag is triggered. The centrally located accelerometer can take one of several forms: a piezoelectric sensor, a piezoresistive device, or a capacitive sensor.

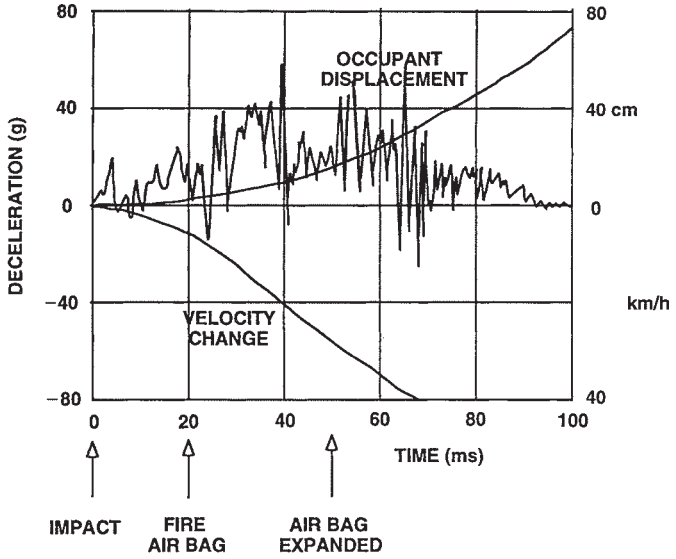


FIGURE 7.20 Typical 48 km/h crash waveform.

The centrally located accelerometer has a number of performance advantages over its mechanical counterpart. These are the reduction in the number of sensors and required bussing, which makes the centrally located system much more cost effective. There is an improvement in sensing and signal-processing accuracy with the single-point sensing accelerometer over the mechanical sensor. This gives a better-defined trigger point and overall improved performance across different chassis types. Capacitive sensors appear to have the edge in this application, because they have the potential of being cost effective, meet the requirements of the application, and have self-test features plus diagnostics available. In this application, a typical accelerometer specification is ±50 *g* full-scale output, accuracy ±5 percent over temperature, bandwidth dc to 750 Hz, and cross-axis sensitivity <3 percent. During impact, the crash sensor can also be used for seat belt locking.

7.5.2 Ride Control Application

In ride control systems, the leaf or coil springs located on the axles are replaced by four *wheel stations,* which form an active suspension. Each wheel station contains an oil-filled cylinder with a piston to set the distance of the frame above the axles and to isolate the frame from axle vibration. This is achieved using a servo feedback system. When a vehicle with conventional suspension encounters a foreign object on the highway, the load on the wheel increases as it moves up to negotiate the obstacle. This load increase makes the vehicle rise up. With a fully active suspension, the increase in load is detected and a servo valve is opened to transfer the necessary amount of oil from the appropriate cylinder to a storage container. Consequently, the load exerted on the chassis by each wheel is maintained at its specific level and

the chassis remains at its static level. After the object has been traversed, oil is pumped back into the cylinder to reestablish the static load conditions.

An alternative to the active suspension is the adaptive suspension system. In this case, information from the front wheels is gathered and used to predict road conditions for the control of the rear wheels. The advantage over the fully active suspension is one of cost, as the number of acceleration sensors is halved. During cornering, oil is also pumped into the outside wheel cylinders to minimize roll angle.[7]

A combination of sensors is used for active suspension. These are accelerometers, wheel speed sensors, chassis-to-ground sensing, and piston-level sensing in the suspension system. The low g accelerometers used on the axles of the four wheels to detect the load changes on the wheels have the following specifications: ±2 *g* full-scale, accuracy ±5 percent over temperature, bandwidth dc to 10 Hz, and cross-axis sensitivity <3 percent. The acceleration information and data from the wheel speed sensors is used to provide the information necessary for the MCU to operate the servo control valves. Hall effect, MRE, and opto sensors have been used for monitoring the level of the pistons in the wheel stations cylinders.

7.5.3 Vibration Applications

Lean-burn engines are being developed for improved emission levels and for better fuel economy (10 to 15 percent improvement). NO_x emissions are greatly reduced to meet federal standards. Lean-burn engines use high stoichiometric ratios; 20:1 and higher are necessary. At these ratios, combustion becomes unstable and torque fluctuations large. Consequently, antiknock and vibration sensors are required to supply the information necessary to the MCU, so that it can adjust the injected fuel amount and ignition timing for stability over widely varying conditions.

There are two types of solid state sensors that can be used in this application: piezoelectric devices and capacitively coupled vibration sensors. A typical vibration sensor contains a number of fingers of varying length which vibrate at their resonant frequencies when those frequencies are encountered. The resonance is capacitively coupled to the sensing circuit, and the outputs as shown are obtained. Optical sensors have also been used as antiknock sensors. In this case, the ignition spectrum is monitored for the detection of misfiring or knocking. Vibration sensors can also be used for vibration monitoring in maintenance applications.

7.5.4 Antilock Brake System Applications

In antilock brake systems, speed sensors are attached to all wheels to determine wheel rotation speed and slip differential between wheels. VR devices, as well as Hall effect and MRE devices, can be used in this application, as zero speed sensing is not required. VR devices have been used and shown to be cost effective in this application, but Hall effect and MRE devices are now being designed into these systems. Pressure sensors are used to monitor brake fluid pressure, and an accelerometer or ground-speed sensor can be used to provide information on changes in the vehicular speed. Brake pedal position and brake fluid pressure information are also required for control. All of this information is fed to an MCU, which processes the data and adjusts the brake fluid pressure to each wheel for optimum braking. Many of the elements of the ABS system can be used for the detection of lateral slippage on high-speed cornering, and can be used for traction and the direction of power to the wheels. Traction control applies in particular to slippery surfaces and with four-wheel-drive vehicles. Additional information over that used in ABS systems is required by the MCU for ASR applications, such as engine speed and throttle angle. In this application, servo feedback to the throttle may also be necessary.

A more cost-effective, but less accurate, system for ABS and ASR is the adaptive control system in which accelerometers are normally used to measure deceleration when braking, and

acceleration when the throttle is opened. If skidding occurs during braking, the brake pressure is reduced and adjusted for maximum deceleration, or the throttle adjusted for maximum traction. Typical specifications for the accelerometer required in this application are: ±1 g full-scale output, accuracy ±5 percent over temperature, bandwidth 0.5 to 50 Hz, and cross-axis sensitivity <3 percent.

7.6 NEW SENSING DEVICES

New cost-effective sensors are continually being developed. The technology and cost are often pushed by the application and volume requirements of the automotive industry and federal mandates. Today's silicon sensors and control electronics are limited in operating temperature to 150 °C to ensure long life of the devices. This operational temperature is adequate for most applications, but higher temperature operation may be required for sensors mounted in the engine compartment. The limit on the operating temperature of silicon devices can be extended to between 200 and 250 °C by the use of special isolation techniques such as dielectric isolation (this operating temperature applies to surface micromachined devices). For higher-temperature operation, alternative materials such as GaAs or SiC are being developed, but the cost of these devices limits their use at present.

A list of semiconductor conductor materials and maximum practical operating temperatures is given in Table 7.3. Higher operating temperatures have been reported but with poor longevity.

TABLE 7.3 Device Operating Temperatures

Material	Maximum practical operating temperature, °C
Si	150
Si (dielectric iso.)	250
GaAs	300
AlGaAs	350
GaP	400
SiC	500

7.6.1 New Rotational Speed-Sensing Devices

A number of new devices are being investigated to detect magnetic fields. These are flux gate, Weigand effect, magnetic transistor, and magnetic diode. The magnetic transistor at present is showing the most promise. The device operates on a similar principle to the Hall effect device. That is, the current division between split collectors (bipolar) or split drains (MOS) can be changed by a magnetic field. This current differential can then be detected and amplified to give an output voltage proportional to magnetic field strength. These devices can use either majority or minority carriers, and can be either vertical or lateral bipolar or MOS devices. The magnetic transistor has the potential of higher sensitivity than the Hall effect device.

Figure 7.21 shows the cross section of a lateral PNP magnetic transistor. The current from each collector is equal until a magnetic field is applied perpendicular to the surface of the device. The magnetic field causes an imbalance of current between the two collectors. Sensitivities with this type of structure have been reported as being an order of magnitude greater than in the Hall effect device.[8] Magnetic transistors and diodes can be directly integrated with the signal-conditioning circuits, which could make them very cost effective in future applications. A comparison of the magnetic transistor to other practical devices is given in Table 7.2.

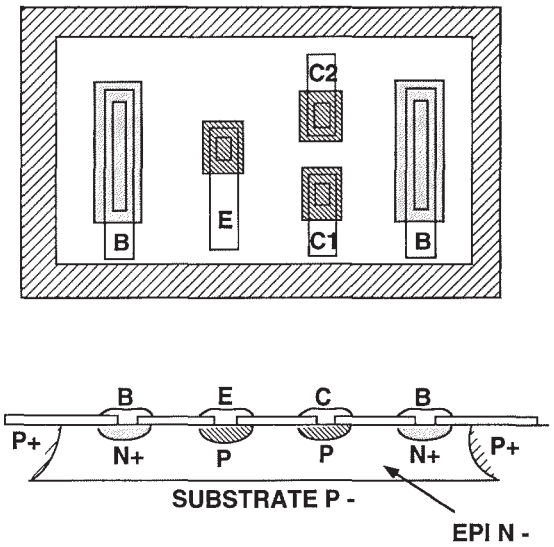


FIGURE 7.21 Cross section of a field-assisted PNP magnetic transistor.

7.6.2 New Linear Speed-Sensing Devices

A number of different sensing technologies can be used for distance, object detection, and approach speed measurements. Shown in Fig. 7.22 are the areas covered by blind-spot, rear, and forward-looking sensors. Ultrasonics, infrared, laser, and microwaves (radar) can be used in the detection of objects behind vehicles and in the blind areas. From a practical standpoint, no technology has come to the forefront. However, with new innovations in technology the situation may change very rapidly. Ultrasonics and infrared sensors are cost effective but degrade with inclement weather conditions such as ice, rain, snow, and the accumulation of road grime. Infrared devices are also color-sensitive, in that the sensitivity to shiny black objects is very low compared to other colors. Microwave devices appear to have the edge

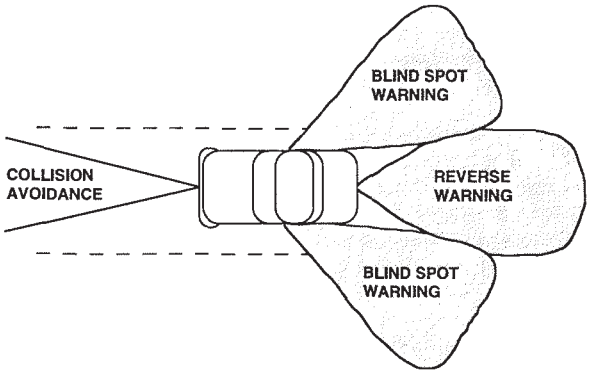


FIGURE 7.22 Collision-avoidance patterns.

when considering environmental conditions,[9] but are expensive, and radar can be affected by false return signals and clutter.

For the detection of obstacles and vehicles in front of a vehicle, the choice is between laser and microwaves due to the distances involved (up to 90 m). Microwaves have the disadvantage of high cost and large antenna size when considering available devices in the 60 GHz range. Frequencies greater than 100 GHz are preferred for acceptable antenna size. However, collision avoidance radar in the 76/77-GHz band has been developed in Europe. Collision avoidance radar in the 77, 94, and 144 GHz is being considered in the United States. A typical system uses a 38.5-GHz VCO (voltage controlled oscillator) with frequency-doubling to obtain about 40 mW of power at 77 GHz. A frequency-modulated continuous wave or pulse-modulated system can be used. The system uses GaAs devices to meet the frequency and power requirements. Lasers can be cost effective in this application, but also have their drawbacks: degradation of performance by fog, reflections from other light sources (sun, etc.), build-up of road grime on sensor surfaces, and poor reflecting surfaces at laser frequencies, such as grimy and shiny black surfaces.

7.6.3 New Inertial and Acceleration-Sensing Devices

Recent developments in solid state technology have made possible very small cost-effective devices to sense angular rotation. The implementation of one such gyroscopic device is shown in Fig. 7.23. This device is fabricated on a silicon substrate using surface micromachining techniques. In this case, three layers of polysilicon are used, with the first and third layers being fixed and the second layer free to vibrate about its center. The center is held in position by four spring arms attached to four mounting posts as shown. This device can sense rotation about the X and Y axes and sense acceleration in the direction of the Z axis. The center layer of polysilicon, driven by electrostatic forces, vibrates about the Z axis. These forces are produced by voltages applied between the fixed comb fingers and the comb fingers of the second polysilicon. Capacitor plates as shown are formed between the first and third polysilicon on the X and Y axes, and the second layer of polysilicon. Differential capacitive sensing techniques can then be used to sense any displacement of the vibrating disc caused by angular rotation. For example, if angular rotation takes place about the X axis, Coriolis forces produce a deflection of the disc about the Y axis. This deflection can be detected by the capacitor plates on the X axis. The sensing of the three functions is achieved by using a common sensing circuit that alternatively senses the X rotation, Y rotation, and acceleration. The gyroscope is designed to have a resolution of <10 degrees per hour for angular rate measurements, and an acceleration resolution of 0.5 mg.

7.7 FUTURE APPLICATIONS

New applications to increase creature comfort and safety are constantly being developed, but their rate of introduction will depend on the cost effectiveness of the technology, demand, and government mandates. Other concerns of automotive manufacturers are size, weight, power requirements, and adverse effects on styling and appearance. Many of the new sensor technologies are in their infancy, and thus are not yet cost effective on medium- and low-priced automobiles, but are being made available as options on luxury cars.

7.7.1 Future Rotational Speed-Sensing Applications

A future application for speed-sensing devices will be in continuously variable transmissions. In this application, engine and wheel speed, as well as torque, will be measured, and the infor-

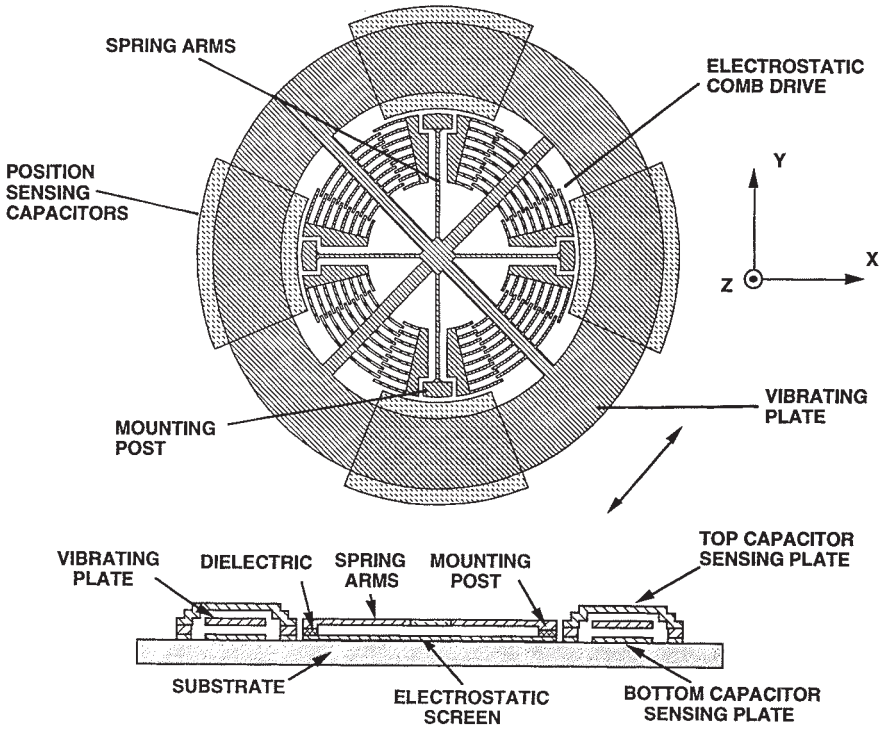


FIGURE 7.23 Solid state gyroscope.

mation processed by an MCU to optimize transmission ratios for engine performance and efficiency. All-wheel steering is also under development, and requires speed-sensing information, in addition to steering, front-, and rear-wheel angle position data for processing and control.

7.7.2 Future Linear Speed-Sensing Applications

Another application that has been developed is the use of speed- and distance-measuring devices for collision avoidance. These devices fall into three categories: near-obstacle detection (rear), blind-spot detection, and semiautomatic frontal object detection and control[9] (see Fig. 7.22).

Near-obstacle detection is used to prevent accidents during reversing. Blind-spot detection is used to prevent accidents due to careless lane changing, and when backing out of a driveway, garage, or alley into traffic. The semiautomatic frontal detection is a long-range system. The distance and closing speed between vehicles, or between a vehicle and a fixed object, can be measured and the speed adjusted as necessary to avoid a collision,[10] or the driver can be warned of impending danger. An addition to this is to monitor road surface conditions for friction—for example, dry roads compared to wet or icy roads—and also to use this information to adjust approach speeds and distance. Without collision avoidance, road condition monitoring can be used to caution vehicle operators. Collision avoidance systems can be used to minimize collisions, or can be used to operate protection systems before an unavoidable collision happens to protect the automobile passengers. In this case, vehicles closing at or approaching an obstacle at 80 km/h will be less than 7 m apart before a collision-is-imminent

determination can be made. This gives 200 ms decision-making time for the system MCU. This is, however, long compared to today's air bag deployment systems, which have 20 ms decision-making time after the event. In the future, an idealistic system may be a combination of the two systems.

7.7.3 Future Acceleration Applications

One of the future applications being considered is the expansion of the air bag system to include side impact protection. The sensor used for crash sensing is unidirectional, so that it can only detect forward impact. A similar sensor, mounted perpendicular to the air bag sensor, can be added to the system to detect side impact and to deploy protection for the passengers. This device will typically require a 250-*g* accelerometer. Another application for accelerometers is to detect slippage during cornering in advanced steering systems. These systems will employ a low-*g* accelerometer (1–2 *g*).

7.7.4 Inertial Navigation Applications

A number of inertial navigation systems are being developed for short- and long-range travel. Long-range inertial navigation systems normally obtain their location by using a triangulation method. This method references three navigation satellites with known locations in fixed orbits. However, there are certain conditions under which contact with all three satellites is lost. This occurs when the vehicle is in the shadow of tall buildings or high hills, and triangulation is not possible. Under these conditions, the guidance system has to rely on such devices as gyroscopes, which sense angular rotation or change in direction, and/or monitor vehicular motion relative to the road.

Short-range inertial navigation systems or inertial measurement units (IMU) rely to a large extent on high-accuracy accelerometers and gyroscopes. A typical accelerometer specification for this application is: ±2-*g* full-scale output, accuracy 0.5 percent over temperature, bandwidth dc to 20 Hz, and a cross-axis sensitivity <0.5 percent. A centrally located IMU can be expanded to cover other applications such as suspension, ABS, ASR, and working with crash avoidance sensors. This may be the way to handle cost-effective system design in the future. The IMU can also be designed to provide location data for intelligent vehicle highway systems. These systems (Prometheus,[11] Amtics[12]) improve travel efficiency and reduce fuel consumption and pollution by selecting the optimum route to a given destination. The route is chosen to avoid traffic congestion, road construction, and accidents (see Chap. 29).

7.8 SUMMARY

In this chapter, a number of speed-sensing devices, both rotary and linear, have been described, together with potential applications. VR, MRE, Hall effect, and opto devices (possibly magnetic transistor in future applications) can be used in rotational applications for engine control, transmission, and wheel speed sensing. Of these devices, Hall effect, VR, and opto have been widely used. With the tendency for direct pickoff, optical devices may become impractical. MRE devices are being designed in and will become a serious contender to the Hall effect device. In linear applications for crash avoidance, microwave devices have the edge over performance and optical devices in terms of cost. However, as the cost of microwave devices declines, they could become cost effective. For blind-area alert and reversing obstacle detection, ultrasonics and infrared devices are cost effective, but performance degrades during inclement weather.

The accelerometer has possibly the greatest potential for applications in the automobile. These applications range from crash sensing, ride control, ABS, and ASR to IMU systems. Accelerometers needed will range in sensitivities from 50 *g* in crash sensing to 1 *g* in the IMU. Advances in technology are providing a number of new sensors that are showing potential, such as the magneto transistor and the micromachined gyroscope. To summarize, Figs. 7.24 and 7.25 show the types of sensors used in specific applications, and the technologies used for specific sensors. As can be seen, one type of sensor can be used in a number of applications. In applications where a sensor output is shared, care must be taken to ensure that a failure in one system does not disable the sensor or other systems. Because of the similarities in several of the systems and the use of shared sensors, the greatest potential for cost-effective system design is a single control system. The IMU shows great potential to be the controller for ride control, ABS, ASR, four-wheel-drive, and steering applications. The rate of introduction of new sensors and systems will depend on federal mandates, customer demand, and the need to improve engine fuel efficiency and to reduce emissions.

GLOSSARY

Adaptive suspension A suspension system that monitors motion of the front wheels and adjusts the suspension of the rear axle accordingly.

Arntics Acronym for Advanced Mobile Traffic Information and Communication System.

ASR (traction) A system to prevent wheel spin on slippery surfaces, to give maximum traction and acceleration.

Sensor \ Application	Air Bag Deployment	Ride Control	ABS	Engine Control	Transmission	4 Wheel Drive	All Wheel Stearing	Engine Vibrarion	Cruise Control	Seat Belts	Power Steering Assist	Traction (ASR)	Collision Avoidance	Obstacle Detection	Inertial Navigation
Speed Rotation	▨	▨	▨	▨	▨	▨	▨		▨		▨	▨			
Speed Linear			□		□				□		□	□	□	□	□
Acceleration	▨	▨	▨	▨						▨		▨			▨
Vibration								▨							
Angular Rotation															▨

▨ USED □ MAY BE USED

FIGURE 7.24 Sensor applications.

Measurand / Sensing Technique	Mechanical	Piezoelectric	Piezoresistive	Capacitive	Optical	Infrared	Radar	Laser	Ultrasonic	Hall Effect	Variable Reluctance	Magnetoresistive
Speed Rotation					▨					▨	▨	▨
Speed Linear						▨	▨	▨	▨			
Acceleration	▨	▨	▨	▨								
Vibration		▨		▨								
Angular Rotation				▨								

FIGURE 7.25 Sensor technologies.

Coriolis forces Forces exerted by a spinning body to oppose any motion at right angles to the axle.

Crash signature Shock waveform projected through a chassis during a collision.

Inertial navigation Guidance system giving accurate location.

IMU (inertial measurement unit) System used for guidance between two locations indicates road hazards and delays.

Lean burn Engine with high compression ratios and high air-to-fuel ratios for increased efficiency and low emissions.

Micromachining Manufacturing technology for micromechanical structures using chemical etching techniques.

NO_x Chemical symbol for oxides of nitrogen.

Prometheus Acronym for PROgraM for a European Traffic with Highest Efficiency and Unprecedented Safety.

Switched capacitor filter Technique for switching capacitors to simulate high-value resistors for low-frequency filters to minimize size.

REFERENCES

1. Osamu Ina, Yoshimi Yoshino, and Makio Lida, "Recent Intelligent Sensor Technology in Japan," S.A.E. paper 891709, 1989.
2. D. E. Bergfried, Mattes, B., and Rutz, R., "Electronic Crash Sensors for Restraint Systems," *Proceedings of the International Congress on Transportation Electronics,* Detroit, Oct. 1990, pp. 169–177.

3. E. Peeters, Vergote, S., Puers, B., and Sansen, W., "A Highly Symmetrical Capacitive Microaccelerometer with Single Degree of Freedom Response," *Tranducers 91,* 1991, pp. 97–103.
4. J. T. Suminto, "A Simple High Performance Piezoresistive Accelerometer," *Transducers 91,* 1991, pp. 104–107.
5. J. C. Cole, "A New Capacitive Technology for Low-Cost Accelerometer Applications," *Sensors Expo. International,* 1989.
6. T. A. Core, Tsang, W. K., and Sherman, S. J., "Fabrication Technology for an Integrated Surface-Micromachined Sensor," *Solid State Technology,* Oct. 1993, pp. 39–47.
7. H. Wallentowitz, "Scope for the Integration of Powertrain and Chassis Control Systems: Traction Control—All Wheel Drive—Active Suspension," *Proceedings of the International Conference on Transportation Electronics,* S.A.E. paper 901168, 1990.
8. H. Kaneko, Muro, H., and French, P. J., "Optimization of Bipolar Magneto-Transistors," *Micro Systems Technologies 90,* 1990, pp. 599–604.
9. L. Raffaelli, Stewart, E., Borelli, J., and Quimby, R., "Monolithic Components for 77 GHz Automotive Collision Avoidance Radars," *Proceedings Sensors Expo,* 1993, pp. 261–268.
10. S. Aono, "Electronic Applications for Enhancing Automotive Safety," *Proceedings of the International Conference on Transportation Electronics,* SAE 901137, 1990, pp. 179–186.
11. J. Hellaker, "Prometheus-Strategy," *Proceedings of the International Congress Transportation Electronics,* SAE 901139, 1990, pp. 195–200.
12. H. Okamoto, and Hase, M., "The Progress of Amtics-Advanced Mobile Traffic Information and Communication System," *Proceedings of the International Congress Transportation Electronics,* SAE 901142, 1990, pp. 217–224.

ABOUT THE AUTHOR

William C. Dunn is a member of the technical staff in the Advanced Custom Technologies group at Motorola's Semiconductor Product Sector in Phoenix, Arizona. He has over 30 years' experience in circuit design and systems engineering. For the past 15 years he has been involved in the development of semiconductor sensors, smart power devices, and control systems for the automotive market. Prior to joining Motorola, he worked for several large corporations in the United Kingdom and United States, has written over 30 papers, and has over 30 patents issued and pending on mechanical sensor structures, semiconductor technology, and circuit design.

CHAPTER 8
ENGINE KNOCK SENSORS

William G. Wolber
Executive Engineer
Cummins Electronics Co., Inc.

8.1 INTRODUCTION

Knock is a phenomenon characterized by undesirable structural vibration and noise generation and is peculiar to spark-ignited engines. Knock is undesirable both from a customer acceptance standpoint and also because severe knock can cause engine damage. The terms *ping* (light, barely observable knock) and *predetonation* (knock caused by ignition of the charge slightly before the full ignition of the flame front by the spark plug) are also commonly used in the industry, and not always with precision.

Historically, knock became an important engine phenomenon in the 1930s. In their search for performance improvements, designers of spark-ignited engines increased compression ratios beyond the capabilities of the gasoline formulations available to provide smooth, knock-free combustion. The discovery of gasoline additives such as tetraethyl lead, which improve the stability of the combustion process primarily by decreasing burn rate, permitted the fuel suppliers to provide a range of fuel knock properties at the pump to match with the range of engine requirements. As a result, by 1950, knock was not regarded as a significant engine performance limiting parameter since the car company could specify what fuel should be selected, ranging from "regular" to "super premium" grades. A blending fuel pump was even provided by one fuel company which allowed the operator of the vehicle to select exactly what grade of fuel he or she wished to use from a selection of six grades.

In the late 1960s, the situation began to change drastically. In response to public concern about air quality and health, legislation on automotive emissions became a new constraint on automotive engine design, first in the United States and more recently worldwide. The use of fuel additives to improve gasoline knock characteristics fell into disfavor for two reasons:

- Many of the better additives resulted in combustion products not desirable in the air. For example, most lead compounds are toxic at low levels to human and animal life.
- As engine designers strove to provide lower engine emissions, the use of the three-way catalytic converter cleanup device in the exhaust system became nearly universal. Unfortunately, most good fuel additives that improve knock characteristics poison the catalyst of the converter, rendering it ineffective. Also, many engine controls began to make use of the oxygen sensor feedback control loop. The catalyst on the oxygen sensor also is poisoned by the better additives.

This situation led to gradually more stringent federal regulation of the fuel industry, a trend which, in the end, legislated fuels with knock-improving additives completely off the

market, so that by the mid 1980s, all spark-ignited automobile engines in the United States were being operated using lead-free gasoline. Of course, as leaded gasoline disappeared, the knock phenomenon returned. The auto industry knew what to do—lower compression ratios—but this step alone has accompanying negative consequences, such as lower fuel economy. Thus, the situation was one driven by the need to optimize between several principal goals: emission regulations, fuel supply characteristics, fuel economy goals, customer acceptability, and vehicle performance.

8.2 THE KNOCK PHENOMENON

8.2.1 Definition

Spark-ignited engine knock has been defined as "an undesirable mode of combustion that originates spontaneously and sporadically in the engine, producing sharp pressure pulses associated with a vibratory movement of the charge and the characteristic sound from which the phenomenon derives its name."[1] This definition at once conveys some of the difficulties of measuring the phenomenon and devising engine controls to minimize it. First, the ultimate standard is the reaction of the human to the vibration felt or the sound heard as a result of the knock. An attempt to measure the cause of the phenomenon leads one to the difficult problem of observing pressure waves in the cylinder. In fact, over the years these difficulties led the industry to devise an experimental comparison measurement technique which measured the octane rating of the *fuel,* not of the engine.

8.2.2 Laboratory Measurements

The CFR Engine. The basic comparison method which evolved for rating fuel knock quality involved the use of a simple, single-cylinder, spark-ignited engine called the CFR engine, and a pure hydrocarbon fuel—100 percent isooctane. Researchers discovered that they could reproduce knock phenomenon, laboratory to laboratory, by running this engine under specified conditions with this fuel.

Gasoline Octane Rating. A system of rating the knock characteristics of fuels, called the octane rating, was developed based on a comparison of the CFR engine performance with the test fuel compared to pure isooctane. On this scale, osooctane has an octane rating of 100; passenger car engine fuels ranged from the high 70s for "regular" to the high 90s for "super premium." (Fuel formulations with an octane rating over 100 were possible but were used mainly for aircraft engines.)

As the industry strove to put knock on a solid engineering basis, methods of measuring the resulting phenomena on the CFR engine, and commercial engines as well, were developed. The parameter on which the industry standardized was "jerk," the third-time derivative of engine block displacement. Scales evolved for comparing the relative knock performance of fuels and of engines based on the output of this kind of sensor. Much work was also done to correlate jerk sensor measurements with cylinder pressure phenomena.

The fundamental resonant frequency of the knock-generated pressure signal is dependent on engine cylinder geometry and the speed of sound in the charge gas. The structural vibration characteristics of the engine block that are excited by the fundamental knock event are determined by the engine block transfer function. Testing conducted on one cylinder of a six-cylinder spark-ignited engine demonstrates that high-frequency structural vibration components are good indicators of knock, despite the relatively low frequency high-pressure excitation event.[2] Structural vibration induced by mechanical events, such as valves opening

and closing, introduces noise that can be confused with knock-induced vibrations. Careful signal analysis is required to overcome a sometimes poor signal-to-noise ratio.

The reverberation resonance of the cylinder typically lies in the range between 2 and 12 kHz. A useful rough estimate of the knock frequency for a given engine cylinder geometry is given by Draper's equation[3]:

$$\text{fr} = \frac{\text{Pmn} \times C}{\pi B} \tag{8.1}$$

where fr is the knock resonant frequency
Pmn is a vibration mode constant
C is the velocity of sound in the gas in the cylinder
and B is the radius of the cylinder

Using this equation, assuming that the average gas temperature is 2000 K so that C is 900 m/s, the first circumferential mode resonant frequency is estimated at 5.75 kHz for a cylinder of 10-cm bore.

As better cylinder pressure sensors have become available, cylinder pressure wave measurements, along with signal analysis techniques for deriving a knock rating number, have supplemented jerk measurements in the laboratory. The most popular signal analysis approach is to declare the maximum amplitude or peak-to-peak value of the bandpass filtered pressure data as a knock number, usually referred to as *knock intensity.*

Other ways of describing knock level are the root-mean square, mean square, or integral of the absolute value of the bandpass pressure oscillations during the knocking portions of the cycle. The spectral power of the pressure transformed into the frequency domain has been used for measurement. Derivative-based methods have also been established, based on rapid changes of cylinder pressure during the knock phenomenon, using the first, second, and third derivatives of the cylinder pressure history.

8.2.3 On-Board Knock Control

Improved Fuel Economy by Lowering Engine Compression Ratio. As engine control using microcomputers became established in the late 1970s and early 1980s for gasoline passenger car engines, and the octane rating of the fuels available for them dropped below 90, the automotive industry responded by lowering engine compression ratios. However, some of these smaller engines were equipped with turbochargers to recover some of the lost performance at wide open throttle. It was found that these engines could experience destructive high-speed knock at maximum speed and power. Not only were the knock reverberations in the cylinder large, but they were modulating a high base pressure wave. This could occur because the extra boost pressure from the turbocharger permits additional fuel to be metered to the engine without violating emissions restrictions on the control, since the amount of oxygen available to react with the fuel is greater. This could result in knock wave peaks exceeding the cylinder head pressure limit, as well as excessive vibration of the entire engine.

The knock occurring at wide open throttle had to be decreased, but retaining the additional power from the turbocharger at part throttle was very desirable. The electronic control was available to modulate the amount of fuel injected during the dangerous engine state; however, a means of sensing engine knock on-board in real time was needed. The sensors used are described in Sec. 8.3. The control operates by sensing that knock is over the permissible limit, and then reduces it either by retarding ignition timing or by opening a wastegate valve on the boosted manifold pressure.

Improved Fuel Economy by Raising Compression Ratio. Following the successful implementation of knock limit control on turbocharged engines, considerable experimentation was carried out to assess whether or not there is value in applying knock limit control to naturally aspirated spark-ignited engines. It was determined that if the engine knock was always kept

less than an empirically determined limit for that type of engine and stored in the microcomputer control memory, then the knock level would not be objectionable even to the more critical car operators. Moreover, this could be accomplished while raising the compression ratio one full number—for example, from 8:1 to 9:1. If used to improve fuel economy, this change results in a 3 percent improvement in the corporate average fleet economy (CAFE) rating for that model/engine combination. At today's CAFE values, this is close to a one-mile-per-gallon improvement.

CAFE values are extremely important to a passenger car original equipment manufacturer (OEM). The larger, heavier luxury cars cannot achieve as low a CAFE value as the smaller, lighter cars. However, the larger cars are more profitable to produce. Thus, the mix of cars produced by an OEM can depend on just how close his CAFE results are coming to the mandated value. A way of improving fuel economy without downsizing, adding significant cost, or losing performance is very valuable.

Starting in the early 1980s, more and more passenger cars had a knock limit control and increased compression ratio. Such a change is usually scheduled at a time when the engine design and the tooling to make the engine are changing for other reasons. By 1990, 25 to 30 percent of the naturally aspirated passenger car engines produced in the United States (and all of the turbocharged engines) featured knock limit control.

8.2.4 Measurement and Control System Considerations

In addition to the problem of developing a suitable on-board sensor for knock, once the parameter to use had been selected, a number of other considerations had to be settled in effecting a satisfactory control. It was already pointed out that engine knock can be reduced either by retarding spark timing or by opening a manifold boost wastegate valve. In the United States, the candidate engines already had electronic control of spark timing, so that all that was needed to implement the knock control was an easily added change in the timing command in the control microcomputer. However, many control strategies exist for processing the knock signal. Perhaps the most elaborate of these was implemented on the first turbocharged engine knock limit control:

- The vibration frequency of the knock is specific to the engine model but lies between 2 and 12 kHz for passenger car engines. The sensor used was mechanically bandpass-tuned to match this characteristic with a Q of about 2.
- The major knock reverberations for a given cylinder occur during a time window that starts shortly after the cylinder reaches top dead center and ends 60 to 90 crank-angle degrees later. The control opened a signal gate to allow the knock signal to pass through and be averaged only when the engine was in these time windows.
- To prevent engine damage, whenever the knock signature exceeded the limit value, the control very rapidly retarded ignition by as much as 10 crank-angle degrees so that the engine would not be in severe knock even for the next few cylinder events. Then the control would very slowly advance the timing until the process repeated. This resulted in somewhat less than best engine performance but assured a comfortable margin of safety for the engine.
- The knock threshold limit in the control was modulated to increase with engine speed so as to compensate for greater noise background at high engine speed.

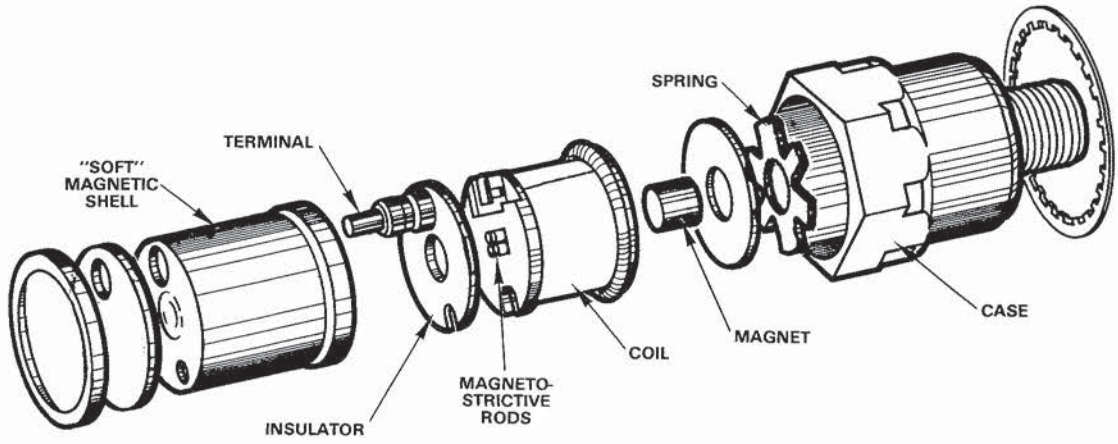
8.3 TECHNOLOGIES FOR SENSING KNOCK

A number of different parameters have been selected for measuring knock in real time on board, and sensors have been developed for this purpose. It must be recognized that these

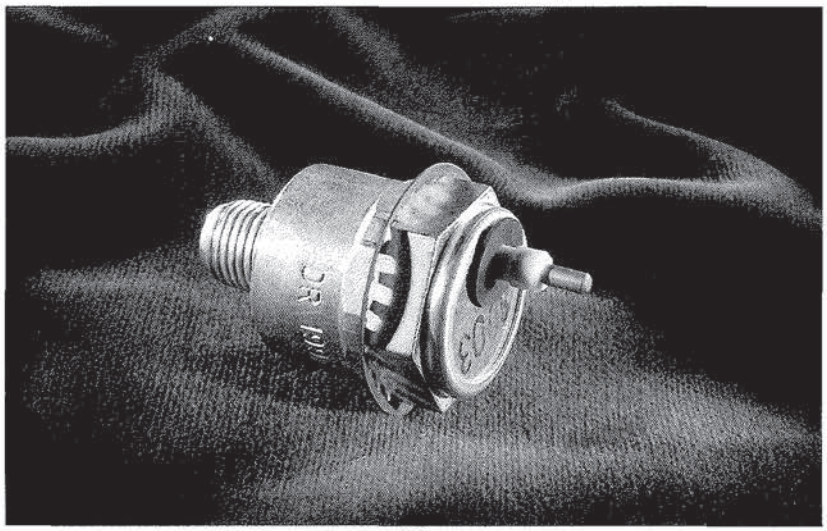
sensors measure the magnitude of a consequential parameter driven by the knock, rather than the knock phenomenon itself. The overall effectiveness of the control is therefore determined not only by the intrinsic performance and stability of the sensor and control, but also by how robust the chain of causality is between the knock phenomenon and the parameter measured.

8.3.1 Jerk Sensors

Naturally enough, the first turbocharged engine knock limit control used a jerk sensor, a productionized version of the kind of sensor used for laboratory CFR engine testing. An exploded view of this sensor is shown in Fig. 8.1.



(a)



(b)

FIGURE 8.1 (*a*) An exploded view of a jerk sensor, and (*b*) a photograph of the sensor.

Referring to the exploded view, it must be understood that when the sensor is fully assembled, the spider spring is preloaded, and all parts of the sensor except the coil cover are in compression. The nickel alloy magnetostrictive rods are biased mechanically in lengthwise compression and magnetically by the field impressed through the rods from the permanent magnet, returning through the ferromagnetic soft iron coil cover. The nickel alloy rods are the highest reluctance element in this magnetic circuit, and are in magnetic saturation.

The vibrations which are picked up and transmitted from the engine block through the mounting stud appear in the nickel alloy rods. These rods are of such a length that they comprise a broad passband mechanically resonant element. The compressive mechanical bias is sufficiently strong that the compression and rarefraction waves picked up are never large enough to take the rods out of net compression. The waves present in the rods therefore linearly modulate the magnetic reluctance of the magnetic circuit.

The many-turn coil wound around the magnetostrictive rods generates a voltage proportional to the rate of change of the magnetic flux; the flux, in turn, is inversely proportional to the magnetic reluctance of the rods. Since the vibrations picked up are already due to accelerations from the knock reverberations transmitted through the engine block, the voltage from the coil represents the third-time derivative of displacement, or jerk.

The vibration signal from a knocking engine is present virtually everywhere on the engine block, with signals from all cylinders superimposed. For passenger car engines, which rarely have more than eight cylinders, the major part of the knock signatures from the successive cylinder events are not superimposed in time, but rather consecutive without overlap. The time delay due to the distance from the cylinder to the sensor is always much less than 1 ms, whereas the cylinder-to-cylinder time is 2.5 ms even for an eight-cylinder engine at 6000 rpm.

However, the *absolute amplitude* of the knock signal does vary from location to location on the block. There is no best place to mount the sensor; rather, a site with a strong amplitude should be selected for a given engine model, and the sensor should always be mounted at that point on that engine model. The knock threshold is experimentally determined on that engine model with the sensor mounted at that location.

The method used for processing the knock signal and effecting control was described in Sec. 8.2.4. As time has gone on, the knock signal processing has tended to become simpler, although in the 1990s this may be changing. Time windows and mechanically resonant sensors have largely disappeared, and much of the filtering is now done digitally.

8.3.2 Accelerometer Sensors

The magnetostrictive jerk sensor, while satisfactory in performance, has too many parts to be a low-cost solution to measuring knock on board. Moreover, its mechanical assembly is not simple. In an effort to achieve lower-cost solutions, the industry found that the second-time derivative of displacement—acceleration—could be measured and used to implement a satisfactory knock control. The drawback to using an accelerometer compared to a jerk sensor is that it does not yield as good a signal-to-noise ratio; however, by using appropriate filtering, a good control signal can be achieved.

The first accelerometers used were mechanically bandpass-tuned, and some still are. However, in the 1990s, the trend has been to broadband sensors, again used with much electronic filtering. This has the advantage that one sensor model can be used for all engines, with the engine-specific frequency filtering characteristics built into the electronics module, which has to be engine-specific anyway. If the crucial filtering features are provided digitally, they can be installed as part of the engine-specific software.

Piezoelectric Accelerometer. Certain crystals of a specific cut or orientation possess the property that, when stressed, they produce a corresponding voltage. If such a crystal is loaded with a mass and spring, vibrations entering the assembly squeeze and pull the crystal with respect to the mass, and result in a signal which is a measure of the acceleration of the body to

which the assembly is attached. These signals can be substantial in size but appear electrically to be sourced at a high (capacitive) driving point impedance. Thus, care has to be taken to ensure that capacitively coupled noise does not contaminate the signal.

There are a variety of single crystals which exhibit the piezoelectric effect, but silicon is not one of them. Efforts to devise a silicon micromachined accelerometer were stymied for a long time by this fact, but a solution will be described below. Single crystal quartz is the material which was used in the single-crystal piezoelectric accelerometers.

In some cases, it is possible to use the mass of the piezoelectric element itself as the mass, and to combine the spring with one of the crystal electrodes. In such a design, the only parts are a spring, a contact, the crystal, and the housing.

Piezoceramic Accelerometer. Another type of piezoelectric element is the piezoceramic device. Piezoceramics are not single crystals. They can be fired into just about any shape, just like a ceramic insulator or a vase. By impressing a high voltage across electrodes on the ceramic while the material is at a high temperature—above its Curie temperature—and then gradually lowering the temperature, the piezoceramic becomes *poled,* or piezoelectric across the electrodes.

The best piezoceramics tend to belong to the lead-zirconate-titanate (PZT) family. These materials have a Curie temperature from 250 to 500 °C. If the sensor ever gets above its Curie temperature, its poling becomes degraded, and its calibration is no longer valid. However, these engine block-mounted sensors will not experience temperatures higher than about 125 to 150 °C, even during hot soak, so they can be used for the knock sensor.

Single-crystal piezoelectric accelerometers also have a Curie temperature, but it is typically much higher than for the piezoceramics. Quartz, for example, has a Curie temperature over 600 °C.

Because the piezoceramic materials can be molded into any shape, handled roughly, sawn and machined, coated with thick-film metallic electrodes, and so on, they lend themselves to easy, low-cost mass production. As a result, they have become pretty much the knock sensor of choice for the automobile industry. A drawing of a piezoceramic accelerometer knock sensor is shown in Fig. 8.2. In this particular design, the spring and top electrode have been combined into one part, but the mass is a separate metallic part between the piezoceramic disc and the spring. The whole assembly is stacked on a mandrel which extends up from the base-stud which forms most of the package.

8.3.3 Silicon Accelerometer

As previously noted, silicon itself cannot be made piezoelectric. However, it is possible to make a silicon micromachined accelerometer by anisotropically machining a cantilevered T-shaped structure into each cell of a silicon wafer, using a chemical etch. Strain gauges are diffused or implanted into the cantilever—the vertical bar of the T. The mass of the horizontal crossbar of the T and the spring constant of the cantilever provide the mechanical-to-strain transduction. Such accelerometers are in mass production for use in antiskid braking systems and as triggers for air bag safety devices. Whether or not they will replace piezoceramic knock sensors is primarily a matter of economics. Technically, the silicon accelerometer is satisfactory; its temperature limit is about 150 °C, which is adequate for on-block mounting.

8.3.4 Other Sensors

Instantaneous Cylinder Pressure Sensor. The direct measurement of instantaneous cylinder pressure permits the extraction of the pressure reverberation signal, which is the direct cause of knock. While this would be a very desirable signal to use for knock detection, it has not been implemented in on-board knock control systems for several reasons:

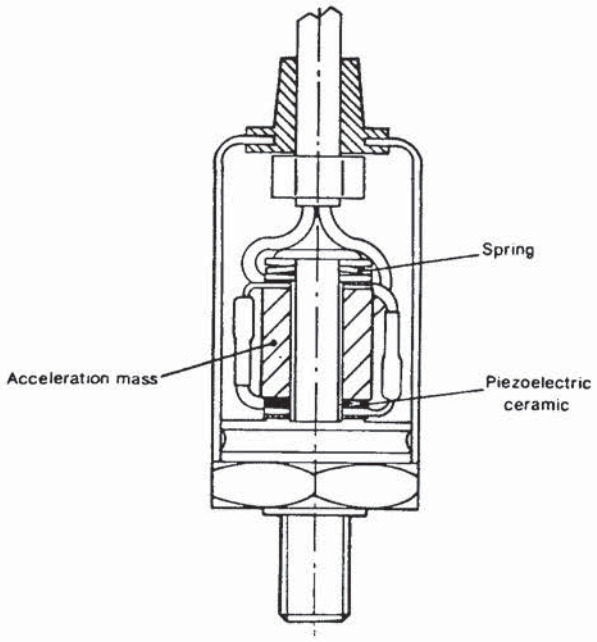


FIGURE 8.2 A piezoceramic accelerometer knock sensor.

- Either the same cylinder must always be the one to experience knock first and most severely, or one must have a sensor on every cylinder.
- The cylinder pressure wave is complex, and knock is only one of many signature elements present. Deriving a unique knock signal requires considerable signal processing in real time.
- While cylinder pressure sensors exist which are suitable for test purposes, a durable, low-cost, mass-producable on-board pressure sensor is not yet available.

In spite of these difficulties, work continues on cylinder pressure sensors for on-board control use, not so much to be used as knock-sensing devices as for use in deriving an instantaneous torque measurement (see Chap. 9). If this approach succeeds, then it is reasonable to suppose that the cylinder pressure sensor will be used to provide the inputs for obtaining both torque and knock measurements.

Since a spark-ignited engine already has a spark plug mounted in each cylinder head, it is possible to convert the spark plug into a kind of "poor man's pressure sensor" by mounting a piezoceramic washer held in compression between the spark plug and the cylinder head when the plug is torqued down. This arrangement has in fact appeared on a commercial passenger car engine. While the signal is not presently being used to extract a knock signal, suitable signal processing would make this possible.

Hydrophone in Coolant. In Sec. 8.3.1, it was mentioned that the knock signal appears virtually everywhere on the block. In fact, it even appears in the coolant fluid faithfully reproduced. This fact led to experiments in which hydrophones were mounted in the coolant and used as knock sensors. This approach has never reached production because it has no advantage over a block-mounted sensor and is in fact more difficult to package and to mount on the engine.

8.4 SUMMARY

In a remarkably short span of time, the on-board sensing of knock and its control have become a rather standard feature of passenger car engine control electronics. At the same time, the on-board sensors have been developed and are now largely mature. Acceleration has become the parameter of choice for most of these control systems, and the required accelerometers are typically based upon piezoceramic technology; it seems likely that this will remain the case at least through the 1990s. The rapid proliferation of these systems was facilitated by three factors:

- A solid background of experimental knowledge about knock and how to control it was established in the laboratory during the several decades before lead-free gas was mandated.
- The three legislated pressures of emissions limits, CAFE fuel economy goals, and no-lead fuel, combined with a customer base used to engines without knock, made solving the knock problem a high priority.
- The already widespread use of digital microcomputer engine control made the addition of knock limit control a straightforward design change.

GLOSSARY

Curie temperature The temperature above which a piezoelectric crystal or piezoceramic element no longer reliably retains its original piezoelectric characteristics.

Jerk sensor A sensor which measures the third-time derivative of displacement.

Knock Acoustic engine noise and vibration of characteristic frequency caused by uneven combustion in the engine cylinder(s).

Octane rating A measure of the resistance of a gasoline fuel to knock, as compared to pure isooctane hydrocarbon.

Retarding Causing engine combustion to occur at a larger angle past top dead center, by delaying the time at which the signal to start the spark event occurs.

Turbocharger A device which compresses engine intake manifold air by using the engine exhaust to drive a turbine and compressor.

Wastegate valve A valve in the exhaust system of a turbocharged diesel engine which causes part of the exhaust to bypass the turbine. The same name is also applied to a valve on a turbocharged gasoline engine which causes part of the boosted intake manifold air to dump to the atmosphere, dropping the manifold pressure.

REFERENCES

1. Paulius V. Puzinauskas, "Examination of methods used to characterize knock," SAE Paper 920808, Superflor Corp.
2. Karp P. Schmillen (FEV Motorentechnik, Aachen, Germany) and Manfred Rechs (Institute for Applied Thermodynamics, Technical University of Aachen, Germany), "Different methods of knock detection and knock control," SAE Paper 910858.
3. Masayoshi Kaneyasu (Hitachi America, Ltd.), Nobuo Kurihara, Kozo Katogi, and Hiroatsu Tokuda (Hitachi Ltd.), "Engine knock detection using multi-spectrum method," SAE Paper 920702.

ABOUT THE AUTHOR

William G. Wolber, executive engineer, Cummins Electronics Co., Inc., has 38 years' experience in the development of sensors, actuators, and instruments for application in automobiles and other products. Since 1981, he has been involved principally in the development of sensors and controls for heavy duty diesel engines for Cummins Electronics Co. He holds 15 U.S. patents and is the author of over 50 technical papers. He has lectured and conducted seminars internationally for the Society of Automotive Engineers and other organizations.

CHAPTER 9
ENGINE TORQUE SENSORS

William G. Wolber
Executive Engineer
Cummins Electronics Co. Inc.

9.1 INTRODUCTION

Torque is one of the primary state parameters of an engine; along with speed it is a fundamental measure related to the output power. Torque can be defined as the moment produced by the engine crankshaft tending to cause the output driveline to turn and thereby deliver power to the load. For rotary motion, the torque multiplied by the rotational speed equals the power delivered by the shaft. In derivative form, Newton's law states that torque T equals the rotational moment of inertia I times the angular acceleration α. Hence, at constant rotational speed, instantaneous power is proportional to instantaneous torque. This is an important relationship because torque and, thus, delivered power can change rapidly compared to rotational speed.

Portraying the engine as a torque-delivering device is a useful concept. When an engine is used in road service as prime mover power, the operator of the vehicle tends to ask the vehicle for positive or negative incremental acceleration, as the operator perceives it to be falling behind or closing upon the vehicle ahead. Hence, the operator asks the engine for more or less torque. If the engine is spark ignited, this happens naturally through the modulation of the engine air intake flow path. In a diesel engine, it happens indirectly. The operator controls fuel rate, which to first approximation is linearly related to power. For slowly varying engine speed and load, an incremental change in power results in a change in torque and therefore a change in acceleration.

9.1.1 Time-Scale Definitions

At this point, it is useful to define two time scales which will be used with regard to the engine. A reciprocating engine is never truly operating steady state. It is a cyclic succession of batch processes, a kind of system which chemical engineers have termed "continuous-by-jerks." Each cylinder absorbs shaft energy during the compression stroke and releases a larger amount of shaft energy through combustion during the power stroke. These actions are so interleaved that the drive shaft is continually accelerating and decelerating in step with cylinder events. Nevertheless, a quasi-steady-state model for the engine can be defined in which these rapid fluctuations are ignored and attention is focused on the more slowly varying state parameters which are nearly stationary during one engine revolution. This is the situation the operator senses and controls and it is also the primary focus of the control algorithms in a microcomputer-based engine control today. (As of 1994, all electronic engine controls operate on the "engine time scale.")

Quasi-Steady-State Torque. Quasi-steady-state torque is defined as that running average value of torque that varies slowly with respect to the cylinder-to-cylinder period, but rapidly with respect to vehicle changes in motion and load. To quantize this state, it is noted that an unloaded engine coupled to flywheel inertia may accelerate from idle to maximum speed in one to five seconds when subjected to wide open throttle and that this performance is essentially reproduced whether the change in throttle position occurs in 20 ms or 1 ms. Quasi-steady-state torque is what the operator is interested in commanding, and so, if we visualize a torque command-feedback engine control, it is the parameter of interest.

Instantaneous Torque. Today's electronic engine control can usefully respond much more rapidly than this for those state parameters having to do with each cylinder batch process itself. The preparation of the fuel and air charges and the timing of ignition take place on a time scale measured in fractions of a cylinder period, and thus the time scale of interest ranges from 30 μs to 20 ms. This time scale corresponds to the instantaneous engine state. It is the time scale of the torque impulses that are termed instantaneous torque.

In order to use instantaneous torque measurements, it must be recognized that the reciprocating engine is a cyclic machine in which major functions and the parameters that characterize them are tied together in sequence mechanically by the crankshaft and camshaft. Proper adjustment of those other parameters that are free to be changed independently of the train of instantaneous cylinder events allows optimization of the overall torque impulse generation in real time. For example, a theoretical examination of the internal combustion engine cycle has shown that the centroid of the cylinder pressure wave should occur about 15 crank-angle degrees after top dead center ("Powell's Magic Angle") to maximize the instantaneous incremental torque from that cylinder.[1,2] No spark or injection timing algorithm can do any better than to effect such timing of the torque impulse. Note, however, that we are stating where the centroid of the pressure wave should be for the best torque impulse, not where the fuel injection or ignition should be commanded.

State-of-the-art scheduled spark timing controls are based upon test data from one or more prototype engines run extensively in the laboratory on a dynamometer under quasi-steady-state conditions. From the data obtained, families of spark plug or injection timing (crank angle) surfaces versus engine speed, load, etc., are derived that correspond to optimum economy and performance while meeting emission requirements. These data then form the schedule embedded in the control microcomputer memory for all engines of that model produced. This type of scheduled computer control is standard throughout the industry today for all electronically controlled engines, both spark-ignited and compression ignition, with just two exceptions. These are the oxygen sensor feedback control loop, which keeps an engine equipped with a three-way catalyst in the stoichiometric "window" of the catalyst efficiency performance, and the speed governing feedback control algorithm used for power takeoff (PTO) operations, diesel all-speed governors, and generator-set frequency control.

By contrast, the addition of an on-board instantaneous torque measurement would permit the direct feedback control of several key engine control parameters. To take maximum advantage of such feedback control, an advanced instantaneous crankshaft position sensor is also required.

The controls which could be implemented through use of an instantaneous torque feedback control system are:

- Quasi-steady-state torque feedback control in response to torque command.
- Spark-ignition or compression-ignition timing control by feedback of instantaneous torque impulse timing compared to the norm of the "Magic Angle." Torque impulse amplitude can also be optimized against timing. These controls can be implemented on a cylinder-by-cylinder basis, for spark-ignited pulse sequential fuel injected engines, and for unit-injected diesel engines.
- Cylinder-by-cylinder torque-leveling feedback control can be implemented by adjusting the amount of fuel injected into each cylinder to obtain equal torque impulses from each one.

This results in a quiet, smoothly running engine with a somewhat improved maximum power, because the torque limit can be set for a more nearly constant cylinder event instead of for the worst case cylinder.

- The instantaneous torque signal is also rich in a variety of diagnostic information—for example, to measure engine power or for use as a miss detector.

9.2 AUTOMOTIVE APPLICATIONS OF TORQUE MEASUREMENT

9.2.1 Off-Board Measurement

The principal automotive use of torque measurement today is in the testing and evaluation of engines using the engine dynamometer. The torque sensor is inserted as a drive shaft between the engine and the dynamometer; the shaft of the sensor is a torsional Hooke's law member and its twist is measured. Typically the torque is actually measured using a strain gage bridge on the drive shaft. Full-scale twist is limited to no more than about a degree per foot of shaft for engines in the 50 to 500 hp range.

In a dc torque sensor, the strain gage bridge is powered from an external source through slip rings, and the offset voltage is taken off the bridge the same way (Fig. 9.1). The sensor is necessarily bulky, large, fragile, and costly. A more rugged ac strain gage sensor can be made using rotational transformers, as shown schematically in Fig. 9.2.

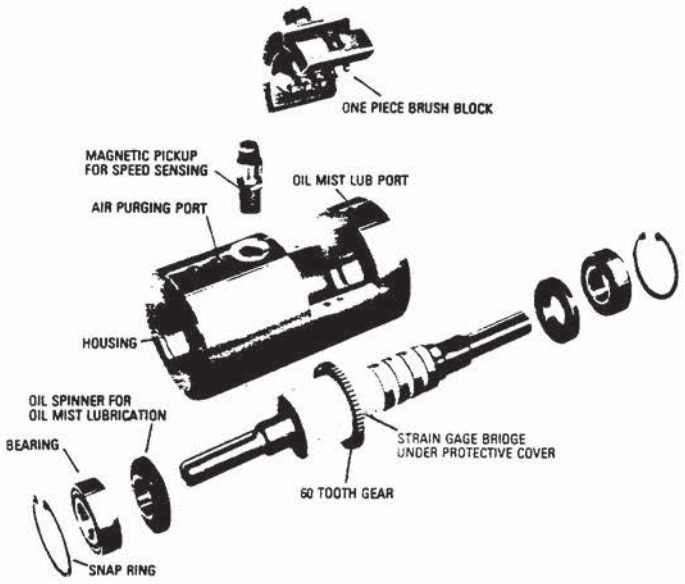


FIGURE 9.1 Exploded view of rotating shaft torque sensor with slip rings. *(Courtesy of the Lebow Div., the Eaton Corp.)*

Developmental Tests. In engine development, a good bit of the work involves fine-tuning the engine to get the maximum power and minimum fuel consumption at various engine conditions. Since power equals torque multiplied by engine speed, if tests are carried out at constant speed, maximum power occurs at maximum torque. Thus, the torque sensor can be used as a direct indicator of whether changes in the engine design or control will improve power.

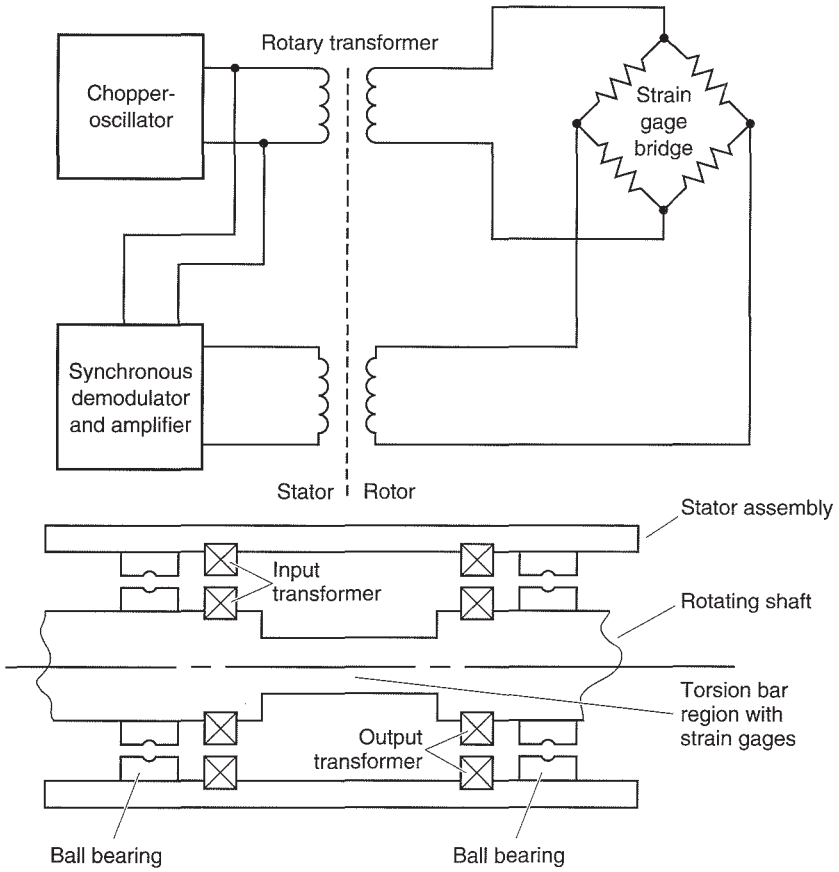


FIGURE 9.2 Rotary transformer rotating shaft torque sensor.

End-of-Line Manufacturing Tests. Particularly in the case of heavy duty diesel engines, the price of the engine is directly related to engine rated horsepower. Such engines are in a real sense sold by the horsepower. The manufacturer certifies that a certain model will deliver no less than a certain horsepower when run at rated conditions. The engine is tested at the end of the assembly line for a number of attributes, chief among which is assurance that it exceeds stated horsepower at rated speed. The assurance is provided by accurate torque and speed measurements.

In-Service Tests. Occasionally, torque sensors may be used with an engine that is being tested under field conditions. This is relatively straightforward when the engine is being used in a stationary application, but not if it is installed in a vehicle.

9.2.2 On-Board Measurements

To be most useful on board, a torque sensor should not only be accurate but also fast in response. Most of the sensors available today do not have this attribute, because they were developed for use in dynamometer tests.

To be a useful on-board engine control sensor, the torque sensor of the future also needs several other attributes:

- It must be extremely rugged, which for practical purposes means noncontacting with respect to the driveline.
- It must be able to withstand torque impulses of 10 to 20 times its full-scale measuring capability without degrading its accuracy. For strain gage type sensors, this means that the sensor would always have to measure in the lowest 5 to 10 percent of its range, where it is not very accurate.
- Installation of the sensor must not cause any significant change in the compliance of the driveline. This is a requirement imposed by engine application engineers to avoid spoiling the tradeoffs that have been made to avoid severe torsional vibrations.
- Ideally, the shaft of the sensor should be very short in length (in part to achieve the preceding characteristics). Unfortunately, for a torsional Hooke's law device, that means the full-scale shaft twist angle will be very small.

In spite of these formidable barriers to a satisfactory design, a wide variety of desirable applications could be made using an on-board instantaneous torque sensor. As a result, many researchers and developers both in academia and industry are working to achieve such a measurement directly or indirectly.

Power Measurement of Large Engines. Means exist for field service personnel to measure engine power in the field for engines installed in a vehicle and rated at up to 400 to 500 hp. For larger vehicle-mounted engines, this measurement is a difficult problem for a number of reasons:

- A rugged on-board sensor is not available.
- These engines are often installed in very large machines such as mine-haul trucks and front-end loaders located off the road in remote areas.
- Dynamometers large enough to load down the engine to full power are not available except in a very few locations.

One of the most frequent customer complaints about heavy duty engines is perceived loss of power. Often such complaints must be resolved at the field location and on a subjective basis. The engine and vehicle can be fully loaded in actual work operation, and engine speed is known accurately. If an on-board torque sensor were available, such complaints could be quickly resolved, and it would be well worth installing on large engines.

Miss Detection. The California Air Resources Board has mandated that all passenger vehicles shall be equipped with an engine diagnostic capability that includes a miss detector. While the regulation is not yet in force as of the time of this writing, the automotive industry knows that every segment will be affected by the requirement by the end of the decade. Many vehicle OEMs, first-tier automotive suppliers, entrepreneurial sensor firms, and academic researchers are working hard to meet the miss detector requirements, mainly basing their efforts on the same technologies useful for measuring or inferring torque.[3]

Torque Feedback Control. The reciprocating engine could be configured in a feedback control mode if equipped with an on-board quasi-steady-state torque sensor capable of responding at least as fast as the engine/driveline. In a torque command drive-by-wire system, the operator input would be a command for a new torque, greater or less than the present torque. The engine control would do its best to provide the new torque as rapidly as possible. The feedback error signal would be the difference between the torque command and the sensor output. Of course, the new torque and the resulting acceleration or deceleration would have to be limited within bounds set by maxima for cylinder pressure, emissions production, exhaust gas temperature, fuel economy, and so on.

Feedback Control of Timing. As was described in the introduction to this chapter, the measurement of instantaneous torque, on board, along with instantaneous crank angle position,

would allow the engine control computer to use signature recognition to measure the crank angle at which the centroid of the torque impulse is located, for each cylinder. From this information, probably with some compensation for seriously off-normal engine conditions as stored in computer memory, the engine control can determine if that impulse was centered at Powell's Magic Angle, and, if not, how much earlier or later the chain of combustion events needs to be started. Most likely, the computer also needs to decide between updating that cylinder based upon that information alone and waiting a whole engine revolution to update, or also basing the correction upon abrupt changes in operator input, load, knock level, etc. In any case, under quasi-steady-state conditions, the best efficiency of that engine for converting chemical energy into shaft power by proper engine timing could be achieved this way whether the engine is spark-ignited or compression-ignition.

Cylinder Torque Leveling. Individual trimming of the amount of fuel for each cylinder can be realized on pulse sequential fuel injected spark-ignited engines and on unit injector diesel engines by matching the torque impulse for each cylinder to the average of all cylinders. Then the average impulse can be raised to very nearly the maximum value the engine is designed for. In today's engines, the control must provide a guard band between the maximum torque for the worst case cylinder and the design limit for the engine (commonly called the maximum cylinder pressure limit or the peak torque limit). If the torque impulses delivered by each cylinder were known to be controlled to be equal, the guard band could be reduced and more power obtained from a given engine.

While cylinder-to-cylinder torque leveling is not the same as compression leveling, an engine controlled cylinder by cylinder for timing and torque leveling intuitively would appear to be as smoothly running and low in acoustic noise emission as can be visualized.

Engine Diagnostics. The instantaneous torque signature is very rich in diagnostic information. A figure of merit for engine roughness can easily be obtained in real time, from which a lean-limit control signal can be derived. The miss detector responds to an extreme case of roughness. Ragged ignition and flame front formation, imperfect injection, poor distribution, and many other engine and control problems can be measured or inferred from such a signal, on board and in real time.

9.3 DIRECT TORQUE SENSORS

9.3.1 Reaction Force

The use of a strain-gaged torsional Hooke's law shaft to measure torque at the dynamometer has been described in Sec. 9.2.1. Torque can also be measured in the test cell by instrumenting the engine or dynamometer mounting with strain gage load cells to measure reaction torque. This is a useful way to measure quasi-steady-state torque, but is not a good measuring system for instantaneous torque, as both the engine and its load have considerable inertia and damping which attenuate the torque impulse and may introduce phase delay. It is also not a good on-board measurement because the engine mounts on a vehicle are designed to attenuate and damp out bouncing from road and load irregularities. Also, as a function of time, temperature, ozone concentration, and other variables, they tend to stiffen substantially and change the deflection of the engine at its mounts as a function of torque.

9.3.2 Torsional Strain (Twist) Sensors

Magnetic Vector Sensors. The strain-gaged torsional Hooke's law sensor was described in Sec. 9.2.1. A more practical approach to an on-board torque sensor is a noncontacting design

called a magnetic vector sensor.[4,5] This sensor operates on the principle that the magnetic domains in a ferromagnetic shaft delivering torque are randomly distributed when no torque is being delivered, but that each domain, on average, is slightly rotated in the tangential direction when torque is delivered by the shaft and twists it. If an ac-driven coil is placed near the shaft and surrounded by four sensing coils arranged mechanically and electrically as a bridge, the amplitude of the bridge offset is proportional to the magnetic vector tangential component, and therefore to twist and torque.

If the magnetic domains were truly statistically distributed over a small range of shaft angle, such a sensor would be able to measure instantaneous torque. Unfortunately, they are not. Over a small shaft angle increment, the average torque vector does not net to zero when torque is zero, and the output is characterized by a fixed pattern noise which repeats every 360°. The magnetic vector sensor is reasonably accurate measuring quasi-steady-state torque, but the signal-to-noise ratio becomes poor when an attempt is made to measure instantaneous torque.

Further work on this sensor simplified and miniaturized it.[6] The bridge was replaced by a single tangential coil, and the sensor was made small enough so that it could be mounted in the rear crankshaft bearing. However, it still was not able to measure instantaneous torque.

Investigation continues on techniques to reduce the fixed pattern noise, and some claims of progress have been made. However, at this writing, a production on-board torque sensor is not available.

Optical Twist Measurement. Research work has been reported on a sensor that changes the duty factor of light pulses sensed after passing through sets of slits at both ends of a shaft.[7] The principle of its operation is shown in Fig. 9.3. Such a sensor is noncontacting and can be made robust, but it requires appreciable shaft length and therefore adds compliance to the driveline. The work has not been engineered to produce an economical on-board device.

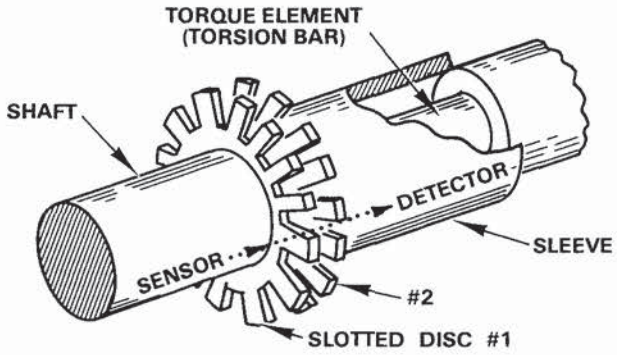


FIGURE 9.3 Optical torque meter. *(Courtesy of The Bendix Corp.)*

Capacitive Twist Sensor. An electrode pattern can be made using interdigitated electrodes spaced one or two degrees apart on two facing flat discs.[8] One of the discs is stationary; the other rotates with the crankshaft. Two such pairs of electrodes can be operated with phase detection measurement to provide a virtually instantaneous signal proportional to the twist of the shaft; the rotating halves of the electrode pairs are attached to the ends of the Hooke's law torsional spring. Obviously, this arrangement has the same drawback as the optical sensor in that it has to have some shaft length to twist and, therefore, some compliance. However, it is robust, has good accuracy, and is a more practical on-board device than the optical twist sen-