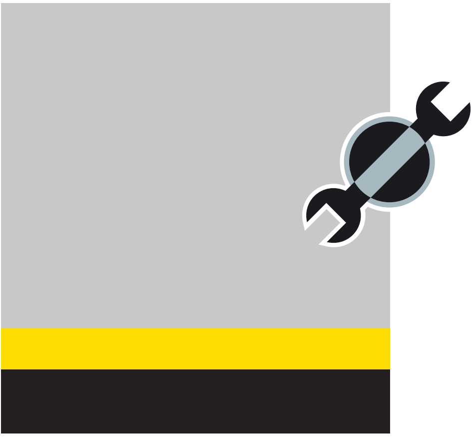
Automobile Electrical and Electronic Systems

Fourth Edition

Tom Denton

Automotive Technology:
Vehicle Maintenance and Repair

B
H



Automobile Electrical and Electronic Systems

Fourth Edition

(Automotive Technology: Vehicle Maintenance and Repair)

This page intentionally left blank

Automobile Electrical and Electronic Systems

Fourth Edition

Automotive Technology:
Vehicle Maintenance and Repair

Tom Denton BA FIMI MSAE
MIRTE Cert Ed

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Fourth Edition first published in 2012
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

First edition published 1995 by Arnold, a member of Hodder Headline plc.
Third edition published 2004 by Elsevier

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
[CIP data]

ISBN: 978-0-08-096942-8 (pbk)
ISBN: 978-0-080-96943-5 (ebk)

Typeset in Helvetica
by RefineCatch Limited, Bungay, Suffolk

Contents

Chapter 10 Engine management 345

Preface

Automobile electrical and electronic systems are at the same time the most complex yet most interesting aspects of a vehicle. Well, they are to me anyway, which is why I am particularly pleased to have produced the fourth edition of this book!

In this edition you will find more details on EVs and HEVs as well as some of the latest ideas about vehicle networks and much more. This book is the second in the ‘Automotive Technology: Vehicle Maintenance and Repair’ series:

- Automobile Mechanical and Electrical Systems
- Automobile Electrical and Electronic Systems, 4th edition
- Automobile Advanced Fault Diagnosis, 3rd edition

Ideally, you will have studied the mechanical book, or have some experience, before starting on this one. If not, it does start with the basics. This is the first book of its type to be published in full colour, and concentrates on electrical and electronic principles as well as comprehensive case studies and examples. It will cover everything you need to advance your studies to a higher level, no matter what qualification (if any) you are working towards.

I hope you find the content useful and informative. Comments, suggestions and feedback are always welcome at my website: www.automotive-technology.co.uk. You will also find links to lots of free online resources to help with your studies.

The final chapter of this book contains assignments, questions, research topics and more. You can look at this at any time or wait until you have studied the rest of the book.

Good luck and I hope you find automotive technology as interesting as I still do.

This page intentionally left blank

Acknowledgements

Over the years many people have helped in the production of my books. I am therefore very grateful to the following companies who provided information and/or permission to reproduce photographs and/or diagrams:

AA Photo Library
AC Delco
Alpine Audio Systems
ATT Training (UK and USA)
Autologic Data Systems
BMW UK
Bosch Gmbh
Bosch Media
C&K Components
Citroën UK
Clarion Car Audio
Delphi Media
Eberspaecher
Fluke Instruments UK
Ford Motor Company
Ford Media
FreeScale Electronics
General Motors
GenRad
Hella UK
Honda Cars UK
Hyundai UK
Jaguar Cars
Kavlico
Loctite
Lucas UK
LucasVarity
Mazda
McLaren Electronic Systems
Mercedes Cars UK
Mitsubishi Cars UK
NGK Plugs
Nissan Cars UK
Most Corporation
Peugeot UK
Philips
PicoTech
Pioneer Radio
Porsche Cars UK
Robert Bosch GmbH.
Robert Bosch UK
Rover Cars
Saab Cars UK
Saab Media
Scandmec
SMSC
Snap-on Tools
Sofanou (France)
Sun Electric UK
Tesla Motors
Thrust SSC Land Speed Team
T&M Auto-Electrical
Toyota Cars UK
Tracker UK
Unipart Group
Valeo
Vauxhall
VDO Instruments
Volvo Media
Volkswagen cars
Wikimedia
ZF Servomatic

If I have used any information, or mentioned a company name that is not listed here, please accept my apologies and let me know so it can be rectified as soon as possible.

This page intentionally left blank

Glossary of abbreviations and acronyms

OBD2/SAE terminology

ABS	*antilock brake system*
AC	*air conditioning*
AC	*air cleaner*
AIR	*secondary air injection*
A/T	*automatic transmission or transaxle*
SAP	*accelerator pedal*
B+	*battery positive voltage*
BARO	*barometric pressure*
CAC	*charge air cooler*
CFI	*continuous fuel injection*
CL	*closed loop*
CKP	*crankshaft position sensor*
CKP REF	*crankshaft reference*
CMP	*camshaft position sensor*
CMP REF	*camshaft reference*
CO	*carbon monoxide*
CO_2	*carbon dioxide*
CPP	*clutch pedal position*
CTOX	*continuous trap oxidizer*
CTP	*closed throttle position*
DEPS	*digital engine position sensor*
DFCO	*decel fuel cut-off mode*
DFI	*direct fuel injection*
DLC	*data link connector*
DPF	*diesel particulate filter*
DTC	*diagnostic trouble code*
DTM	*diagnostic test mode*
EBCM	*electronic brake control module*
EBTCM	*electronic brake traction control module*
EC	*engine control*
ECM	*engine control module*
ECL	*engine coolant level*
ECT	*engine coolant temperature*
EEPROM	*electrically erasable programmable read only memory*
EFE	*early fuel evaporation*
EGR	*exhaust gas recirculation*
EGRT	*EGR temperature*
EI	*electronic ignition*
EM	*engine modification*

EPROM	*erasable programmable read only memory*
ESC	*electronic stability control*
EVAP	*evaporative emission system*
FC	*fan control*
FEEPROM	*flash electrically erasable programmable read only memory*
FF	*flexible fuel*
FP	*fuel pump*
FPROM	*flash erasable programmable read only memory*
FT	*fuel trim*
FTP	*federal test procedure*
GCM	*governor control module*
GEN	*generator*
GND	*ground*
H_20	*water*
HO2S	*heated oxygen sensor*
HO2S1	*upstream heated oxygen sensor*
HO2S2	*up or downstream heated oxygen sensor*
HO2S3	*downstream heated oxygen sensor*
HC	*hydrocarbon*
HVS	*high voltage switch*
HVAC	*heating ventilation and air conditioning system*
IA	*intake air*
IAC	*idle air control*
IAT	*intake air temperature*
IC	*ignition control circuit*
ICM	*ignition control module*
IFI	*indirect fuel injection*
IFS	*inertia fuel shutoff*
I/M	*inspection/maintenance*
IPC	*instrument panel cluster*
ISC	*idle speed control*
KOEC	*key on, engine cranking*
KOEO	*key on, engine off*
KOER	*key on, engine running*
KS	*knock sensor*
KSM	*knock sensor module*
LTFT	*long term fuel trim*
MAF	*mass airflow sensor*
MAP	*manifold absolute pressure sensor*
MC	*mixture control*
MDP	*manifold differential pressure*
MFI	*multiport fuel injection*
MIL	*malfunction indicator lamp*
MPH	*miles per hour*
MST	*manifold surface temperature*
MVZ	*manifold vacuum zone*
NVRAM	*non-volatile random access memory*
NO_x	*oxides of nitrogen*
O2S	*oxygen sensor*
OBD	*on-board diagnostics*
OBDI	*on-board diagnostics generation one*
OBD II	*on-board diagnostics, second generation*
OC	*oxidation catalyst*

ODM	*output device monitor*
OL	*open loop*
OSC	*oxygen sensor storage*
PAIR	*pulsed secondary air injection*
PCM	*powertrain control module*
PCV	*positive crankcase ventilation*
PNP	*park/neutral switch*
PROM	*program read only memory*
PSA	*pressure switch assembly*
PSP	*power steering pressure*
PTOX	*periodic trap oxidizer*
RAM	*random access memory*
RM	*relay module*
ROM	*read only memory*
rpm	*revolutions per minute*
SC	*supercharger*
SCB	*supercharger bypass*
SDM	*sensing diagnostic mode*
SFI	*sequential fuel injection*
SRI	*service reminder indicator*
SRT	*system readiness test*
STFT	*short term fuel trim*
TB	*throttle body*
TBI	*throttle body injection*
TC	*turbocharger*
TCC	*torque converter clutch*
TCM	*transmission or transaxle control module*
TFP	*throttle fluid pressure*
TP	*throttle position*
TPS	*throttle position sensor*
TVV	*thermal vacuum valve*
TWC	*three way catalyst*
TWC+OC	*three way + oxidation catalytic converter*
VAF	*volume airflow*
VCM	*vehicle control module*
VR	*voltage regulator*
VS	*vehicle sensor*
VSS	*vehicle speed sensor*
WU-TWC	*warm up three way catalytic converter*
WOT	*wide open throttle*

OEM and other terminology

A	*amps*
AC	*air conditioning*
A/F	*air/fuel ratio*
A/T	*automatic transmission*
AAV	*anti-afterburn valve (Mazda)*
ABS	*antilock brake system*
ABSV	*air bypass solenoid valve (Mazda)*
AC	*alternating current*
ACTS	*air charge temperature sensor (Ford)*

AERA	*Automotive Engine Rebuilders Assn.*
AFM	*air flow meter*
AFS	*air flow sensor (Mitsubishi)*
AIR	*Air Injection Reaction (GM)*
AIS	*Air Injection System (Chrysler)*
AIS	*automatic idle speed motor (Chrysler)*
ALCL	*assembly line communications link (GM)*
ALDL	*assembly line data link (GM)*
API	*American Petroleum Institute*
APS	*absolute pressure sensor (GM)*
APS	*atmospheric pressure sensor (Mazda)*
ASD	*automatic shutdown relay (Chrysler)*
ASDM	*airbag system diagnostic module (Chrysler)*
ASE	*Automotive Service Excellence*
ATC	*after top centre*
ATDC	*after top dead centre*
ATF	*automatic transmission fluid*
ATMC	*Automotive Training Managers Council*
ATS	*air temperature sensor (Chrysler)*
AWD	*all-wheel drive*
BARO	*barometric pressure sensor (GM)*
BAT	*battery*
BCM	*body control module (GM)*
BHP	*brake horsepower*
BID	*Breakerless Inductive Discharge (AMC)*
BMAP	*barometric/manifold absolute pressure sensor (Ford)*
BP	*backpressure sensor (Ford)*
BPS	*barometric pressure sensor (Ford & Nissan)*
BPT	*back-pressure transducer*
BTC	*before top centre*
BTDC	*before top dead centre*
Btu	*British thermal units*
C	*Celsius*
C3	*Computer Command Control system (GM)*
C3I	*Computer Controlled Coil Ignition (GM)*
C4	*Computer Controlled Catalytic Converter system (GM)*
CAAT	*Council of Advanced Automotive Trainers*
CAFE	*corporate average fuel economy*
CALPAK	*calibration pack*
CANP	*canister purge solenoid valve (Ford)*
CARB	*California Air Resources Board*
CAS	*Clean Air System (Chrysler)*
CAS	*crank angle sensor*
CC	*catalytic converter*
CC	*cubic centimetres*
CCC	*Computer Command Control system (GM)*
CCD	*computer controlled dwell (Ford)*
CCEI	*Coolant Controlled Idle Enrichment (Chrysler)*
CCEV	*Coolant Controlled Engine Vacuum Switch (Chrysler)*
CCOT	*clutch cycling orifice tube*
CCP	*controlled canister purge (GM)*
CCV	*canister control valve*
CDI	*Capacitor Discharge Ignition (AMC)*

CEAB	*cold engine air bleed*
CEC	*Crankcase Emission Control System (Honda)*
CECU	*central electronic control unit (Nissan)*
CER	*cold enrichment rod (Ford)*
CESS	*cold engine sensor switch*
CFC	*Chlorofluorocarbons*
CFI	*Cross Fire Injection (Chevrolet)*
cfm	*cubic feet per minute*
CID	*cubic inch displacement*
CID	*cylinder identification sensor (Ford)*
CIS	*Continuous Injection System (Bosch)*
CMP	*camshaft position sensor (GM)*
COP	*Coil On Plug ignition*
CP	*canister purge (GM)*
CP	*crankshaft position sensor (Ford)*
CPI	*Central Port Injection (GM)*
CPU	*central processing unit*
CSC	*Coolant Spark Control (Ford)*
CSSA	*Cold Start Spark Advance (Ford)*
CSSA	*Cold Start Spark Advance (Ford)*
CSSH	*Cold Start Spark Hold (Ford)*
CTAV	*Cold Temperature Actuated Vacuum (Ford)*
CTO	*Coolant Temperature Override Switch (AMC)*
CTS	*charge temperature switch (Chrysler)*
CTS	*coolant temperature sensor (GM)*
CTVS	*choke thermal vacuum switch*
CVCC	*Compound Vortex Controlled Combustion system (Honda)*
CVR	*control vacuum regulator (Ford)*
dB	*decibels*
DC	*direct current*
DEFI	*Digital Electronic Fuel Injection (Cadillac)*
DERM	*diagnostic energy reserve module (GM)*
DFS	*deceleration fuel shutoff (Ford)*
DIS	*Direct Ignition System (GM)*
DIS	*Distributorless Ignition System (Ford)*
DLC	*data link connector (GM)*
DOHC	*dual overhead cams*
DOT	*Department of Transportation*
DPF	*diesel particulate filter*
DRBII	*Diagnostic Readout Box (Chrysler)*
DRCV	*distributor retard control valve*
DSSA	*Dual Signal Spark Advance (Ford)*
DVDSV	*differential vacuum delay and separator valve*
DVDV	*distributor vacuum delay valve*
DVOM	*digital volt ohm meter*
EACV	*electronic air control valve (Honda)*
EBCM	*electronic brake control module (GM)*
EBM	*electronic body module (GM)*
ECA	*electronic control assembly*
ECCS	*Electronic Concentrated Control System (Nissan)*
ECM	*electronic control module (GM)*
ECS	*Evaporation Control System (Chrysler)*
ECT	*engine coolant temperature (Ford & GM)*

ECU	*electronic control unit (Ford, Honda & Toyota)*
EDIS	*Electronic Distributorless Ignition System (Ford)*
EEC	*Electronic Engine Control (Ford)*
EEC	*Evaporative Emission Controls (Ford)*
EECS	*Evaporative Emissions Control system (GM)*
EEPROM	*electronically erasable programmable read only memory chip*
EFC	*electronic feedback carburettor (Chrysler)*
EFC	*electronic fuel control*
EFCA	*electronic fuel control assembly (Ford)*
EFE	*Early Fuel Evaporation system (GM)*
EFI	*electronic fuel injection*
EGO	*exhaust gas oxygen sensor (Ford)*
EGRPS	*EGR valve position sensor (Mazda)*
EGR-SV	*EGR solenoid valve (Mazda)*
EGRTV	*EGR thermo valve (Chrysler)*
EI	*electronic ignition (GM)*
ELB	*Electronic Lean Burn (Chrysler)*
EMI	*electromagnetic interference*
EOS	*exhaust oxygen sensor*
EPA	*Environmental Protection Agency*
EPOS	*EGR valve position sensor (Ford)*
EPROM	*erasable programmable read only memory chip*
ESA	*Electronic Spark Advance (Chrysler)*
ESC	*Electronic Spark Control (GM)*
ESS	*Electronic Spark Selection (Cadillac)*
EST	*Electronic Spark Timing (GM)*
EVP	*EGR valve position sensor (Ford)*
EVRV	*electronic vacuum regulator valve for EGR (GM)*
F	*Fahrenheit*
FBC	*feedback carburettor system (Ford & Mitsubishi)*
FBCA	*feedback carburettor actuator (Ford)*
FCA	*fuel control assembly (Chrysler)*
FCS	*fuel control solenoid (Ford)*
FDC	*fuel deceleration valve (Ford)*
FI	*fuel injection*
FLS	*fluid level sensor (GM)*
FMVSS	*Federal Motor Vehicle Safety Standards*
ft.lb.	*foot pound*
FUBAR	*Fracked Up Beyond All Repair*
FWD	*front-wheel drive*
gal	*gallon*
GND	*ground*
GPM	*grams per mile*
HAIS	*Heated Air Intake System (Chrysler)*
HEGO	*heated exhaust gas oxygen sensor*
HEI	*High Energy Ignition (GM)*
Hg	*mercury*
hp	*horsepower*
I/P	*instrument panel*
IAC	*idle air control (GM)*
IAT	*inlet air temperature sensor (Ford)*
IATS	*intake air temperature sensor (Mazda)*
IC	*integrated circuit*

ICS	*idle control solenoid (GM)*
ID	*inside diameter*
IGN	*ignition*
IIIBDFI	*If it isn't broke don't fix it*
IM240	*inspection/maintenance 240 program*
IMI	*Institute of the Motor Industry*
ISC	*idle speed control (GM)*
ISC	*idle speed control (GM)*
ISO	*International Standards Organization*
ITCS	*Ignition Timing Control System (Honda)*
ITS	*idle tracking switch (Ford)*
JAS	*Jet Air System (Mitsubishi)*
kHz	*kilohertz*
KISS	*Keep It Simple Stupid!*
Km	*kilometres*
kPa	*kilopascals*
KS	*knock sensor*
KV	*kilovolts*
L	*litters*
lb. ft.	*pound feet*
LCD	*liquid crystal display*
LED	*light emitting diode*
MACS	*Mobile Air Conditioning Society*
MAF	*mass airflow sensor*
MAMA	*Midwest Automotive Media Assn.*
MAP	*manifold absolute pressure*
MAP	*Motorist Assurance Program*
MAT	*manifold air temperature*
MCS	*mixture control solenoid (GM)*
MCT	*manifold charge temperature (Ford)*
MCU	*Microprocessor Controlled Unit (Ford)*
MFI	*multiport fuel injection*
MIL	*malfunction indicator lamp*
MISAR	*Microprocessed Sensing and Automatic Regulation (GM)*
mm	*millimetres*
MPFI	*multi point fuel injection*
MPG	*miles per gallon*
MPH	*miles per hour*
MPI	*multi-port injection*
ms	*millisecond*
MSDS	*material safety data sheet*
mV	*millivolts*
NACAT	*National Assn. of College Automotive Teachers*
NATEF	*National Automotive Technician's Education Foundation*
NHTSA	*National Highway Traffic Safety Administration*
Nm	*Newton meters*
OBD	*on-board diagnostics*
OC	*oxidation converter (GM)*
OD	*outside diameter*
OE	*original equipment*
OEM	*original equipment manufacture*
OHC	*overhead cam*
ORC	*oxidation reduction catalyst (GM)*

OS	*oxygen sensor*
OSAC	*Orifice Spark Advance Control (Chrysler)*
P/B	*power brakes*
P/N	*part number*
PA	*pressure air (Honda)*
PAFS	*Pulse Air Feeder System (Chrysler)*
PAIR	*Pulsed Secondary Air Injection system (GM)*
PCM	*powertrain control module (supersedes ECM)*
PECV	*power enrichment control valve*
PERA	*Production Engine Rebuilders Assn.*
PFI	*port fuel injection (GM)*
PGM-FI	*Programmed Gas Management Fuel Injection (Honda)*
PIP	*profile ignition pickup (Ford)*
PPM	*parts per million*
PROM	*program read only memory computer chip*
PS	*power steering*
PSI	*pounds per square inch*
pt.	*pint*
PVA	*ported vacuum advance*
PVS	*ported vacuum switch*
PVS	*ported vacuum switch*
QS9000	*Quality assurance standard for OEM part suppliers*
Qt.	*quart*
RABS	*Rear wheel Antilock Brake System (Ford)*
RFI	*radio frequency interference*
rpm	*revolutions per minute*
RPO	*regular production option*
RWAL	*Rear Wheel Antilock brake system (GM)*
RWD	*rear-wheel drive*
SAE	*Society of Automotive Engineers*
SAVM	*spark advance vacuum modulator*
SCC	*Spark Control Computer (Chrysler)*
SDI	*Saab Direct Ignition*
SES	*service engine soon indicator (GM)*
SFI	*Sequential Fuel Injection (GM)*
SIR	*Supplemental Inflatable Restraint (air bag)*
SMPI	*Sequential Multiport Fuel Injection (Chrysler)*
SOHC	*single overhead cam*
SPOUT	*Spark Output signal (Ford)*
SRDV	*spark retard delay valve*
SRS	*Supplemental Restraint System (air bag)*
SS	*speed sensor (Honda)*
SSI	*Solid State Ignition (Ford)*
STS	*Service Technicians Society*
TA	*temperature air (Honda)*
TABPV	*throttle air bypass valve (Ford)*
TAC	*thermostatic air cleaner (GM)*
TACH	*tachometer*
TAD	*Thermactor air diverter valve (Ford)*
TAV	*temperature actuated vacuum*
TBI	*throttle body injection*
TCC	*torque converter clutch (GM)*
TCCS	*Toyota Computer Controlled System*

TCS	*Transmission Controlled Spark (GM)*
TDC	*top dead centre*
TIC	*thermal ignition control (Chrysler)*
TIV	*Thermactor idle vacuum valve (Ford)*
TKS	*throttle kicker solenoid (Ford)*
TP	*throttle position sensor (Ford)*
TPI	*Tuned Port Injection (Chevrolet)*
TPMS	*Tire Pressure Monitor System*
TPP	*throttle position potentiometer*
TPS	*throttle position sensor*
TPT	*throttle position transducer (Chrysler)*
TRS	*Transmission Regulated Spark (Ford)*
TSP	*throttle solenoid positioner (Ford)*
TV	*throttle valve*
TVS	*thermal vacuum switch*
TVS	*thermal vacuum switch (GM)*
TVV	*thermal vacuum valve (GM)*
V	*volts*
VAC	*volts alternating current*
VAF	*vane airflow sensor*
VCC	*viscous converter clutch (GM)*
VDC	*volts direct current*
VDV	*vacuum delay valve*
VIN	*vehicle identification number*
VSM	*vehicle security module*
VSS	*vehicle speed sensor*
WOT	*wide open throttle*
WOT	*wide open throttle switch (GM)*
WSS	*wheel speed sensor*

This page intentionally left blank

CHAPTER 1

Development of the automobile electrical system

1.1 A short history

1.1.1 Where did it all begin?

The story of electric power can be traced back to around 600 BC, when the Greek philosopher Thales of Miletus found that amber rubbed with a piece of fur would attract lightweight objects such as feathers. This was due to static electricity. It is thought that, around the same time, a shepherd in what is now Turkey discovered magnetism in lodestones, when he found pieces of them sticking to the iron end of his crook.

William Gilbert, in the sixteenth century, proved that many other substances are ‘electric’ and that they have two electrical effects. When rubbed with fur, amber acquires ‘resinous electricity’; glass, however, when rubbed with silk, acquires ‘vitreous electricity’. Electricity repels the same kind and attracts the opposite kind of electricity. Scientists thought that the friction actually created the electricity (their word for charge). They did not realize that an equal amount of opposite electricity remained on the fur or silk.

Key fact

The story of electric power can be traced back to around 600 BC.

A German, Otto Von Guerick, invented the first electrical device in 1672. He charged a ball of sulphur with static electricity by holding his hand against it as it rotated on an axle. His experiment was, in fact, well ahead of the theory developed in the 1740s by William Watson, an English physician, and the American statesman Benjamin Franklin, that electricity is in all matter and that it can be transferred by rubbing. Franklin, in order to prove that lightning was a form of electricity, flew a kite during a thunder-storm and produced sparks from a key attached to the string! Some good did come from this dangerous experiment though, as Franklin invented the lightning conductor.

Key fact

Alessandro Volta, an Italian aristocrat, invented the first battery.

Alessandro Volta, an Italian aristocrat, invented the first battery. He found that by placing a series of glass jars containing salt water, and zinc and copper electrodes connected in the correct order, he could get an electric shock by touching the wires. This was the first wet battery and is indeed the forerunner of the accumulator, which was developed by the French physicist Gaston Planche in 1859. This was a lead-acid battery in which the chemical reaction that produces electricity could be reversed by feeding current back in the opposite direction. No battery or storage cell can supply more than a small amount of power and inventors soon realized that they needed a continuous source of current. Michael Faraday, a Surrey blacksmith’s son and an assistant to Sir Humphrey Davy, devised the first electrical generator. In 1831 Faraday made a machine in which a copper disc rotated between the poles of a large magnet. Copper strips provided contacts with the rim of the disc and the axle on which it turned; current flowed when the strips were connected.

William Sturgeon of Warrington, Lancashire, made the first working electric motor in the 1820s. He also made the first working electromagnets and used battery-powered electromagnets in a generator in place of permanent magnets. Several inventors around 1866, including two English electricians – Cromwell Varley and Henry Wilde – produced permanent magnets. Anyos Jedlik, a Hungarian physicist, and the American pioneer electrician, Moses Farmer, also worked in this field. The first really successful generator was the work of a German, Ernst Werner Von Siemens. He produced his generator, which he called a dynamo, in 1867. Today, the term dynamo is applied only to a generator that provides direct current. Generators, which produce alternating current, are called alternators.

The development of motors that could operate from alternating current was the work of an American engineer, Elihu Thomson. Thomson also invented the transformer, which changes the voltage of an electric supply. He demonstrated his invention in 1879 and, 5 years later, three Hungarians, Otto Blathy, Max Deri and Karl Zipernowksy, produced the first commercially practical transformers.

It is not possible to be exact about who conceived particular electrical items in relation to the motor car. Innovations in all areas were thick and fast in the latter half of the nineteenth century.

In the 1860s, Ettiene Lenoir developed the first practical gas engine. This engine used a form of electric ignition employing a coil developed by Ruhmkorff in 1851. In 1866, Karl Benz used a type of magneto that was belt driven. He found this to be unsuitable though, owing to the varying speed of his engine. He solved the problem by using two primary cells to provide an ignition current.

In 1889, Georges Bouton invented contact breakers for a coil ignition system, thus giving positively tuned ignition for the first time. It is arguable that this is the ancestor of the present day ignition system. Emile Mors used electric ignition on a low-tension circuit supplied by accumulators that were recharged from a belt-driven dynamo. This was the first successful charging system and can be dated to around 1895.

The now formidable Bosch empire was started in a very small way by Robert Bosch. His most important area of early development was in conjunction with his foreman, Fredrich Simms, when they produced the low-tension magneto at the end of the nineteenth century. Bosch introduced the high-tension magneto

Key fact

William Sturgeon of Warrington, Lancashire, made the first working electric motor in the 1820s.

Key fact

In the 1860s, Ettiene Lenoir developed the first practical gas engine.

Key fact

In 1889, Georges Bouton invented contact breakers for a coil ignition system.

Key fact

Bosch created the first working magneto in 1897.

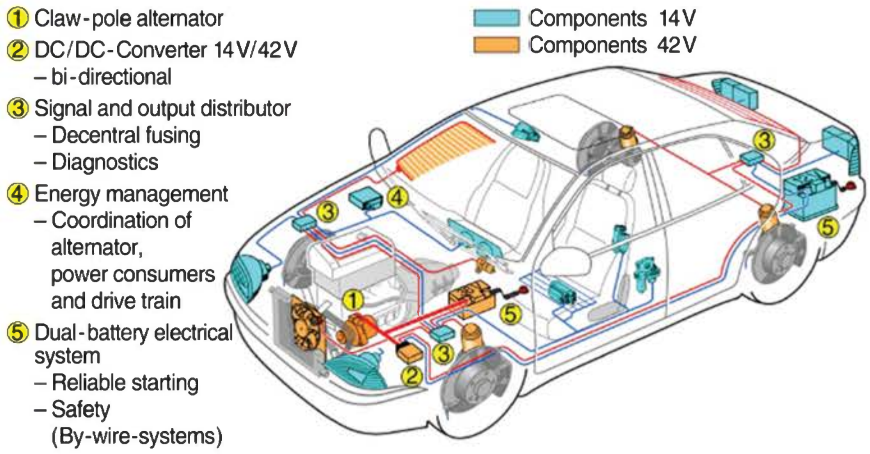


Figure 1.1 Future electronic systems (Source: Bosch Media)

Figure 1.2 1897 The De Dion-Bouton three-wheeler, with Bosch magneto

to almost universal acceptance in 1902. The 'H' shaped armature of the very earliest magneto is now used as the Bosch trademark on all the company's products. Bosch actually created the first working magneto in 1897.

From this period onwards, the magneto was developed to a very high standard in Europe, while in the USA the coil and battery ignition system took the lead. Charles F. Kettering played a vital role in this area working for the Daytona electrical company (Delco), when he devised the ignition, starting and lighting system for the 1912 Cadillac. Kettering also produced a mercury-type voltage regulator.

The third-brush dynamo, first produced by Dr Hans Leitner and R.H. Lucas, first appeared in about 1905. This gave the driver some control over the charging system. It became known as the constant current charging system. By today's standards this was a very large dynamo and could produce only about 8 A.

Many other techniques were tried over the next decade or so to solve the problem of controlling output on a constantly varying speed dynamo. Some novel control methods were used, some with more success than others. For example, a drive system, which would slip beyond a certain engine speed, was used with limited success, while one of my favourites had a hot wire in the main output line which, as it became red hot, caused current to bypass it and flow through a 'bucking' coil to reduce the dynamo field strength. Many variations of the 'field warp' technique were used. The control of battery charging current for all these constant current systems was poor and often relied on the driver to switch from high to low settings. In fact, one of the early forms of instrumentation was a dashboard hydrometer to check the battery state of charge!

The two-brush dynamo and compensated voltage control unit was used for the first time in the 1930s. This gave far superior control over the charging system and paved the way for the many other electrical systems to come.

In 1936, the much-talked about move to positive earth took place (in the UK mostly). Lucas played a major part in this change. It was done to allow reduced spark plug firing voltages and hence prolong electrode life – however, there is much debate over the reasons. It was also hoped to reduce corrosion between the battery terminals and other contact points around the car.

The 1950s was the era when lighting began to develop towards today's complex arrangements. Flashing indicators were replacing the semaphore arms and the

Figure 1.3 Magneto (Source: Bosch Media)

twin filament bulb allowed more suitable headlights to be made. The quartz halogen bulb, however, did not appear until the early 1970s.

Great improvements now started to take place with the fitting of essential items such as heaters, radios and even cigar lighters! Also in the 1960s and 1970s, many more optional extras became available, such as windscreen washers and two-speed wipers. Cadillac introduced full air conditioning and even a time switch for the headlights.

The negative earth system was re-introduced in 1965 with complete acceptance. This did, however, cause some teething problems, particularly with the growing DIY fitment of radios and other accessories. It was also good, of course, for the established auto-electrical trade!

Key fact

The 1970s hailed the era of fuel injection and electronic ignition.

The 1970s also hailed the era of fuel injection and electronic ignition. Instrumentation became far more complex and the dashboard layout was now an important area of design. Heated rear windows that worked were fitted as standard to some vehicles. The alternator, first used in the USA in the 1960s, became the norm by about 1974 in Britain.

The extra power available and the stable supply of the alternator was just what the electronics industry was waiting for and, in the 1980s, the electrical system of the vehicle changed beyond all recognition.

Key fact

The advances in micro-computing and associated technology have now made control of all vehicle functions possible by electrical means.

The advances in micro-computing and associated technology have now made control of all vehicle functions possible by electrical means. That is what the rest of this book is about, so read on.

1.1.2 A chronological history

The electrical and electronic systems of the motor vehicle are often the most feared, but at the same time can be the most fascinating aspects of an

Figure 1.4 A 1913 Bosch headlight

automobile. The complex circuits and systems now in use have developed in a very interesting way.

For many historical developments it is not possible to be certain exactly who ‘invented’ a particular component, or indeed when, as developments were taking place in parallel, as well as in series.

It is interesting to speculate on who we could call the founder of the vehicle electrical system. Michael Faraday of course deserves much acclaim, but then of course so does Ettiene Lenoir and so does Robert Bosch and so does Nikolaus Otto and so does …

Perhaps we should go back even further to the ancient Greek philosopher Thales of Miletus who, whilst rubbing amber with fur, discovered static electricity. The Greek word for amber is ‘elektron’.

Key fact

The Greek word for amber is ‘elektron’.

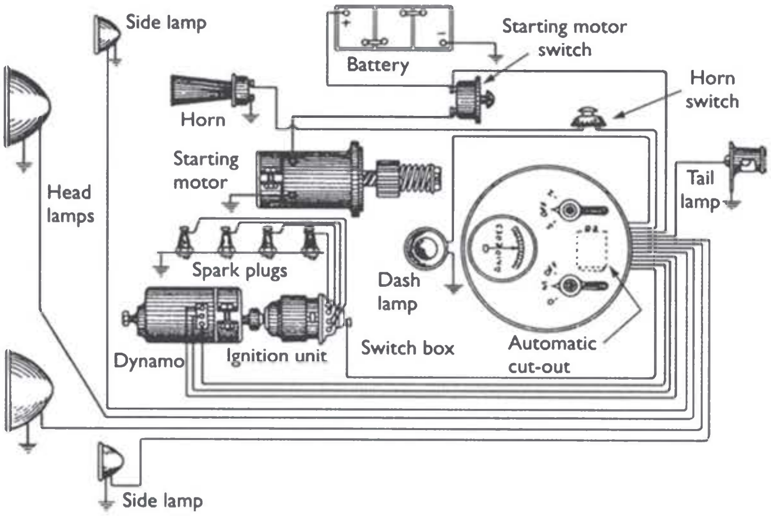


Figure 1.5 A complete circuit diagram

c600 BC Thales of Miletus discovers static electricity by rubbing amber with fur.

cAD1550 William Gilbert showed that many substances contain ‘electricity’ and that, of the two types of electricity he found different types attract while like types repel.

1672 Otto Von Guerick invented the first electrical device, a rotating ball of sulphur.

1742 Andreas Gordon constructed the first static generator.

1747 Benjamin Franklin flew a kite in a thunderstorm!

1769 Cugnot built a steam tractor in France made mostly from wood.

1780 Luigi Galvani started a chain of events resulting in the invention of the battery.

1800 The first battery was invented by Alessandro Volta.

1801 Trevithick built a steam coach.

1825 Electromagnetism was discovered by William Sturgeon.

1830 Sir Humphery Davy discovered that breaking a circuit causes a spark.

1831 Faraday discovered the principles of induction.

1851 Ruhmkorff produced the first induction coil.

1859 The accumulator was developed by the French physicist Gaston Planche.

1860 Lenoir built an internal-combustion gas engine.

1860 Lenoir developed ‘in cylinder’ combustion.

1860 Lenoir produced the first spark-plug.

1861 Lenoir produced a type of trembler coil ignition.

1861 Robert Bosch was born in Albeck near Ulm in Germany.

1870 Otto patented the four-stroke engine.

1875 A break spark system was used in the Seigfried Marcus engine.

1876 Otto improved the gas engine.

1879 Hot-tube ignition was developed by Leo Funk.

1885 Benz fitted his petrol engine to a three-wheeled carriage.

1885 The motor car engine was developed by Gottlieb Daimler and Karl Benz.

1886 Daimler fitted his engine to a four-wheeled carriage to produce a four-wheeled motorcar.

1887 The Bosch low-tension magneto was used for stationary gas engines.

1887 Hertz discovered radio waves.

1888 Professor Ayrton built the first experimental electric car.

1889 E. Martin used a mechanical system to show the word ‘STOP’ on a board at the rear of his car.

1889 Georges Bouton invented contact breakers.

1891 Panhard and Levassor started the present design of cars by putting the engine in the front.

1894 The first successful electric car.

1895 Emile Mors used accumulators that were recharged from a belt-driven dynamo.

1895 Georges Bouton refined the Lenoir trembler coil.

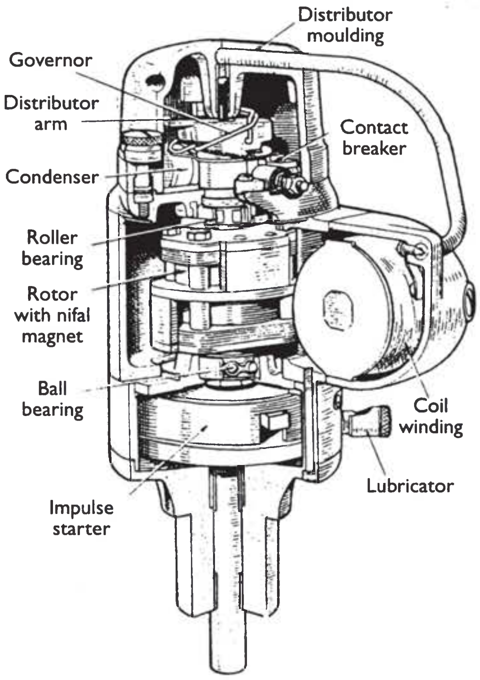


Figure 1.6 Sectional view of the Lucas type 6VRA Magneto

1896 Lanchester introduced epicyclic gearing, which is now used in automatic transmission.

1897 The first radio message was sent by Marconi.

1897 Bosch and Simms developed a low-tension magneto with the ‘H’ shaped armature, used for motor vehicle ignition.

1899 Jenatzy broke the 100 kph barrier in an electric car.

1899 First speedometer introduced (mechanical).

1899 World speed record 66 mph – in an electric powered vehicle!

1901 The first Mercedes took to the roads.

1901 Lanchester produced a flywheel magneto.

1902 Bosch introduced the high-tension magneto, which was almost universally accepted.

1904 Rigolly broke the 100 mph barrier.

1905 Miller Reese invented the electric horn.

1905 The third-brush dynamo was invented by Dr Hans Leitner and R.H. Lucas.

1906 Rolls-Royce introduced the Silver Ghost.

1908 Ford used an assembly-line production to manufacture the Model T.

1908 Electric lighting appeared, produced by C.A. Vandervell.

1910 The Delco prototype of the electric starter appeared.

1911 Cadillac introduced the electric starter and dynamo lighting.

1912 Bendix invented the method of engaging a starter with the flywheel.

1912 Electric starting and lighting used by Cadillac. This ‘Delco’ electrical system was developed by Charles F. Kettering.

1913 Ford introduced the moving conveyor belt to the assembly line.

1914 Bosch perfected the sleeve induction magneto.

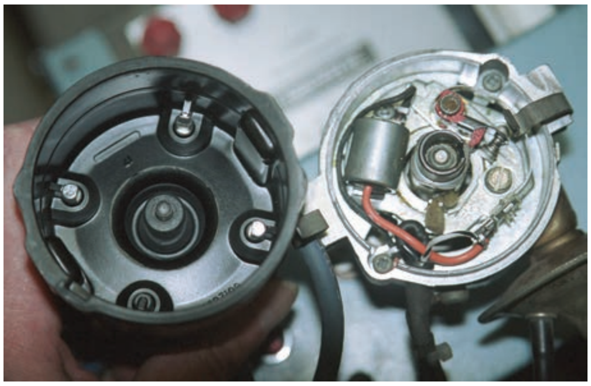


Figure 1.7 Distributor with contact breakers

1914 A buffer spring was added to starters.

1920 Duesenberg began fitting four-wheel hydraulic brakes.

1920 The Japanese made significant improvements to magnet technology.

1921 The first radio set was fitted in a car by the South Wales Wireless Society.

1922 Lancia used a unitary (all-in-one) chassis construction and independent front suspension.

1922 The Austin Seven was produced.

1925 Dr D.E. Watson developed efficient magnets for vehicle use.

1927 Segrave broke the 200 mph barrier in a Sunbeam.

1927 The last Ford model T was produced.

1928 Cadillac introduced the synchromesh gearbox.

1928 The idea for a society of engineers specializing in the auto-electrical trade was born in Huddersfield, Yorkshire, UK.

1929 The Lucas electric horn was introduced.

1930 Battery coil ignition begins to supersede magneto ignition.

1930 Magnet technologies are further improved.

1931 Smiths introduced the electric fuel gauge.

1931 The Vertex magneto was introduced.

1932 The Society of Automotive Electrical Engineers held its first meeting in the Constitutional Club, Hammersmith, London, 21 October at 3.30 pm.

1934 Citroën pioneered front-wheel drive in their 7CV model.

1934 The two-brush dynamo and compensated voltage control unit was first fitted.

1936 An electric speedometer was used that consisted of an AC generator and voltmeter.

1936 Positive earth was introduced to prolong spark-plug life and reduce battery corrosion.

1937 Coloured wires were used for the first time.

1938 Germany produced the Volkswagen Beetle.

1939 Automatic advance was fitted to ignition distributors.

1939 Car radios were banned in Britain for security reasons.

1939 Fuse boxes start to be fitted.

Figure 1.8 Thrust SSC

1939 Tachograph recorders were first used in Germany.

1940 The DC speedometer was used, as were a synchronous rotor and trip meter.

1946 Radiomobile company formed.

1947 The transistor was invented.

1948 Jaguar launched the XK120 sports car and Michelin introduced a radial-ply tyre.

1948 UK manufacturers start to use 12 V electrical system.

1950 Dunlop announced the disc brake.

1951 Buick and Chrysler introduced power steering.

1951 Development of petrol injection by Bosch.

1952 Rover's gas-turbine car set a speed record of 243 kph.

1954 Bosch introduced fuel injection for cars.

1954 Flashing indicators were legalized.

1955 Citroën introduced a car with hydro-pneumatic suspension.

1955 Key starting becomes a standard feature.

1957 Wankel built his first rotary petrol engine.

1957 Asymmetrical headlamps were introduced.

1958 The first integrated circuit was developed.

1959 BMC (now Rover Cars) introduced the Mini.

1960 Alternators started to replace the dynamo.

1963 The electronic flasher unit was developed.

1965 Development work started on electronic control of anti-locking braking system (ABS).

1965 Negative earth system reintroduced.

1966 California brought in legislation regarding air pollution by cars.

1966 In-car record players are not used with great success in Britain due to inferior suspension and poor roads!

1967 The Bosch Jetronic fuel injection system went into production.

1967 Electronic speedometer introduced.

1970 Gabelich drove a rocket-powered car, ‘Blue Flame’, to a new record speed of 1001.473 kph.

1970 Alternators began to appear in British vehicles as the dynamo began its demise.

1972 Dunlop introduced safety tyres, which seal themselves after a puncture.

1972 Lucas developed head-up instrumentation display.

1974 The first maintenance free breakerless electronic ignition was produced.

1976 Lambda oxygen sensors were produced.

1979 Barrett exceeded the speed of sound in the rocket-engined ‘Budweiser Rocket’ (1190.377 kph).

1979 Bosch started series production of the Motronic fuel injection system.

1980 The first mass-produced car with four-wheel drive, the Audi Quattro, was available.

1981 BMW introduced the on-board computer.

1981 Production of ABS for commercial vehicles started.

1983 Austin Rover introduced the Maestro, the first car with a talking dashboard.

1983 Richard Noble set an official speed record in the jet-engined ‘Thrust 2’ of 1019.4 kph.

1987 The solar-powered ‘Sunraycer’ travelled 3000 km.

1988 California’s emission controls aim for use of zero emission vehicles (ZEVs) by 1998.

1989 The Mitsubishi Gallant was the first mass-produced car with four-wheel steering.

1989 Alternators, approximately the size of early dynamos or even smaller, produced in excess of 100 A.

1990 Fiat of Italy and Peugeot of France launched electric cars.

1990 Fibre-optic systems used in Mercedes vehicles.

1991 The European Parliament voted to adopt stringent control of car emissions.

1991 Gas discharge headlamps were in production.

1992 Japanese companies developed an imaging system that views the road through a camera.

1993 A Japanese electric car reached a speed of 176 kph.

1993 Emission control regulations force even further development of engine management systems.

1994 Head-up vision enhancement systems were developed as part of the Prometheus project.

1995 Greenpeace designed an environmentally friendly car capable of doing 67–78 miles to the gallon (100 km per 3–3.5 litres).

1995 The first edition of *Automobile Electrical and Electronic Systems* was published!

1996 Further legislation on control of emissions.

1997 GM developed a number of its LeSabres for an Automated Highway System.

Figure 1.9 Ford Mustang

1998 Thrust SSC broke the sound barrier. 1998 Blue vision headlights started to be used.

1998 Mercedes ‘S’ class had 40 computers and over 100 motors.

1999 Mobile multimedia became an optional extra.

2000 Second edition of *Automobile Electrical and Electronic Systems* published!

2001 Global positioning systems start to become a popular optional extra.

2002 Full X-by-wire concept cars produced.

2003 Bosch celebrates 50 years of fuel injection.

2003 Ford develop the Hydrogen Internal Combustion Engine (H2ICE).

2004 Third edition of *Automobile Electrical and Electronic Systems* published!

2005 FreeScale Semiconductor paved the way for the autonomous car by becoming the first company to offer both integrated and stand-alone FlexRay(TM) controllers.

2006 More sensors such as yaw are integrated into a single control chip.

2007 Tesla Roadster EV first on sale.

2008 BMW’s safety and assistance telematics service, ConnectedDrive released in the UK.

2009 Experimentation with car platooning by Volvo and others as part of the SARTRE project.

2009 KERS first used in formula 1.

2010 Twin motor wipers go into production.

2011 This fourth edition of *Automobile Electrical and Electronic Systems* was published!

2012 The semantic web changes automotive training systems…

2013 F1 uses light hybrid engines…

20-- And the story continues with you…

1.2 Where next?

1.2.1 Current developments

Greater electronic control continues to be a key area of development on the car. However, an emerging technology is the networked car and the endless possibilities that this could bring – some good, some bad of course. Hybrid cars are now mainstream and full EVs will not be far behind.

An area that I speculate may become more significant is the use of satellite navigation systems for far more than getting you to a destination. One area of research is looking at adapting the engine operating characteristics and maximum speed – based on where the car is on the map. Scary!

The next three sections are reprints from the first three editions of this book where I speculated, to the point of fantasy, where the automotive electrical system would end up. Somewhat spookily many of the ideas I came up with are now available.

Section 1.2.4 dares to speculate even further but it is not a very long story…

1.2.2 Auto-electrical systems in the next millennium

(First published 1995)

Imagine what a vehicle will be like which is totally controlled by electronic systems. Imagine a vehicle with total on board diagnostic systems to pin-point any fault and the repairs required. Imagine a vehicle controlled by a 64 bit computer system with almost limitless memory. Imagine a vehicle with artificial intelligence to take all the operating decisions for you which also learns what you like and where you are likely to go. Finally imagine all of the above ideas combined with an automatic guidance system which works from cables laid under the road surface. Imagine what it would be like when it really went wrong!

However, picture this: Monday morning 15 January 2020, 08:00 hours. You are due at work by 09:00 which is just enough time to get there even though it is only fifteen miles away (the fourteen lane M25 soon filled to capacity), but at least access to the wire guided lane helps.

A shiver of cold as you walk from the door of your house through the layer of snow makes you glad you paid the extra for the XYZ version of the 'car'. As you would expect the windows of the car are already defrosted and as you touch the thumb print recognition padd and the door opens slowly a comforting waft of warm air hits you. It is still a little difficult to realise that the car anticipated that you would need it this morning and warmed the interior ready for your arrival.

Once the door is closed and the seat belts lift ready for you to snap into place a message appears on the windscreen. 'Good morning Tom', you find that a little irritating as usual, 'All systems are fully operational except the rear collision avoidance radar' (again). 'I have taken the liberty of switching to a first line back-up system and have made a booking with the workshop computer via the radio modem link.' You can't help but feel some control has been lost, but still it's one less thing for you to worry about. 'Shall we begin the journey, I have laid in a course for your work, is this correct?' Being able to speak to your car was odd at first but one soon gets used to these things. 'Yes', you say, and the journey begins.

It is always comforting to know that the tyre pressures and treads automatically adjust to the road and weather conditions. Even the suspension and steering is fine tuned. The temperature as usual is now just right, without you even having to touch a control. This is because the temperature and climate control system soon learned that you prefer to feel very warm when you first sit in the car but like to cool the temperature down as the journey progresses. A small adjustment to the humidity would seem to be in order so you tell the car. 'I will ensure I remember the change in future', appears on the screen.

Part way in to the journey the car slows down and takes a turning not part of your usual route to work. The car decides to override the block you placed on audio communication, as it knows you will be wondering what happened. 'Sorry about the change of route Tom but the road report transmission suggested this way would be quicker due to snow clearing.' 'We will still be at work on time.'

The rest of the journey is uneventful and as usual you spend time working on some papers but can't resist seeing if you can hear when the diesel engine takes over from the electric. It's very difficult though because the active noise reduction is so good these days.

The car arrives at your place of work and parks in its usual place. For a change you remember to take the control unit with you so the car doesn't have to remind you again. It's very good really though, as the car will not work without it and you can use it to tell the car when you need it next and so on. The car can also contact you if for example unauthorised entry is attempted.

Finally one touch on the outside control padd and the doors close and lock setting the alarm system at the same time.

While you are at work the car runs its fifth full diagnostic check of the day and finds no further faults. The sodium batteries need topping up so the car sets a magnetic induction link with the underground transformer and the batteries are soon fully charged.

The car now drops into standby mode after having set the time to start preparing for your journey home which it has learnt has an 85% probability of being via the local pub...

Key fact

The car now drops into standby mode after having set the time to start preparing for your journey home which it has learnt has an 85% probability of being via the local pub...

1.2.3 Automobile systems in the next millennium – 'The modern driver'

(First published 2000)

The best thing about wire guided car systems is that you can still take control from time to time! I often do some work as my car is taking me on a journey, which is good. Today though – is a day off.

'Please enter details of your journey,' said the car in its uncannily human voice. The voice can be adjusted, but it's even worse when it sounds like Professor Stephen Hawking. 'I'm going to drive for a change' I told it, but as usual it persisted, 'Would you like me to plan the best route?' 'No!' I said, 'not today.' 'All diagnostic routines have been run during the night and no faults found,' it continued. At this point I told it quite succinctly, not to speak again unless in an emergency. After accessing its 'colloquial database', it appeared to understand – and stopped talking!

Today I wanted to really drive. Pulling out of the garage I set off towards my favourite test track. It was the proactive suspension that I wanted to put through its paces. As well as the obvious surface scanning lasers, the new system uses magnetoelastic springs. This system could, in theory, not only change the suspension stiffness on each wheel instantly, it could also change the damping characteristics. We will see!

As usual I tried to feel when the electric motor cut out and the turbine cut in, but as usual, I couldn't. The high performance electromechanical torque storage system made sure of that.

Passing other cars on the road reminded me of my first time driving with a joystick instead of a steering wheel, it was weird, too much like a three-dimensional computer simulation. However, now I am used to it I don't think I could go back!

I was about half way to the test track, according to the guidance system, when the unthinkable happened - the car stopped. 'What's going on' I demanded, and, as the car had interpreted this event as an emergency, it answered, 'An unknown system error has occurred, please wait for further details.' I explained that it should proceed with all haste. Again the 'colloquial database' must have been useful because it said 'Accessing at maximum speed, please be patient.'

Three minutes later the system stated up again like nothing had happened. 'All systems fully functional using first line backups' the car announced, with what could only be described as a little pride in its artificial voice. 'What was wrong?' I asked, which seemed like a reasonable question at the time. 'A comparative run time error occurred in the second parallel processor line due to an incorrect digital signal response from the main sensor area network data bus responsible for critical system monitoring,' the car replied. 'You mean a wire fell off' I said. 'Yes' it admitted after consulting its 'concise lexicographical response database'. I think it's about time somebody invented a system that could bypass faults to repair itself, without having to stop the car. That three minutes could have been important!

At last I reached the test track and switched the car into full sports mode. 'All vehicle control systems adjusted to optimum settings for test track seven,' the car told me. Test track seven is great for putting the car through its paces. It has banked corners, 'S' bends, cobbled surface sections and even a water splash. There were only a few other drivers on the track so today was going to be the day.

I pulled out on to the track and floored the pedal. The car took off like crazy with the traction control allowing just enough wheel spin to gain maximum possible acceleration. The active steering felt great on the first corner; I could feel it fighting the tendency to oversteer by adjusting the four-wheel steering as well as diverting drive from one wheel to another. Plunging into the water trap at full speed nearly fooled the steering – but not quite. The wipers even switched on just before the water hit the screen. As I accelerated out of the 'S' bends another car pulled out of a side lane right in front of me – I noticed just in time. I hit the brakes as hard as I could and the ABS stopped me in plenty of time. Off I went again, this time on to the cobbled section although it didn't feel any different to the rest of the smooth track. I was just about to tell the car to check its magnetoelastic suspension system, when I realised that it must have been working! Just as I was about to finish my first lap, the head up display flashed 'Automatic Overtake?' in front of me. 'Go for it!' I shouted and the car overtook the one in front like it was standing still. This was a great day for driving.

On the way home, as usual, the car had predicted that I would be going via the local pub and had set a route accordingly. I parked the car, well it parked itself really, in the inductive recharge slot, and I went in for a well-earned drink. I couldn't wait to tell my friends what real driving was all about.

Key fact

I couldn't wait to tell my friends what real driving was all about.

1.2.4 An eye on the future

(First published 2004)

Evidently, my new car, which is due to arrive later today, has a digital camera that will watch my eyes. Something to do with stopping me from falling asleep I think. However, unless it pokes me in the eye with a sharp stick, it has its work cut out! Anyway, it seems like a pointless system in a car that drives itself most of the time. I can't wait for my new car to arrive.

The thing is I intend to spend as much time sleeping in my car as possible, well, when travelling long distances anyway. The whole point of paying the extra money for the 'Professional' instead of the 'Home' edition of the onboard software was so I could sleep or at least work on long journeys. The fully integrated satellite broadband connection impressed me too. The global positioning system is supposed to be so accurate you can even use it for parking in a tight spot. Not that you need it to, because the auto park and recharge was good even on my old car. The data transfer rate, up to or down from the satellite, is blistering – or so the 3D sales brochure said anyway. This means I will be able to watch the latest HoloVids when travelling if I'm not working or sleeping. It will even be useful for getting data to help with my work as a writer. Thing is though, the maximum size of most Macrosoft HoloWord documents is only about 4Tb. A Terabyte is only a million Megabytes so I won't be using even half of the available bandwidth. I hope my new car arrives soon.

I still like my existing car but it has broken down on a number of occasions. In my opinion three breakdowns in two years is not acceptable. And, on the third occasion, it took the car almost four and a half minutes to fix itself. I have come to expect a better level of service than that. I do hope however, that the magnetic gas suspension is as good as the MagnetoElastic system that I have become used to.

It took me a long time to decide whether to go for the hybrid engine or to go fully electric. I decided in the end that as the range of the batteries was now over two hundred miles, it would be worth the chance. After all, the tax breaks for a zero emission car are considerable.

I will still take my new car down to the test track because it is so much fun but this time I have really gone for comfort rather than performance. Still a 0 to 60 time of six seconds is not bad for a big comfortable, electric powered family car. The gadget I am going to enjoy most is the intelligent seat adjustment system. Naturally, the system will remember and adjust to previous settings when I unlock the car (and it recognises me of course). However, the new system even senses tension or changes in your body as you sit down and makes appropriate adjustments to the seat. Subtle temperature changes and massage all take place without you even saying anything. I can't wait much longer. Why isn't the car here yet?

My previous voice control system was good but a bit slow at times. It had to use its colloquial database every time I got mad with it and its built in intelligence was a bit limited. The new system is supposed to be so smart that it even knows when to argue with the driver. This will be useful for when I decide to override the guidance system, as I have done a number of occasions, and ended up getting lost every time. Well not really lost because when I let the car take over again, we got back on the route within ten minutes, but you know what I mean.

I'm also looking forward to using the computer enhanced vision system. Not that I will need to see where I'm going most of the time but it will be fun being able to look into other people's cars when they think I can't see. I wonder how well the recording facility works.

Having a multi-flavour drinks dispenser will be nice but unfortunately, it doesn't fill itself up so if it runs out between services I will have to learn how to fill the water tank. I hope that improves for the next model.

Servicing the new car is going to be much easier. Evidently, all you have to do is take the car to the local service centre (or send it on its own) and they change the complete powertrain system for a new one. Apparently, it is cheaper to import new fully integrated powertrain and chassis systems from overseas than it is for our technicians to repair or service the old ones! I expect it will take over an hour for this though so I will probably send the car during the night or when I am working at home. Surely the car should be here by now.

The most radical design aspect of my new car, if it ever arrives, is the ability to switch off every single driving aid and do it yourself! I can't wait to try this. However, I am led to believe that the insurance cover is void if you use the car on the 'Wired-Roads' (wi-ro for short). Evidently, the chance of having an accident increases a thousand fold when people start driving themselves. Still I'm going to try it at some point! Problem is over ninety eight percent of the roads are wi-ro now so I will have to take care. The few that aren't wi-ro have been taken over by that group of do-gooders the 'Friends of the Classic Car'. You know those people who still like to drive things like the ancient Mondeo or Escort. To be safe I will just use one of the test tracks. It's here, my new car it's here!

It was a bit weird watching it turn up in my garage with no driver but everything looks just fine. It was also a bit sad seeing my old car being towed away by the Recovery Drone but at least the data transfer to the new one went off without a problem. You know, I will miss my old car. Hey, is that an unlisted feature of my new car? I must check the ReadMe.HoloTxt file.

As I jumped in the car, the seat moved and it felt like it was adjusting itself to my inner soul – it was even better than I had hoped, it was just so comfortable. 'Welcome sir,' said the car, and it made me jump as it always does the first time! 'Hello' I replied after a moment, 'oh and please call me Tom.' No problem,' it answered without any noticeable delay. 'Would you like to go for a test drive Tom?' it asked after a short but carefully calculated delay. I liked its attitude so I said, 'Yes, let's go and see the boys down at the test track.' 'Would that be track five as usual Tom?' it continued. 'Yes!' I answered, a bit sharper than I had intended to, well, for this early in our relationship anyway. 'If you prefer, I will deactivate my intelligence subroutines or adjust them – you don't need to get cross with me!' 'I'm not cross,' I told it crossly, and then realised I was arguing with my car! 'Just take me to track five,' I told it firmly.

On the way, it was just so smooth and comfortable that I almost fell asleep. Still we got there, me and my new friend the car, in less than half an hour so that was good. This was it then, I uncovered the master driving aid control switch, keyed in my PIN and told it to deactivate all assistance systems, engage the steering stick and then leave it to me. I like my new car!

I set off round the track, slowly at first, because it felt so strange but it was just fantastic to be able to control the car myself. It was even possible to steer as well as speed up and slow down. Fantastic, yawn, awesome… However, I still, yawn, stretch, can't figure out why the car has cameras watching my eyes. I mean, yawn, I've only been driving for a few minutes and, yawn, I'm not sleepy at…

Ouch! What was that? It felt like a sharp stick.

Safety first

Do not poke sharp sticks in your eyes!

1.2.5 The death of the car – Energise?

(First published 2011)

Most of the time now I don't go to work; this is not because I don't have a car anymore, but because I don't really need to travel. Anything I can do at work I can do at home because of the 'uber-fast' Internet connection that allows full 3D conferencing and access to all the data in the world via the semantic web.

Of course when I do need a car I just call one up and it delivers itself within 10 or 15 minutes and, when I tell it my destination, off I go under full GPS guided control.

Mind you, when the new point-to-point particle transporter (PTPPT) system is finished even the GPS car will not be needed.

However, I am glad I kept my old BMW 5 series M-Sport to play in at the week-ends; it only has something like ten ECUs, three communication networks, GPRS connected systems, central ISO electronic control and throttle by wire so it is pretty easy to repair.

Ah, those were the days…

Definition

PTPPT: Point-to-point particle transporter.

Figure 1.10 Robert Bosch built his first magneto ignition device in 1887 for a mechanical engineer. The ignition unit, used in a stationary petrol/gasoline engine, aroused Robert Bosch's technical curiosity

Figure 1.11 Sony concept vehicle interior (Source: Visteon)

This page intentionally left blank

CHAPTER 2

Electrical and electronic principles

2.1 Safe working practices

2.1.1 Introduction

Safe working practices in relation to electrical and electronic systems are essential, for your safety as well as that of others. You only have to follow two rules to be safe.

- Use your common sense – don't fool about.
- If in doubt – seek help.

The following section lists some particular risks when working with electricity or electrical systems, together with suggestions for reducing them. This is known as risk assessment.

Safety first

The key to safe working: Common sense.

2.1.2 Risk assessment and reduction

Table 2.1 lists some identified risks involved with working on vehicles, in particular the electrical and electronic systems. The table is by no means exhaustive but serves as a good guide.

2.2 Basic electrical principles

2.2.1 Introduction

To understand electricity properly we must start by finding out what it really is. This means we must think very small (Figure 2.1 shows a representation of an atom). The molecule is the smallest part of matter that can be recognized as that particular matter. Sub-division of the molecule results in atoms, which are the smallest part of matter. An element is a substance that comprises atoms of one kind only.

The atom consists of a central nucleus made up of protons and neutrons. Around this nucleus orbit electrons, like planets around the sun. The neutron is a very small part of the nucleus. It has equal positive and negative charges and is therefore neutral and has no polarity. The proton is another small part of the nucleus, it is positively charged. The neutron is neutral and the proton

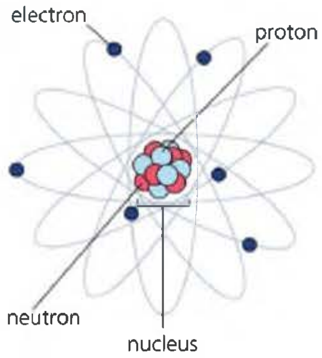


Figure 2.1 The atom

Table 2.1 Risks and risk reduction

Identified risk	Reducing the risk
Electric shock	Ignition HT is the most likely place to suffer a shock, up to 40 000 Volts is quite normal. Use insulated tools if it is necessary to work on HT circuits with the engine running. Note that high voltages are also present on circuits containing windings due to back emf as they are switched off, a few hundred Volts is common. Mains supplied power tools and their leads should be in good condition and using an earth leakage trip is highly recommended. Only work on HEV and EVs if training in the high voltage systems.
Battery acid	Sulphuric acid is corrosive so always use good PPE. In this case, overalls and if necessary rubber gloves. A rubber apron is ideal, as are goggles if working with batteries a lot.
Raising or lifting vehicles	Apply brakes and/or chock the wheels and when raising a vehicle on a jack or drive on lift. Only jack under substantial chassis and suspension structures. Use axle stands in case the jack fails.
Running engines	Do not wear loose clothing, good overalls are ideal. Keep the keys in your possession when working on an engine to prevent others starting it. Take extra care if working near running drive belts.
Exhaust gases	Suitable extraction must be used if the engine is running indoors. Remember it is not just the CO that might make you ill or even kill you, other exhaust components could cause asthma or even cancer.
Moving loads	Only lift what is comfortable for you; ask for help if necessary and/or use lifting equipment. As a general guide, do not lift on your own if it feels too heavy!
Short circuits	Use a jump lead with an in-line fuse to prevent damage due to a short when testing. Disconnect the battery (earth lead off first and back on last) if any danger of a short exists. A very high current can flow from a vehicle battery, it will burn you as well as the vehicle.
Fire	Do not smoke when working on a vehicle. Fuel leaks must be attended to immediately. Remember the triangle of fire – (Heat/Fuel/Oxygen) – don't let the three sides come together.
Skin problems	Use a good barrier cream and/or latex gloves. Wash skin and clothes regularly.

is positively charged, which means that the nucleus of the atom is positively charged. The electron is an even smaller part of the atom, and is negatively charged. It orbits the nucleus and is held in orbit by the attraction of the positively charged proton. All electrons are similar no matter what type of atom they come from.

When atoms are in a balanced state, the number of electrons orbiting the nucleus equals the number of protons. The atoms of some materials have electrons that are easily detached from the parent atom and can therefore join an adjacent atom. In so doing these atoms move an electron from the parent atom to another atom (like polarities repel) and so on through material. This is a random movement and the electrons involved are called free electrons.

Materials are called conductors if the electrons can move easily. In some materials it is extremely difficult to move electrons from their parent atoms. These materials are called insulators.

2.2.2 Electron flow and conventional flow

If an electrical pressure (electromotive force or voltage) is applied to a conductor, a directional movement of electrons will take place (for example when connecting a battery to a wire). This is because the electrons are attracted to the positive side and repelled from the negative side.

Certain conditions are necessary to cause an electron flow:

- A pressure source, e.g. from a battery or generator.
- A complete conducting path in which the electrons can move (e.g. wires).

Figure 2.2 Electronic components have made technology such as the 200+ km/h Tesla Roadster possible (Source: Tesla Motors)

An electron flow is termed an electric current. Figure 2.3 shows a simple electric circuit where the battery positive terminal is connected, through a switch and lamp, to the battery negative terminal. With the switch open the chemical energy of the battery will remove electrons from the positive terminal to the negative terminal via the battery. This leaves the positive terminal with fewer electrons and the negative terminal with a surplus of electrons. An electrical pressure therefore exists between the battery terminals.

With the switch closed, the surplus electrons at the negative terminal will flow through the lamp back to the electron-deficient positive terminal. The lamp will light and the chemical energy of the battery will keep the electrons moving in this circuit from negative to positive. This movement from negative to positive is called the electron flow and will continue whilst the battery supplies the pressure – in other words whilst it remains charged.

- Electron flow is from negative to positive.

It was once thought, however, that current flowed from positive to negative and this convention is still followed for most practical purposes. Therefore, although this current flow is not correct, the most important point is that we all follow the same convention.

- Conventional current flow is said to be from positive to negative.

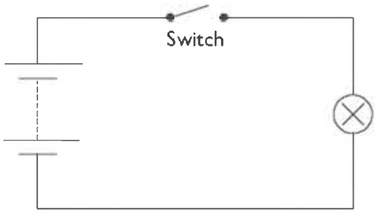


Figure 2.3 A simple electrical circuit

Key fact

Conventional current flow is said to be from positive to negative.

2.2.3 Effects of current flow

When a current flows in a circuit, it can produce only three effects:

- Heat
- Magnetism
- Chemical.

The heating effect is the basis of electrical components such as lights and heater plugs. The magnetic effect is the basis of relays and motors and generators. The chemical effect is the basis for electroplating and battery charging.

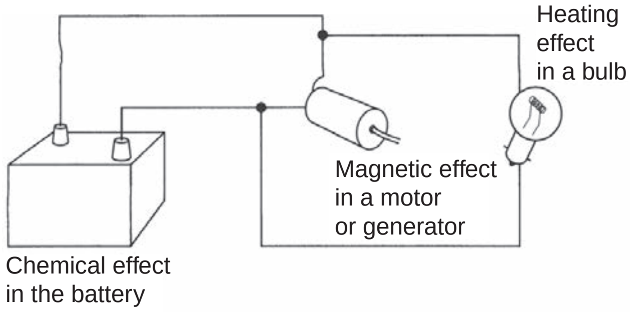


Figure 2.4 A bulb, motor and battery – heat, magnetic and chemical effects

In the circuit shown in Figure 2.4 the chemical energy of the battery is first converted to electrical energy, and then into heat energy in the lamp filament.

The three electrical effects are reversible. Heat applied to a thermocouple will cause a small electromotive force and therefore a small current to flow. Practical use of this is mainly in instruments. A coil of wire rotated in the field of a magnet will produce an electromotive force and can cause current to flow. This is the basis of a generator. Chemical action, such as in a battery, produces an electromotive force, which can cause current to flow.

Key fact

The three electrical effects are reversible.

2.2.4 Fundamental quantities

In Figure 2.5, the number of electrons through the lamp every second is described as the rate of flow. The cause of the electron flow is the electrical pressure. The lamp produces an opposition to the rate of flow set up by the electrical pressure. Power is the rate of doing work, or changing energy from one form to another. These quantities, as well as several others, are given names as shown in Table 2.2 on page 28.

If the voltage pressure applied to the circuit was increased but the lamp resistance stayed the same, then the current would also increase. If the voltage was maintained constant but the lamp was changed for one with a higher resistance the current would decrease. Ohm's Law describes this relationship.

Ohm's law states that in a closed circuit 'current is proportional to the voltage and inversely proportional to the resistance'. When 1 volt causes 1 ampere to flow the power used (P) is 1 watt.

Using symbols this means:

Voltage = Current × Resistance

($V = IR$) or ($R = V/I$) or ($I = V/R$)

Power = Voltage × Current

($P = VI$) or ($I = P/V$) or ($V = P/I$)

2.2.5 Describing electrical circuits

Three descriptive terms are useful when discussing electrical circuits.

- **Open circuit.** This means the circuit is broken therefore no current can flow.
- **Short circuit.** This means that a fault has caused a wire to touch another conductor and the current uses this as an easier way to complete the circuit.
- **High resistance.** This means a part of the circuit has developed a high resistance (such as a dirty connection), which will reduce the amount of current that can flow.

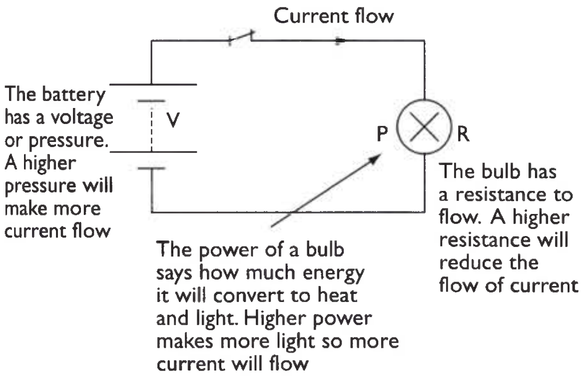


Figure 2.5 An electrical circuit demonstrating links between voltage, current, resistance and power

2.2.6 Conductors, insulators and semiconductors

All metals are conductors. Silver, copper and aluminium are among the best and are frequently used. Liquids that will conduct an electric current are called electrolytes. Insulators are generally non-metallic and include rubber, porcelain, glass, plastics, cotton, silk, wax paper and some liquids. Some materials can act as either insulators or conductors depending on conditions. These are called semiconductors and are used to make transistors and diodes.

2.2.7 Factors affecting the resistance of a conductor

In an insulator, a large voltage applied will produce a very small electron movement. In a conductor, a small voltage applied will produce a large electron flow or current. The amount of resistance offered by the conductor is determined by a number of factors (Figure 2.6).

- Length – the greater the length of a conductor the greater is the resistance.
- Cross-sectional area (CSA) – the larger the cross-sectional area the smaller the resistance.
- The material from which the conductor is made – the resistance offered by a conductor will vary according to the material from which it is made. This is known as the resistivity or specific resistance of the material.
- Temperature – most metals increase in resistance as temperature increases.

2.2.8 Resistors and circuit networks

Good conductors are used to carry the current with minimum voltage loss due to their low resistance. Resistors are used to control the current flow in a circuit or to set voltage levels. They are made of materials that have a high resistance. Resistors intended to carry low currents are often made of carbon. Resistors for high currents are usually wire wound.

Resistors are often shown as part of basic electrical circuits to explain the principles involved. The circuits shown as Figure 2.7 are equivalent. In other words, the circuit just showing resistors is used to represent the other circuit.

When resistors are connected so that there is only one path (Figure 2.8), for the same current to flow through each bulb they are connected in series and the following rules apply.

Key fact

Resistors are used to control the current flow in a circuit or to set voltage levels.

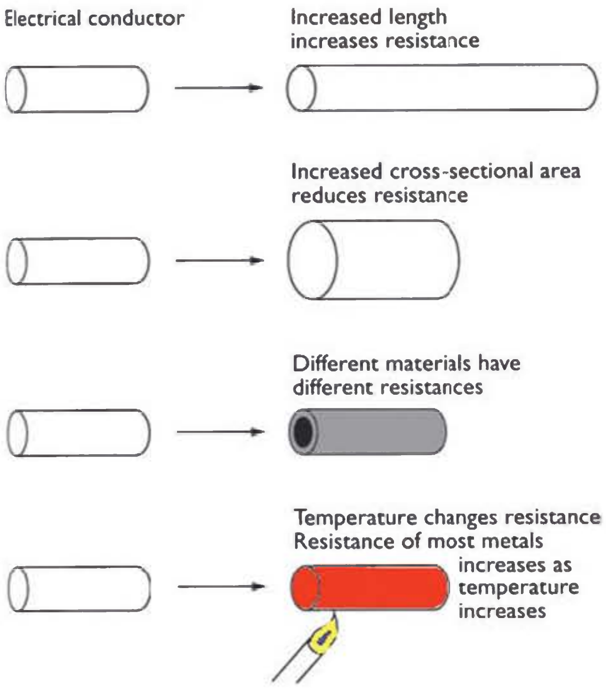


Figure 2.6 Factors affecting electrical resistance

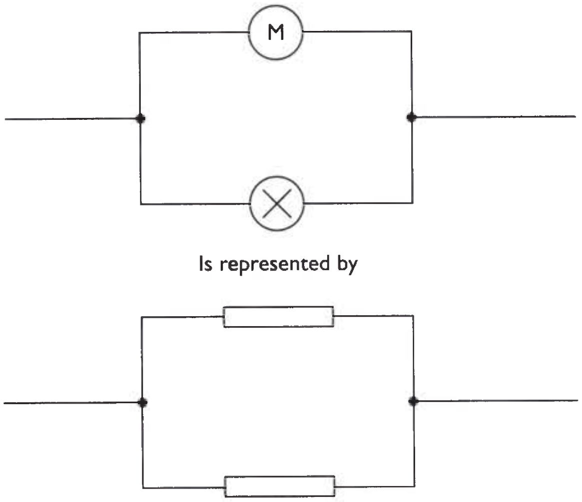


Figure 2.7 An equivalent circuit

Figure 2.8 Series circuit

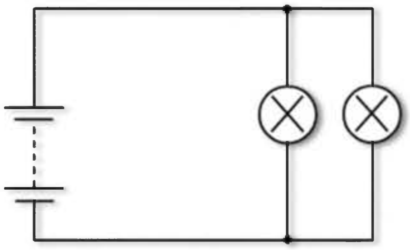


Figure 2.9 Parallel circuit

- Current is the same in all parts of the circuit.
- The applied voltage equals the sum of the volt drops around the circuit.
- Total resistance of the circuit (RT) equals the sum of the individual resistance values ($R_1 + R_2$ etc.).

When resistors or bulbs are connected such that they provide more than one path (Figure 2.9 shows two paths) for the current to flow through and have the same voltage across each component they are connected in parallel and the following rules apply.

- The voltage across all components of a parallel circuit is the same.
- The total current equals the sum of the current flowing in each branch.

- The current splits up depending on each component resistance.
- The total resistance of the circuit (RT) can be calculated by

$1/R_T = 1/R_1 + 1/R_2$ or
$R_T = (R_1 \times R_2)/(R_1 + R_2)$

2.2.9 Magnetism and electromagnetism

Magnetism can be created by a permanent magnet or by an electromagnet (it is one of the three effects of electricity remember). The space around a magnet in which the magnetic effect can be detected is called the magnetic field. The shape of magnetic fields in diagrams is represented by flux lines or lines of force.

Some rules about magnetism:

- Unlike poles attract. Like poles repel.
- Lines of force in the same direction repel sideways, in the opposite direction they attract.
- Current flowing in a conductor will set up a magnetic field around the conductor. The strength of the magnetic field is determined by how much current is flowing.
- If a conductor is wound into a coil or solenoid, the resulting magnetism is the same as a permanent bar magnet.

Electromagnets are used in motors, relays and fuel injectors, to name just a few applications. Force on a current-carrying conductor in a magnetic field is caused because of two magnetic fields interacting. This is the basic principle of how a motor works. Figure 2.10 shows a representation of these magnetic fields.

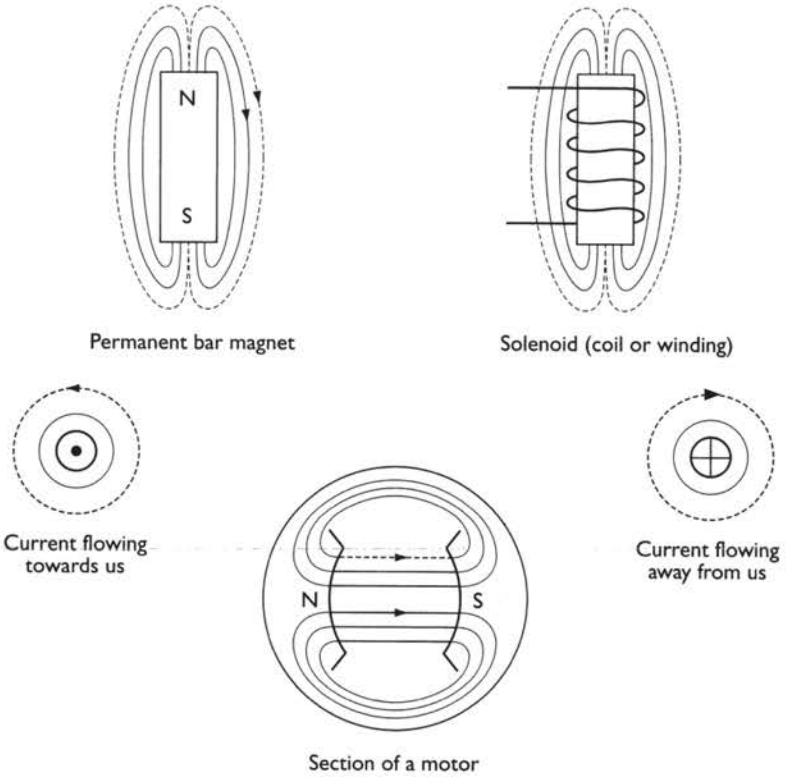


Figure 2.10 Magnetic fields

2.2.10 Electromagnetic induction

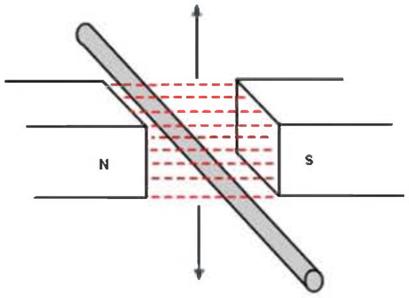


Figure 2.11 Induction

Basic laws:

- When a conductor cuts or is cut by magnetism, a voltage is induced in the conductor.
- The direction of the induced voltage depends upon the direction of the magnetic field and the direction in which the field moves relative to the conductor.
- The voltage level is proportional to the rate at which the conductor cuts or is cut by the magnetism.

This effect of induction, meaning that voltage is made in the wire, is the basic principle of how generators such as the alternator on a car work. A generator is a machine that converts mechanical energy into electrical energy. Figure 2.11 shows a wire moving in a magnetic field.

Definition

A generator is a machine that converts mechanical energy into electrical energy.

2.2.11 Mutual induction

If two coils (known as the primary and secondary) are wound on to the same iron core then any change in magnetism of one coil will induce a voltage in to the other. This happens when a current to the primary coil is switched on and off. If the number of turns of wire on the secondary coil is more than the primary, a higher voltage can be produced. If the number of turns of wire on the secondary coil is less than the primary a lower voltage is obtained. This is called ‘transformer action’ and is the principle of the ignition coil. Figure 2.12 shows the principle of mutual induction. The value of this ‘mutually induced’ voltage depends on:

- The primary current.
- The turns ratio between primary and secondary coils.
- The speed at which the magnetism changes.

Key fact

Transformer action is the principle of the ignition coil.

2.2.12 Definitions and laws

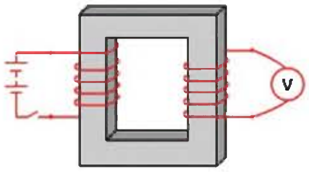


Figure 2.12 Mutual induction

Ohm’s law

- For most conductors, the current which will flow through them is directly proportional to the voltage applied to them.

The ratio of voltage to current is referred to as resistance. If this ratio remains constant over a wide range of voltages, the material is said to be ‘ohmic’.

$$V = I/R$$

where: I = Current in amps, V = Voltage in volts, R = Resistance in ohms.

Georg Simon Ohm was a German physicist, well known for his work on electrical currents.

Lenz’s law

- The emf induced in an electric circuit always acts in a direction so that the current it creates around the circuit will oppose the change in magnetic flux which caused it.

Lenz’s law gives the direction of the induced emf resulting from electromagnetic induction. The ‘opposing’ emf is often described as a ‘back emf’.

The law is named after the Estonian physicist Heinrich Lenz.

Kirchhoff's laws

Kirchhoff's 1st law:

- The current flowing into a junction in a circuit must equal the current flowing out of the junction.

This law is a direct result of the conservation of charge; no charge can be lost in the junction, so any charge that flows in must also flow out.

Kirchhoff's 2nd law:

- For any closed loop path around a circuit the sum of the voltage gains and drops always equals zero.

This is effectively the same as the series circuit statement that the sum of all the voltage drops will always equal the supply voltage.

Gustav Robert Kirchhoff was a German physicist; he also discovered caesium and rubidium.

Faraday's law

- Any change in the magnetic field around a coil of wire will cause an emf (voltage) to be induced in the coil.

It is important to note here that no matter how the change is produced, the voltage will be generated. In other words, the change could be produced by changing the magnetic field strength, moving the magnetic field towards or away from the coil, moving the coil in or out of the magnetic field, rotating the coil relative to the magnetic field and so on!

Michael Faraday was a British physicist and chemist, well known for his discoveries of electromagnetic induction and of the laws of electrolysis.

Fleming's rules

- In an electrical machine, the First Finger lines up with the magnetic Field, the seCond finger lines up with the Current and the thuMb lines up with the Motion.

Fleming's rules relate to the direction of the magnetic field, motion and current in electrical machines. The left hand is used for motors, and the right hand for generators (remember gener-righters).

The English physicist John Fleming devised these rules.

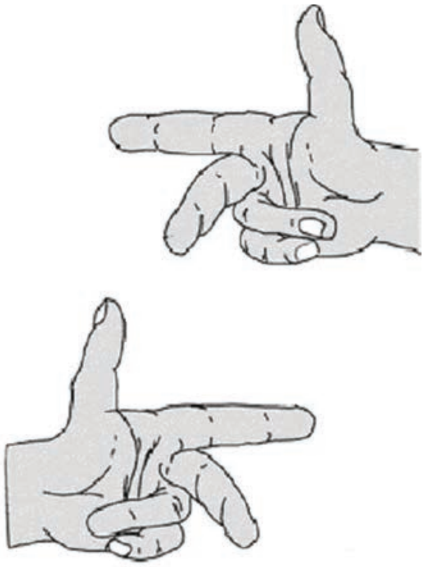


Figure 2.13 Fleming's rules

Ampere's law

- For any closed loop path, the sum of the length elements times the magnetic field in the direction of the elements is equal to the permeability times the electric current enclosed in the loop.

In other words, the magnetic field around an electric current is proportional to the electric current which creates it and the electric field is proportional to the charge which creates it.

André Marie Ampère was a French scientist, known for his significant contributions to the study of electrodynamics.

Summary

It was tempting to conclude this section by stating some of Murphy's laws, for example:

- If anything can go wrong, it will go wrong …
- You will always find something in the last place you look …

- In a traffic jam, the lane on the motorway that you are not in always goes faster …

… but I decided against it!

Table 2.2 Quantities, symbols and units

Name	Definition	Symbol	Common formula	Unit name	Abbreviation
Electrical charge	One coulomb is the quantity of electricity conveyed by a current of one ampere in one second.	Q	$Q = It$	coulomb	C
Electrical flow or current	The number of electrons past a fixed point in one second	I	$I = V/R$	ampere	A
Electrical pressure	A pressure of 1 volt applied to a circuit will produce a current flow of 1 amp if the circuit resistance is 1 ohm.	V	$V = IR$	volt	V
Electrical resistance	This is the opposition to current flow in a material or circuit when a voltage is applied across it.	R	$R = V/I$	ohm	Ω
Electrical conductance	Ability of a material to carry an electrical current. One siemens equals one ampere per volt. It was formerly called the mho or reciprocal ohm.	G	$G = 1/R$	siemens	S
Current density	The current per unit area. This is useful for calculating the required conductor cross sectional areas	J	$J = I/A$ (A = area)		A m^{-2}
Resistivity	A measure of the ability of a material to resist the flow of an electric current. It is numerically equal to the resistance of a sample of unit length and unit cross-sectional area, and its unit is the ohmmeter. A good conductor has a low resistivity (1.7×10^{-8} Ω m copper); an insulator has a high resistivity (10^{15} Ω m polyethane)	ρ (rho)	$R = \rho L/A$ (L = length A = area)	ohm meter	Ω m
Conductivity	The reciprocal of resistivity	σ (sigma)	$\sigma = 1/\rho$	ohm^{-1} meter^{-1}	$\Omega^{-1}\,\text{m}^{-1}$
Electrical power	When a voltage of 1 volt causes a current of 1 amp to flow the power developed is 1 watt.	P	$P = IV$ $P = I^2R$ $P = V^2/R$	watt	W
Capacitance	Property of a capacitor that determines how much charge can be stored in it for a given potential difference between its terminals	C	$C = Q/V$ $C = \varepsilon A/d$ (A = plate area, d = distance between, ε = permittivity of dielectric)	farad	F
Inductance	Where a changing current in a circuit builds up a magnetic field which induces an electromotive force either in the same circuit and opposing the current (self-inductance) or in another circuit (mutual inductance)	L	$i = \frac{V}{R}\left(1 - e^{-Rt/L}\right)$ (i = instantaneous current, R = resistance, L = inductance, t = time, e = base of natural logs)	henry	H

(Continued)

Table 2.2 *(Continued)*

Name	Definition	Symbol	Common formula	Unit name	Abbreviation
Magnetic field strength or intensity	Magnetic field strength is one of two ways that the intensity of a magnetic field can be expressed. A distinction is made between magnetic field strength H and magnetic flux density B.	H	$H = B/\mu_0$ (μ_0 being the magnetic permeability of space)	amperes per meter	A/m (An older unit for magnetic field strength is the oersted: 1 A/m = 0.01257 oersted)
Magnetic flux	A measure of the strength of a magnetic field over a given area.	Φ (phi)	$\Phi = \mu HA$ (μ = magnetic permeability, H = magnetic field intensity, A = area)	weber	Wb
Magnetic flux density	The density of magnetic flux, one tesla is equal to one weber per square meter. Also measured in Newton-meters per ampere (Nm/A)	B	$B = H/A$ $B = H \times \mu$ (μ = magnetic permeability of the substance, A = area)	tesla	T

2.3 Electronic components and circuits

2.3.1 Introduction

This section, describing the principles and applications of various electronic circuits, is not intended to explain their detailed operation. The intention is to describe briefly how the circuits work and, more importantly, how and where they may be utilized in vehicle applications.

The circuits described are examples of those used and many pure electronics books are available for further details. Overall, an understanding of basic electronic principles will help to show how electronic control units work, ranging from a simple interior light delay unit, to the most complicated engine management system.

2.3.2 Components

The main devices described here are often known as discrete components. Figure 2.14 shows the symbols used for constructing the circuits shown later in this section. A simple and brief description follows for many of the components shown.

Resistors are probably the most widely used component in electronic circuits. Two factors must be considered when choosing a suitable resistor, namely the ohms value and the power rating. Resistors are used to limit current flow and provide fixed voltage drops. Most resistors used in electronic circuits are made from small carbon rods, and the size of the rod determines the resistance. Carbon resistors have a negative temperature coefficient (NTC) and this must be considered for some applications. Thin film resistors have more stable temperature properties and are constructed by depositing a layer of carbon onto an insulated former such as glass. The resistance value can be manufactured very accurately by spiral grooves cut into the carbon film. For higher power applications, resistors are usually wire wound. This can, however, introduce

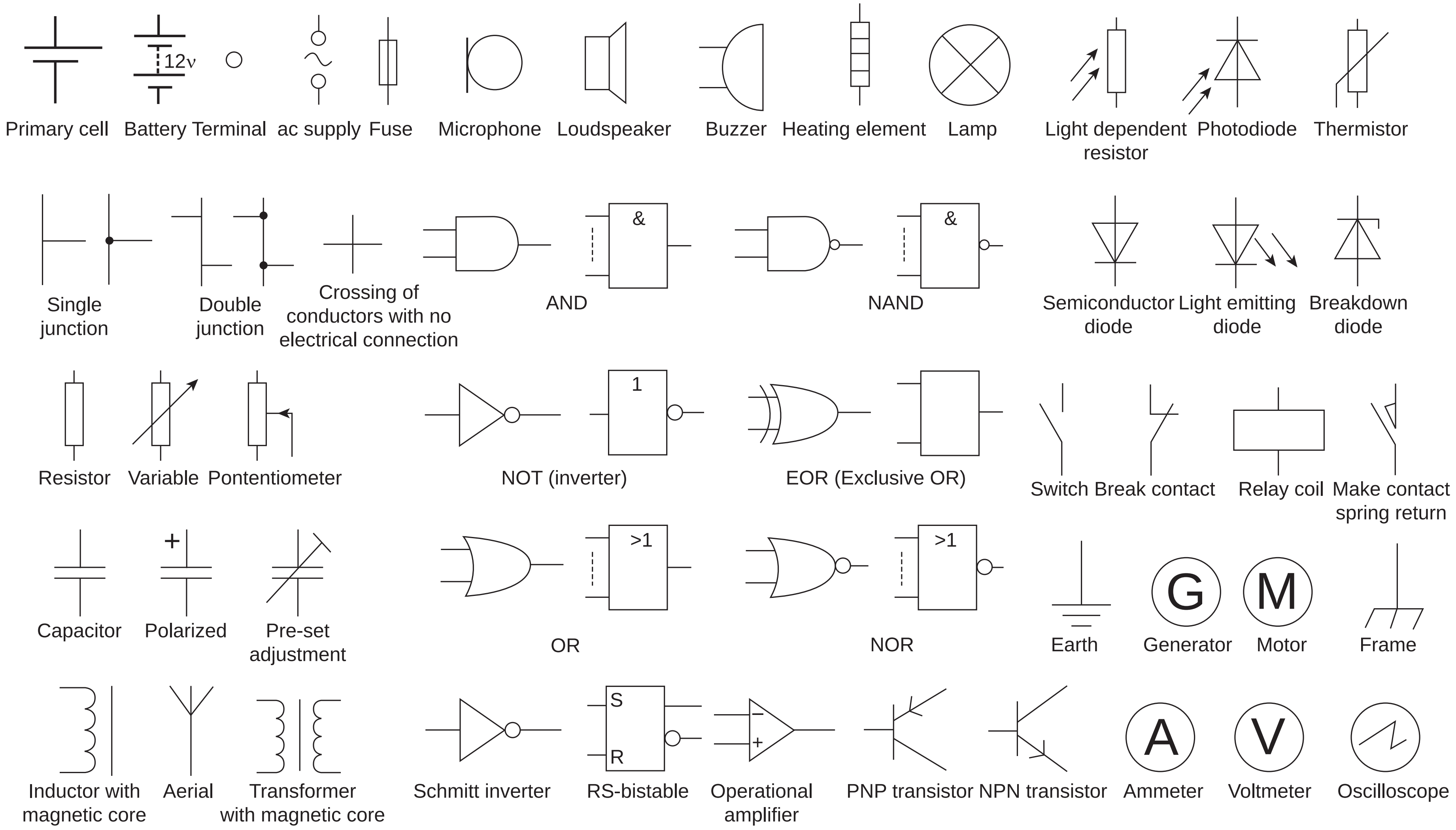


Figure 2.14 Circuit symbols

inductance into a circuit. Variable forms of most resistors are available in either linear or logarithmic forms. The resistance of a circuit is its opposition to current flow.

A capacitor is a device for storing an electric charge. In its simple form it consists of two plates separated by an insulating material. One plate can have excess electrons compared to the other. On vehicles, its main uses are for reducing arcing across contacts and for radio interference suppression circuits as well as in electronic control units. Capacitors are described as two plates separated by a dielectric. The area of the plates A, the distance between them d, and the permittivity (ε), of the dielectric, determine the value of capacitance. This is modelled by the equation:

$$C = \varepsilon A/d$$

Metal foil sheets insulated by a type of paper are often used to construct capacitors. The sheets are rolled up together inside a tin can. To achieve higher values of capacitance it is necessary to reduce the distance between the plates in order to keep the overall size of the device manageable. This is achieved by immersing one plate in an electrolyte to deposit a layer of oxide typically 104 mm thick, thus ensuring a higher capacitance value. The problem, however, is that this now makes the device polarity conscious and only able to withstand low voltages. Variable capacitors are available that are varied by changing either of the variables given in the previous equation. The unit of capacitance is the farad (F). A circuit has a capacitance of one farad (1 F) when the charge stored is one coulomb and the potential difference is 1 V. Figure 2.15 shows a capacitor charged up from a battery.

Diodes are often described as one-way valves and, for most applications, this is an acceptable description. A diode is a simple PN junction allowing electron flow from the N-type material (negatively biased) to the P-type material (positively biased). The materials are usually constructed from doped silicon. Diodes are not perfect devices and a voltage of about 0.6 V is required to switch the diode on in its forward biased direction. Zener diodes are very similar in operation, with the exception that they are designed to breakdown and conduct in the reverse

Definition

Negative temperature coefficient (NTC): As temperature increases, resistance decreases.

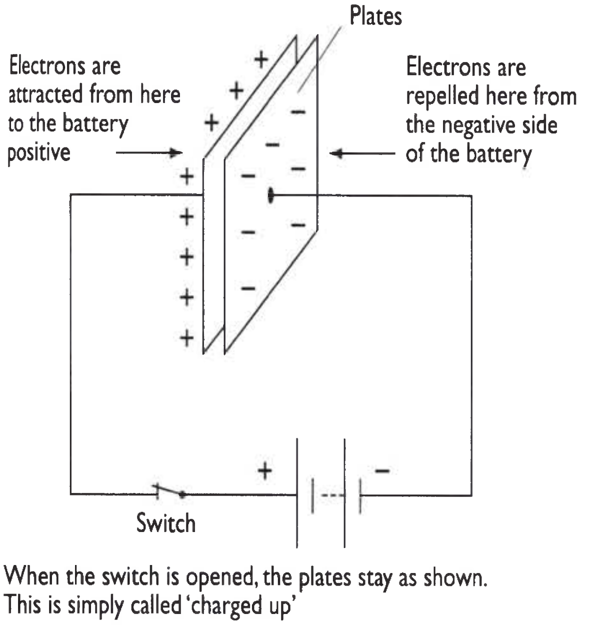


Figure 2.15 A capacitor charged up

direction at a pre-determined voltage. They can be thought of as a type of pressure relief valve.

Transistors are the devices that have allowed the development of today's complex and small electronic systems. They replaced the thermal-type valves. The transistor is used as either a solid-state switch or as an amplifier. Transistors are constructed from the same P- and N-type semiconductor materials as the diodes, and can be either made in NPN or PNP format. The three terminals are known as the base, collector and emitter. When the base is supplied with the correct bias the circuit between the collector and emitter will conduct. The base current can be of the order of 200 times less than the emitter current. The ratio of the current flowing through the base compared with the current through the emitter (I_e/I_b), is an indication of the amplification factor of the device and is often given the symbol.

Another type of transistor is the FET or field effect transistor. This device has higher input impedance than the bipolar type described above. FETs are constructed in their basic form as n-channel or p-channel devices. The three terminals are known as the gate, source and drain. The voltage on the gate terminal controls the conductance of the circuit between the drain and the source.

A further and important development in transistor technology is the insulate gate bipolar transistor (IGBT). The insulated gate bipolar transistor (Figure 2.16) is a three-terminal power semiconductor device, noted for high efficiency and fast switching. It switches electric power in many modern appliances: electric cars, trains, variable speed refrigerators, air-conditioners and even stereo systems with switching amplifiers. Since it is designed to rapidly turn on and off, amplifiers that use it often synthesize complex waveforms with pulse width modulation and low-pass filters.

Inductors are most often used as part of an oscillator or amplifier circuit. In these applications, it is essential for the inductor to be stable and to be of reasonable size. The basic construction of an inductor is a coil of wire wound on a former. It is the magnetic effect of the changes in current flow that gives this device the properties of inductance. Inductance is a difficult property to control, particularly as the inductance value increases due to magnetic coupling with other devices. Enclosing the coil in a can will reduce this, but eddy currents are then induced in the can and this affects the overall inductance value. Iron cores are used to increase the inductance value as this changes the permeability of the core. However, this also allows for adjustable devices by moving the position of the core. This only allows the value to change by a few per cent but is useful for tuning a circuit. Inductors, particularly of higher values, are often known as chokes and may be used in DC circuits to smooth the voltage. The value of

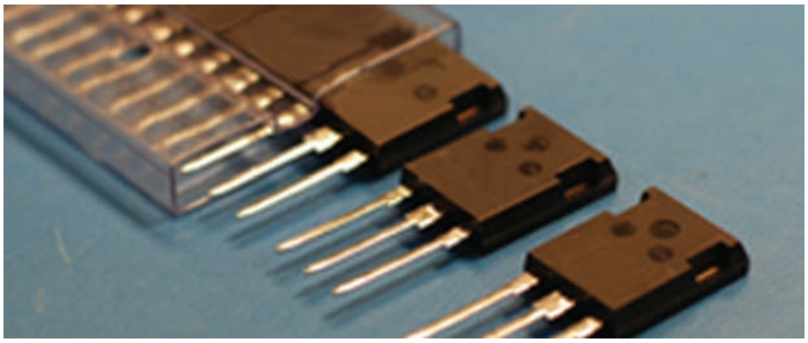


Figure 2.16 IGBT packages (Source: Telsa Motors)

inductance is the henry (H). A circuit has an inductance of one henry (1 H) when a current, which is changing at one ampere per second, induces an electromotive force of one volt in it.

2.3.3 Integrated circuits

Integrated circuits (ICs) are constructed on a single slice of silicon often known as a substrate. In an IC, Some of the components mentioned previously can be combined to carry out various tasks such as switching, amplifying and logic functions. In fact, the components required for these circuits can be made directly on the slice of silicon. The great advantage of this is not just the size of the ICs but the speed at which they can be made to work due to the short distances between components. Switching speeds in excess of 1 MHz is typical.

There are four main stages in the construction of an IC. The first of these is oxidization by exposing the silicon slice to an oxygen stream at a high temperature. The oxide formed is an excellent insulator. The next process is photo-etching where part of the oxide is removed. The silicon slice is covered in a material called a photoresist which, when exposed to light, becomes hard. It is now possible to imprint the oxidized silicon slice, which is covered with photoresist, by a pattern from a photographic transparency. The slice can now be washed in acid to etch back to the silicon those areas that were not protected by being exposed to light. The next stage is diffusion, where the slice is heated in an atmosphere of an impurity such as boron or phosphorus, which causes the exposed areas to become p- or n-type silicon. The final stage is epitaxy, which is the name given to crystal growth. New layers of silicon can be grown and doped to become n- or p-type as before. It is possible to form resistors in a similar way and small values of capacitance can be achieved. It is not possible to form any useful inductance on a chip. Figure 2.18 shows a representation of the ‘packages’ that integrated circuits are supplied in for use in electronic circuits.

The range and types of integrated circuits now available are so extensive that a chip is available for almost any application. The integration level of chips has now reached, and in many cases is exceeding, that of VLSI (very large scale integration). This means there can be more than 100 000 active elements on

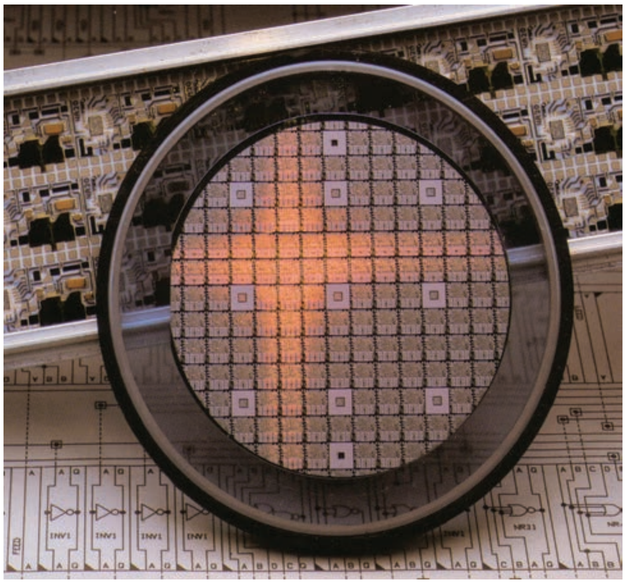


Figure 2.17 Integrated circuit components

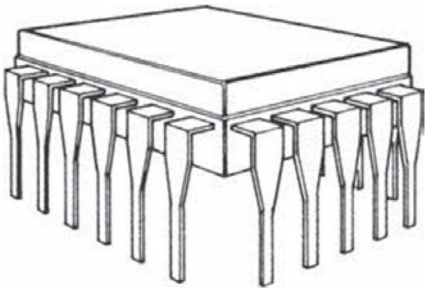


Figure 2.18 Typical integrated circuit package

Key fact

Today's microprocessors have many millions of gates and billions of individual transistors (well in excess of VLSI).

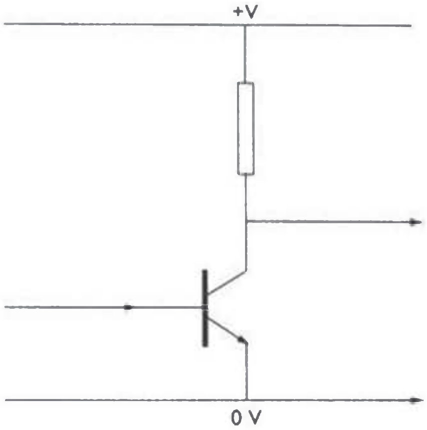


Figure 2.19 Simple amplifier circuit

one chip. Development in this area is moving so fast that often the science of electronics is now concerned mostly with choosing the correct combination of chips, and discreet components are only used as final switching or power output stages.

2.3.4 Amplifiers

The simplest form of amplifier involves just one resistor and one transistor, as shown in Figure 2.19. A small change of current on the input terminal will cause a similar change of current through the transistor and an amplified signal will be evident at the output terminal. Note however that the output will be inverted compared with the input. This very simple circuit has many applications when used more as a switch than an amplifier. For example, a very small current flowing to the input can be used to operate, say, a relay winding connected in place of the resistor.

One of the main problems with this type of transistor amplifier is that the gain of a transistor (β) can be variable and non-linear. To overcome this, some type of feedback is used to make a circuit with more appropriate characteristics. Figure 2.20 shows a more practical AC amplifier.

Resistors Rb_1 and Rb_2 set the base voltage of the transistor and, because the base–emitter voltage is constant at 0.6 V, this in turn will set the emitter voltage. The standing current through the collector and emitter resistors (Rc and Re) is hence defined and the small signal changes at the input will be reflected in an amplified form at the output, albeit inverted. A reasonable approximation of the voltage gain of this circuit can be calculated as: Rc/Re.

Capacitor C_1 is used to prevent any change in DC bias at the base terminal and C_2 is used to reduce the impedance of the emitter circuit. This ensures that Re does not affect the output.

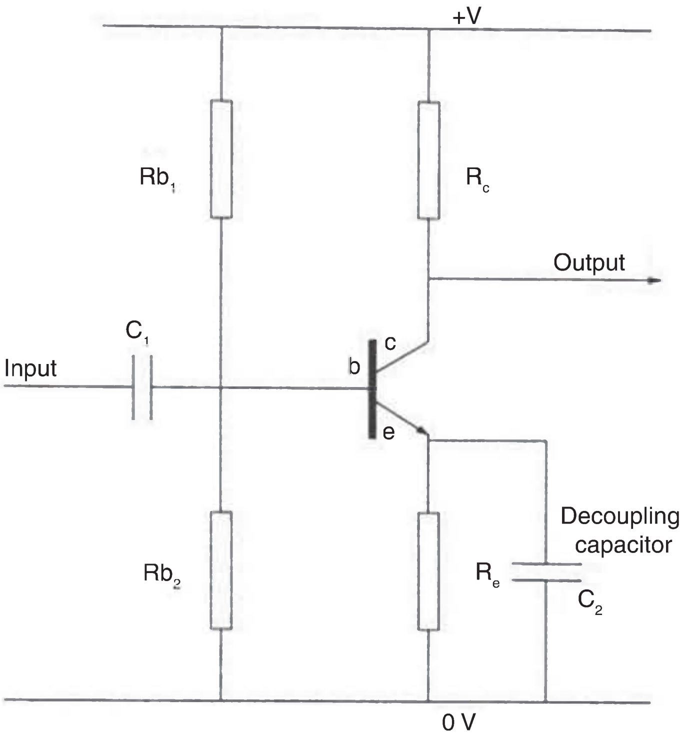


Figure 2.20 Practical AC amplifier circuit

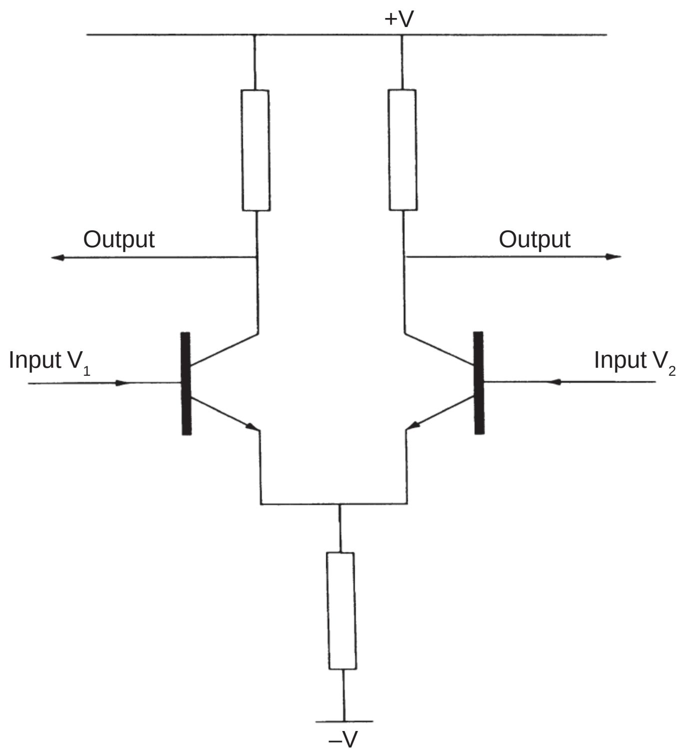


Figure 2.21 DC amplifier, long tail pair

For amplification of DC signals, a differential amplifier is often used. This amplifies the voltage difference between two input terminals. The circuit shown in Figure 2.21, known as the long tail pair, is used almost universally for DC amplifiers.

The transistors are chosen such that their characteristics are very similar. For discreet components, they are supplied attached to the same heat sink and, in integrated applications, the method of construction ensures stability. Changes in the input will affect the base–emitter voltage of each transistor in the same way, such that the current flowing through Re will remain constant. Any change in the temperature, for example, will affect both transistors in the same way and therefore the differential output voltage will remain unchanged. The important property of the differential amplifier is its ability to amplify the difference between two signals but not the signals themselves.

Integrated circuit differential amplifiers are very common, one of the most common being the 741 op-amp. This type of amplifier has a DC gain in the region of 100 000. Operational amplifiers are used in many applications and, in particular, can be used as signal amplifiers. A major role for this device is also to act as a buffer between a sensor and a load such as a display. The internal circuit of these types of device can be very complicated, but external connections and components can be kept to a minimum. It is not often that a gain of 100 000 is needed so, with simple connections of a few resistors, the characteristics of the op-amp can be changed to suit the application. Two forms of negative feedback are used to achieve an accurate and appropriate gain.

These are shown in Figure 2.22 and are often referred to as shunt feedback and proportional feedback operational amplifier circuits.

The gain with shunt (parallel) feedback is: $-R_2/R_1$

The gain with proportional feedback is: $R_2/(R_1 + R_2)$

Key fact

Integrated circuit differential amplifiers can have a DC gain in the region of 100 000.

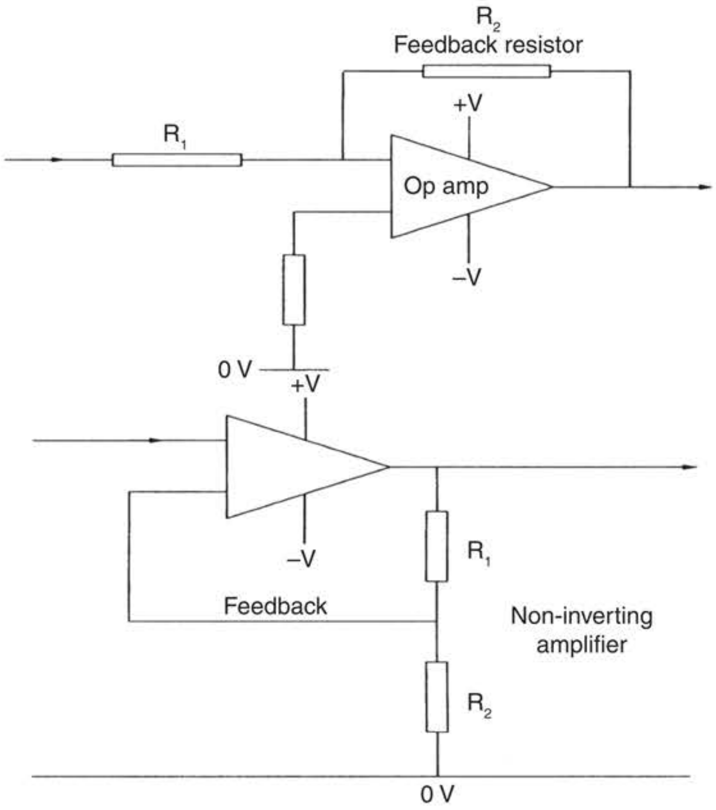


Figure 2.22 Operational amplifier feedback circuits

Key fact

Operational amplifier gain is dependent on frequency.

An important point to note with this type of amplifier is that its gain is dependent on frequency. This, of course, is only relevant when amplifying AC signals. Figure 2.23 shows the frequency response of a 741 amplifier. Op-amps are basic building blocks of many types of circuit, and some of these will be briefly mentioned later in this section.

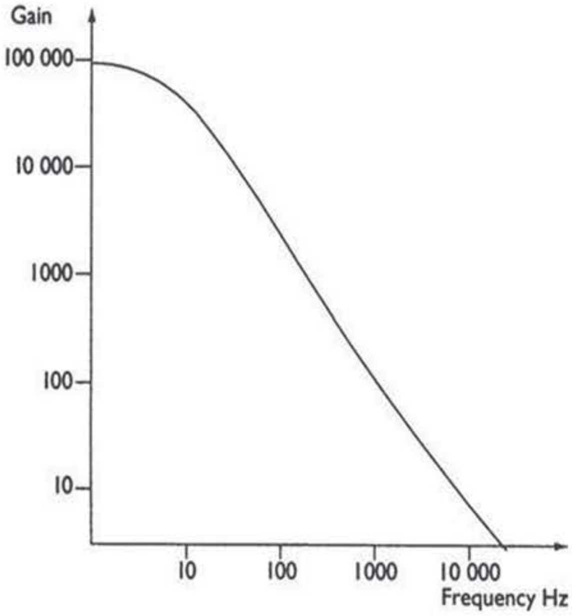


Figure 2.23 Frequency response of a 741 amplifier

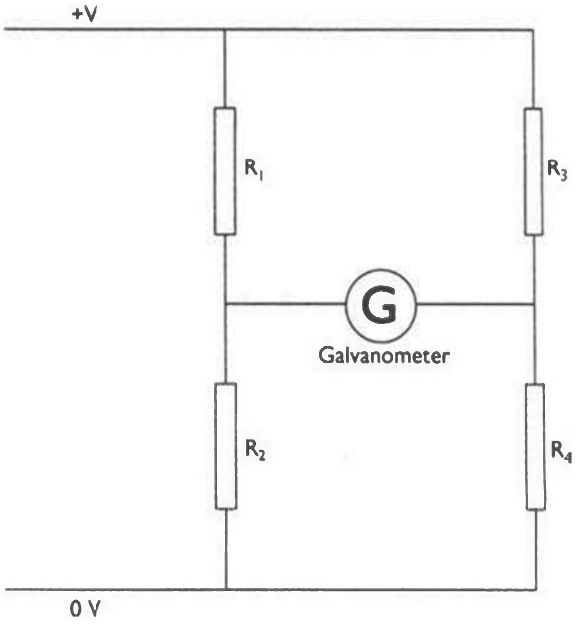


Figure 2.24 Wheatstone bridge

2.3.5 Bridge circuits

There are many types of bridge circuits but they are all based on the principle of the Wheatstone bridge, which is shown in Figure 2.24. The meter shown is a very sensitive galvanometer. A simple calculation will show that the meter will read zero when:

$$R_1/R_2 = R_3/R_4$$

To use a circuit of this type to measure an unknown resistance very accurately (R1), R3 and R4 are pre-set precision resistors and R2 is a precision resistance box. The meter reads zero when the reading on the resistance box is equal to the unknown resistor. This simple principle can also be applied to AC circuits to determine unknown inductance and capacitance.

A bridge and amplifier circuit, which may be typical of a motor vehicle application, is shown in Figure 2.25. In this circuit R_1 has been replaced by a temperature measurement thermistor. The output of the bridge is then amplified with a differential operational amplifier using shunt feedback to set the gain.

2.3.6 Schmitt trigger

The Schmitt trigger is used to change variable signals into crisp square-wave type signals for use in digital or switching circuits. For example, a sine wave fed

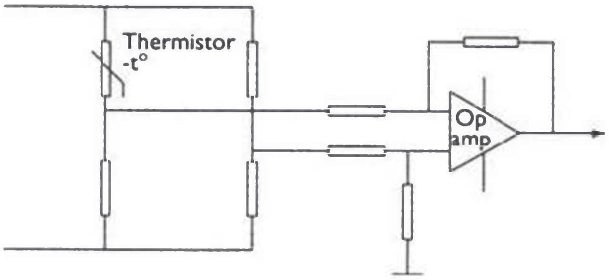


Figure 2.25 Bridge and amplifier circuit.

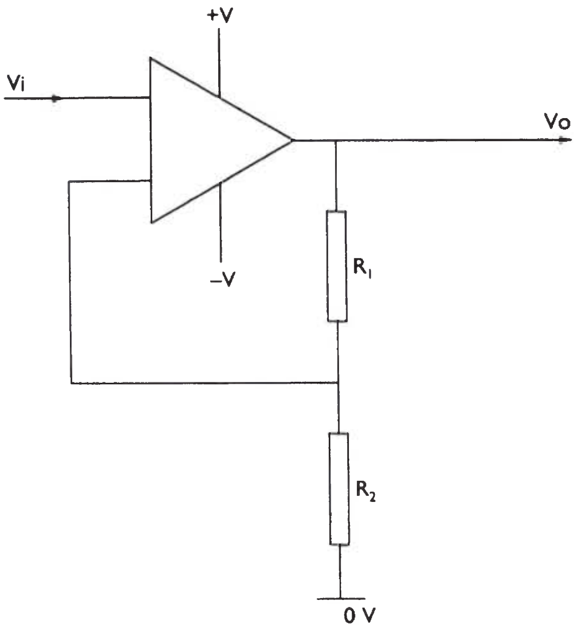


Figure 2.26 Schmitt trigger circuit utilizing an operational amplifier

Key fact

A Schmitt trigger is used to change variable signals square-wave type signals.

Definition

UTP and LTP: Upper and lower trigger points.

into a Schmitt trigger will emerge as a square wave with the same frequency as the input signal. Figure 2.26 shows a simple Schmitt trigger circuit utilizing an operational amplifier.

The output of this circuit will be either saturated positive or saturated negative due to the high gain of the amplifier. The trigger points are defined as the upper and lower trigger points (UTP and LTP) respectively. The output signal from an inductive type distributor or a crank position sensor on a motor vehicle will need to be passed through a Schmitt trigger. This will ensure that either further processing is easier, or switching is positive. Schmitt triggers can be purchased as integrated circuits in their own right or as part of other ready-made applications.

2.3.7 Timers

In its simplest form, a timer can consist of two components, a resistor and a capacitor. When the capacitor is connected to a supply via the resistor, it is accepted that it will become fully charged in 5CR seconds, where R is the resistor value in ohms and C is the capacitor value in farads. The time constant of this circuit is CR, often-denoted τ.

The voltage across the capacitor (Vc), can be calculated as follows:

$$V_c = V(I - e^{-t/CR})$$

where: V = supply voltage; t = time in seconds; C = capacitor value in farads; R = resistor value in ohms; e = exponential function.

These two components with suitable values can be made to give almost any time delay, within reason, and to operate or switch off a circuit using a transistor. Figure 2.27 shows an example of a timer circuit using this technique.

2.3.8 Filters

A filter that prevents large particles of contaminates reaching, for example, a fuel injector is an easy concept to grasp. In electronic circuits the basic idea is just

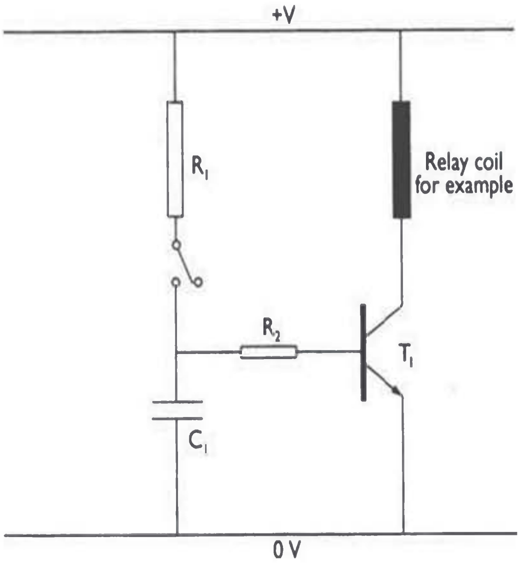


Figure 2.27 Example of a timer circuit

the same except the particle size is the frequency of a signal. Electronic filters come in two main types: a low pass filter, which blocks high frequencies, and a high pass filter, which blocks low frequencies.

Many variations of these filters are possible to give particular frequency response characteristics, such as band pass or notch filters. Here, just the basic design will be considered. The filters may also be active, in that the circuit will include amplification, or passive, when the circuit does not. Figure 2.28 shows the two main passive filter circuits.

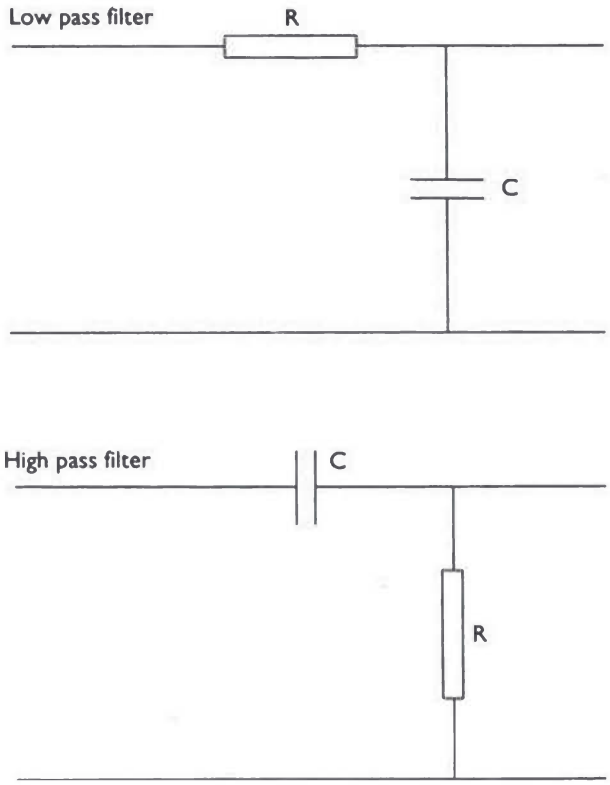


Figure 2.28 Low pass and high pass filter circuits

Definition

Reactance: The non-resistive component of impedance in an AC circuit, due to the effect of inductance or capacitance or both. It also causes the current to be out of phase with the voltage.

The principle of the filter circuits is based on the reactance of the capacitors changing with frequency. In fact, capacitive reactance, X_c decreases with an increase in frequency. The roll-off frequency of a filter can be calculated as shown:

$$f = \frac{1}{2\pi RC}$$

where: f = frequency at which the circuit response begins to roll off, R = resistor value; C = capacitor value.

It should be noted that the filters are far from perfect (some advanced designs come close though), and that the roll-off frequency is not a clear-cut 'off' but the point at which the circuit response begins to fall.

2.3.9 Darlington pair

A Darlington pair is a simple combination of two transistors that will give a high current gain, of typically several thousand. The transistors are usually mounted on a heat sink and, overall, the device will have three terminals marked as a single transistor – base, collector and emitter. The input impedance of this type of circuit is of the order of 1MΩ, hence it will not load any previous part of a circuit connected to its input. Figure 2.29 shows two transistors connected as a Darlington pair.

Key fact

The Darlington pair configuration is used for many switching applications.

The Darlington pair configuration is used for many switching applications. A common use of a Darlington pair is for the switching of the coil primary current in the ignition circuit.

2.3.10 Stepper motor driver

A later section gives details of how a stepper motor works. In this section it is the circuit used to drive the motor that is considered. For the purpose of this

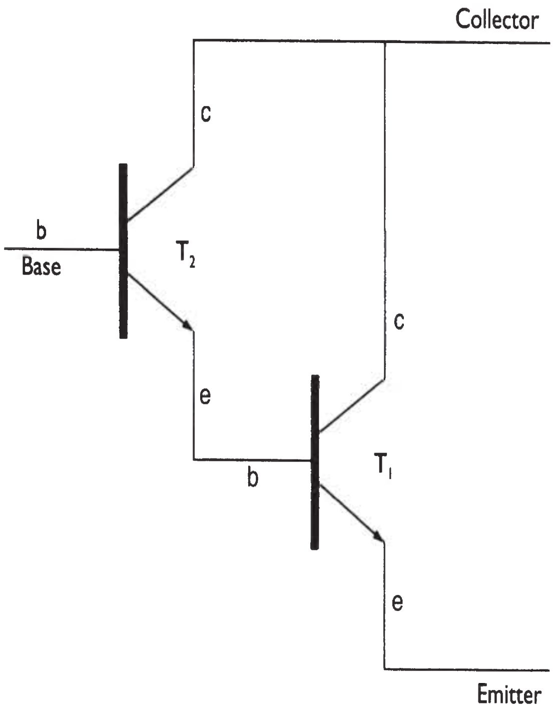


Figure 2.29 Darlington pair

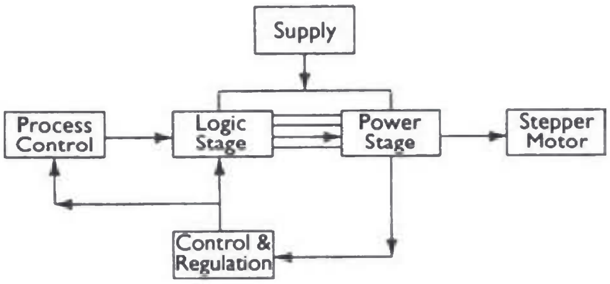


Figure 2.30 Stepper motor control system

explanation, a driver circuit for a four-phase unipolar motor is described. The function of a stepper motor driver is to convert the digital and 'wattless' (no significant power content) process control signals into signals to operate the motor coils. The process of controlling a stepper motor is best described with reference to a block diagram of the complete control system, as shown in Figure 2.30.

The process control block shown represents the signal output from the main part of an engine management ECU (electronic control unit). The signal is then converted in a simple logic circuit to suitable pulses for controlling the motor. These pulses will then drive the motor via a power stage. Figure 2.31 shows a simplified circuit of a power stage designed to control four motor windings.

2.3.11 Digital to analogue conversion

Conversion from digital signals to an analogue signal is a relatively simple process. When an operational amplifier is configured with shunt feedback the input and feedback resistors determine the gain.

$$\text{Gain} = -R_f/R_1$$

If the digital-to-analogue converted circuit is connected as shown in Figure 2.32 then the 'weighting' of each input line can be determined by choosing suitable resistor values. In the case of the four-bit digital signal, as shown, the most significant bit will be amplified with a gain of one. The next bit will be amplified with a gain of 1/2, the next bit 1/4 and, in this case, the least significant bit will be amplified with a gain of 1/8. This circuit is often referred to as an adder. The output signal produced is therefore a voltage proportional to the value of the digital input number.

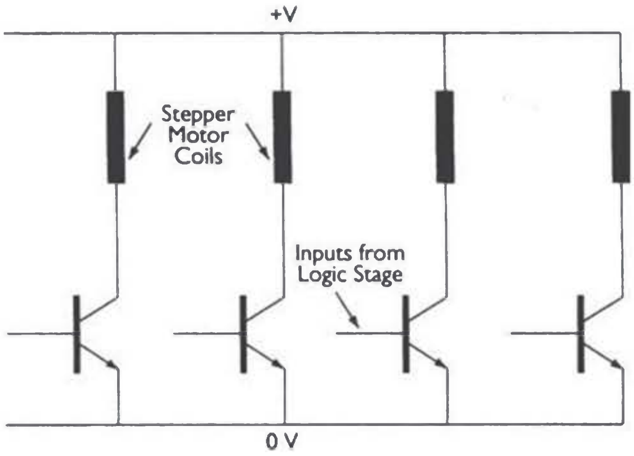


Figure 2.31 Stepper motor driver circuit (power stage)

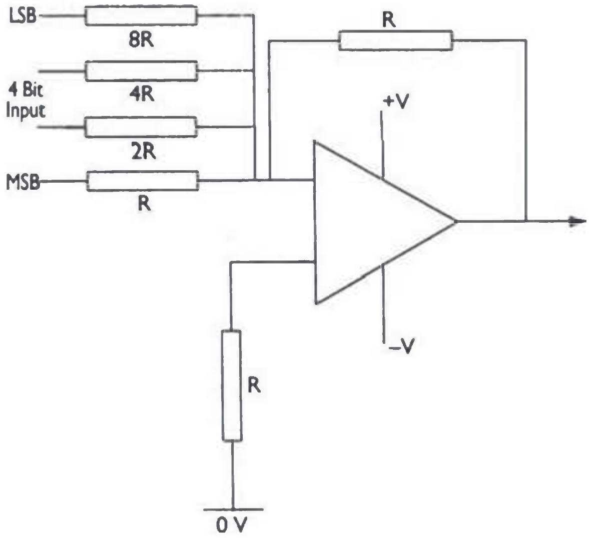


Figure 2.32 Digital-to-analogue converter

The main problem with this system is that the accuracy of the output depends on the tolerance of the resistors. Other types of digital-to-analogue converter are available, such as the R2R ladder network, but the principle of operation is similar to the above description.

2.3.12 Analogue to digital conversion

The purpose of this circuit is to convert an analogue signal, such as that received from a temperature thermistor, into a digital signal for use by a computer or a logic system. Most systems work by comparing the output of a digital-to-analogue converter (DAC) with the input voltage. Figure 2.33 is a ramp analogue-to-digital converter (ADC). This type is slower than some others but is simple in

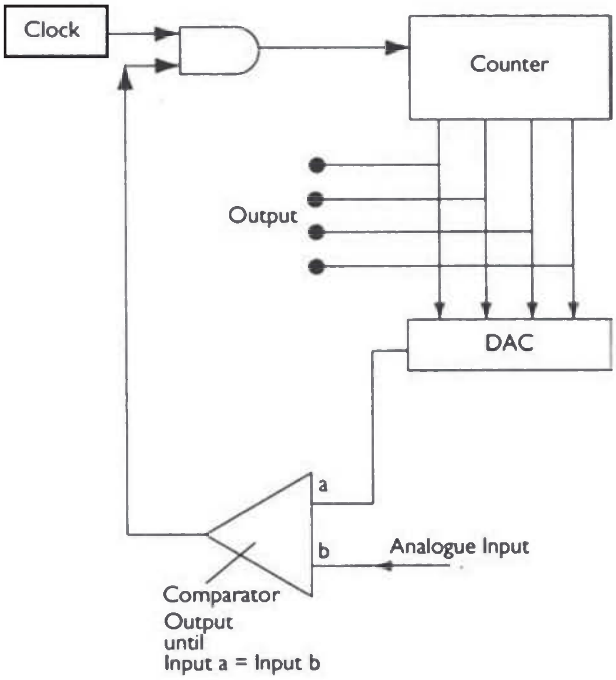


Figure 2.33 Ramp analogue to digital converter

operation. The output of a binary counter is connected to the input of the DAC, the output of which will be a ramp. This voltage is compared with the input voltage and the counter is stopped when the two are equal. The count value is then a digital representation of the input voltage. The operation of the other digital components in this circuit will be explained in the next section.

ADCs are available in IC form and can work to very high speeds at typical resolutions of one part in 4096 (12-bit word). The speed of operation is critical when converting variable or oscillating input signals. As a rule, the sampling rate must be at least twice the frequency of the input signal.

Definition

DAC: Digital-to-analogue converter.
ADC: Analogue-to-digital converter.

2.4 Digital electronics

2.4.1 Introduction to digital circuits

With some practical problems, it is possible to express the outcome as a simple yes/no or true/false answer. Let us take a simple example: if the answer to either the first or the second question is ‘yes’, then switch on the brake warning light, if both answers are ‘no’ then switch it off.

1. Is the handbrake on?
2. Is the level in the brake fluid reservoir low?

In this case, we need the output of an electrical circuit to be ‘on’ when either one or both of the inputs to the circuit are ‘on’. The inputs will be via simple switches on the handbrake and in the brake reservoir. The digital device required to carry out the above task is an OR gate, which will be described in the next section.

Once a problem can be described in logic states then a suitable digital or logic circuit can also determine the answer to the problem. Simple circuits can also be constructed to hold the logic state of their last input – these are, in effect, simple forms of ‘memory’. By combining vast quantities of these basic digital building blocks, circuits can be constructed to carry out the most complex tasks in a fraction of a second. Due to integrated circuit technology, it is now possible to create hundreds of thousands if not millions of these basic circuits on one chip. This has given rise to the modern electronic control systems used for vehicle applications as well as all the countless other uses for a computer.

In electronic circuits, true/false values are assigned voltage values. In one system, known as TTL (transistor-transistor-logic), true or logic ‘1’, is represented by a voltage of 5 V and false or logic ‘0’, by 0V.

Key fact

If a problem can be described in logic states then a digital or logic circuit can also determine the answer to the problem.

2.4.2 Logic gates

The symbols and truth tables for the basic logic gates are shown in Figure 2.34. A truth table is used to describe what combination of inputs will produce a particular output.

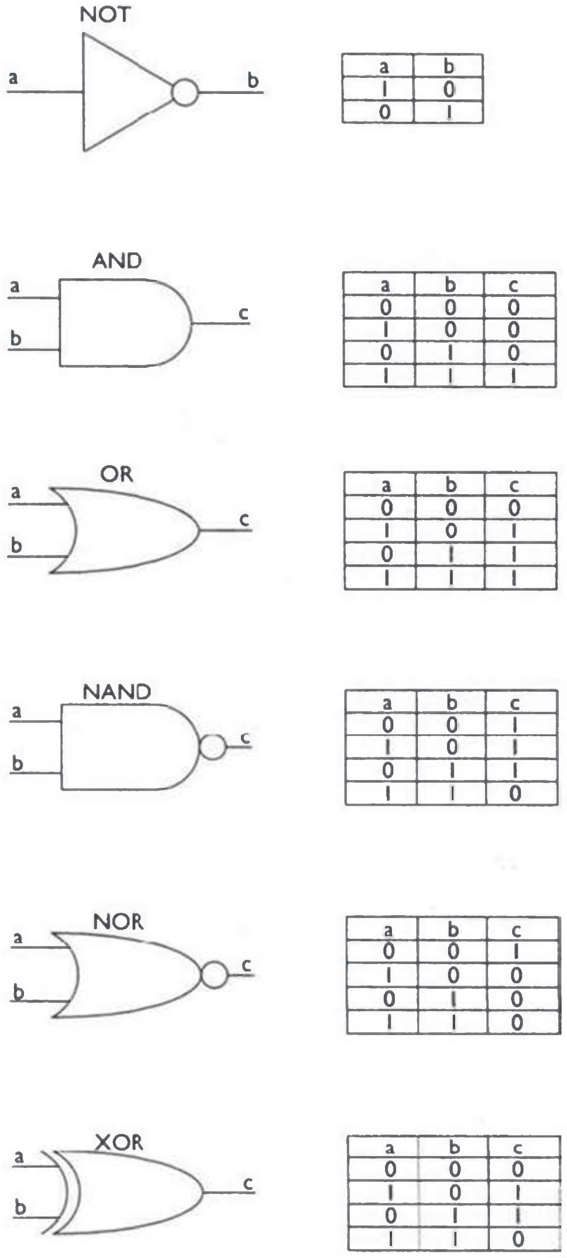
The AND gate will only produce an output of ‘1’ if both inputs (or all inputs as it can have more than two) are also at logic ‘1’. Output is ‘1’ when inputs A AND B are ‘1’.

The OR gate will produce an output when either A OR B (OR both), are ‘1’. Again more than two inputs can be used.

A NOT gate is a very simple device where the output will always be the opposite logic state from the input. In this case A is NOT B and, of course, this can only be a single input and single output device.

Key fact

A truth table is used to describe what combination of inputs will produce a particular output.



a	b
1	0
0	1

a	b	c
0	0	0
1	0	0
0	1	0
1	1	1

a	b	c
0	0	0
1	0	1
0	1	1
1	1	1

a	b	c
0	0	1
1	0	1
0	1	1
1	1	0

a	b	c
0	0	1
1	0	0
0	1	0
1	1	0

a	b	c
0	0	0
1	0	1
0	1	1
1	1	0

Figure 2.34 Logic gates and truth tables

The AND and OR gates can each be combined with the NOT gate to produce the NAND and NOR gates, respectively. These two gates have been found to be the most versatile and are used extensively for construction of more complicated logic circuits. The output of these two is the inverse of the original AND and OR gates.

The final gate, known as the exclusive OR gate, or XOR, can only be a two-input device. This gate will produce an output only when A OR B is at logic ‘1’ but not when they are both the same.

2.4.3 Combinational logic

Circuits consisting of many logic gates, as described in the previous section, are called combinational logic circuits. They have no memory or counter circuits and can be represented by a simple block diagram with N inputs and Z outputs. The

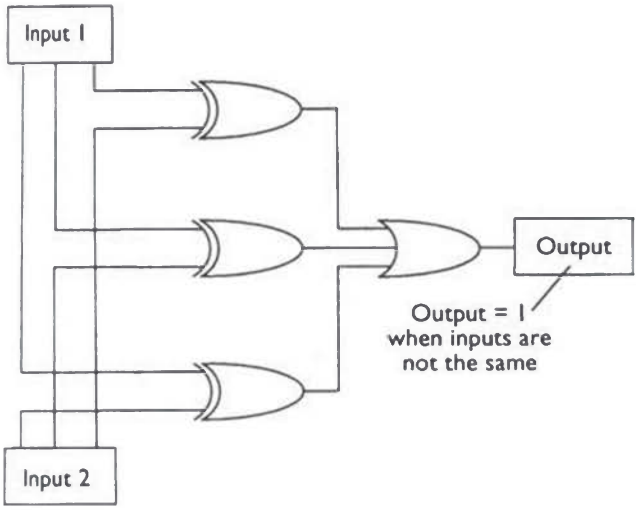


Figure 2.35 Combinational logic to compare inputs

first stage in the design process of creating a combinational logic circuit is to define the required relationship between the inputs and outputs.

Let's consider a situation where we need a circuit to compare two sets of three inputs and, if they are not the same, to provide a single logic '1' output. This is oversimplified, but could be used to compare the actions of a system with twin safety circuits, such as an ABS electronic control unit. The logic circuit could be made to operate a warning light if a discrepancy exists between the two safety circuits. Figure 2.35 shows the block diagram and one suggestion for how this circuit could be constructed.

Referring to the truth tables for basic logic circuits, the XOR gate seemed the most appropriate to carry out the comparison: it will only produce a '0' output when its inputs are the same. The outputs of the three XOR gates are then supplied to a three-input OR gate which, providing all its inputs are '0', will output '0'. If any of its inputs change to '1' the output will change to '1' and the warning light will be illuminated.

Other combinations of gates can be configured to achieve any task. A popular use is to construct an adder circuit to perform addition of two binary numbers. Subtraction is achieved by converting the subtraction to addition (4 – 3 = 1 is the same as 4 + [–3] = 1). Adders are also used to multiply and divide numbers, as this is actually repeated addition or repeated subtraction.

2.4.4 Sequential logic

The logic circuits discussed above have been simple combinations of various gates. The output of each system was only determined by the present inputs. Circuits that have the ability to memorize previous inputs or logic states, are known as sequential logic circuits. In these circuits the sequence of past inputs determines the current output. Because sequential circuits store information after the inputs are removed, they are the basic building blocks of computer memories.

Basic memory circuits are called bistables as they have two steady states. They are, however, more often referred to as flip-flops.

There are three main types of flip-flop: an RS memory, a D-type flip-flop and a JK-type flip-flop. The RS memory can be constructed by using two NAND and

Key fact

Circuits that have the ability to memorize previous inputs or logic states, are known as sequential logic circuits.

two NOT gates, as shown in Figure 2.33 next to the actual symbol. If we start with both inputs at ‘0’ and output X is at ‘1’ then as output X goes to the input of the other NAND gate its output will be ‘0’. If input A is now changed to ‘1’ output X will change to ‘0’, which will in turn cause output Y to go to ‘1’. The outputs have changed over. If A now reverts to ‘1’ the outputs will remain the same until B goes to ‘1’, causing the outputs to change over again. In this way the circuit remembers which input was last at ‘1’. If it was A then X is ‘0’ and Y is ‘1’, if it was B then X is ‘1’ and Y is ‘0’. This is the simplest form of memory circuit. The RS stands for set–reset.

The second type of flip-flop is the D-type. It has two inputs labelled CK (for clock) and D; the outputs are labelled Q and Q–. These are often called ‘Q’ and ‘not Q’. The output Q takes on the logic state of D when the clock pulse is applied. The JK-type flip-flop is a combination of the previous two flip-flops. It has two main inputs like the RS type but now labelled J and K and it is controlled by a clock pulse like the D-type. The outputs are again ‘Q’ and ‘not Q’. The circuit remembers the last input to change in the same way as the RS memory did. The main difference is that the change-over of the outputs will only occur on the clock pulse. The outputs will also change over if both J and K are at logic ‘1’, this was not allowed in the RS type.

2.4.5 Timers and counters

A device often used as a timer is called a ‘mono-stable’ as it has only one steady state. Accurate and easily controllable timer circuits are made using this device. A capacitor and resistor combination is used to provide the delay. Figure 2.36 shows a monostable timer circuit with the resistor and capacitor attached.

Every time the input goes from 0 to 1, the output Q will go from 0 to 1 for t seconds. The other output Q will do the opposite. Many variations of this type of timer are available. The time delay (t) is usually 0.7RC.

Counters are constructed from a series of bistable devices (Figure 2.37). A binary counter will count clock pulses at its input. Figure 2.38 shows a four-bit counter constructed from D-type flip-flops. These counters are called ‘ripple through’ or non-synchronous, because the change of state ripples through from the least significant bit and the outputs do not change simultaneously. The type of triggering is important for the system to work as a counter. In this case, negative edge triggering is used, which means that the devices change state when the clock pulse changes from ‘1’ to ‘0’. The counters can be configured to count up or down.

Key fact

Counters are constructed from a series of bistable devices.

In low-speed applications, ‘ripple through’ is not a problem but at higher speeds the delay in changing from one number to the next may be critical. To get over this asynchronous problem a synchronous counter can be constructed from

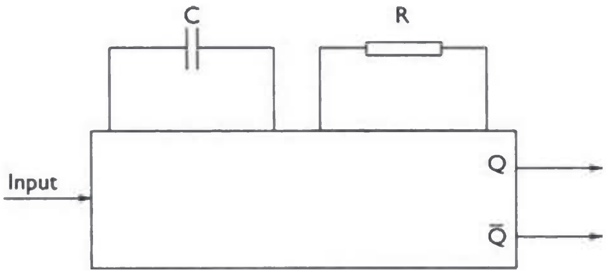


Figure 2.36 Monostable timer circuit with a resistor and capacitor attached

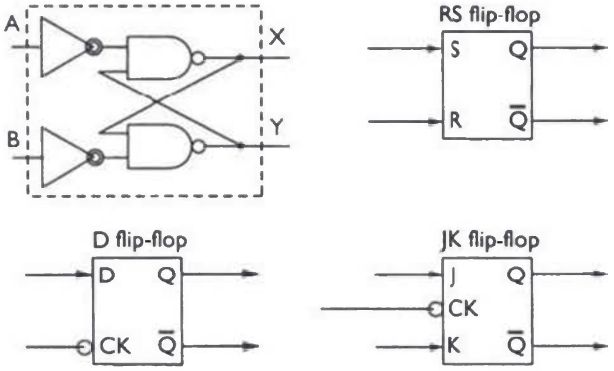


Figure 2.37 D-type and JK-type flip-flop (bistables). A method using NAND gates to make an RS type is also shown

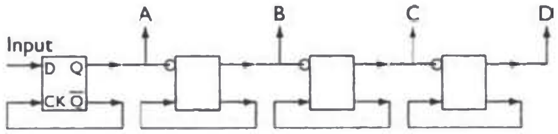


Figure 2.38 Four-bit counter from D-type flip-flops

JK-type flip-flops, together with some simple combinational logic. Figure 2.39 shows a four-bit synchronous up-counter.

With this arrangement, all outputs change simultaneously because the combinational logic looks at the preceding stages and sets the JK inputs to a ‘1’ if a toggle is required. Counters are also available ‘ready-made’ in a variety of forms including counting to non-binary bases in the up or down mode.

2.4.6 Memory circuits

Electronic circuits constructed using flip-flops as described above are one form of memory. If the flip-flops are connected as shown in Figure 2.40 they form a simple eight-bit word memory. This, however, is usually called a register rather than memory.

Eight bits (binary digits) are often referred to as one byte. Therefore, the register shown has a memory of one byte. When more than one register is used, an address is required to access or store the data in a particular register. Figure 2.41 shows a block diagram of a four-byte memory system. Also shown is an address bus, as each area of this memory is allocated a unique address. A control bus is also needed as explained below.

In order to store information (write), or to get information (read), from the system shown, it is necessary first to select the register containing the required data.

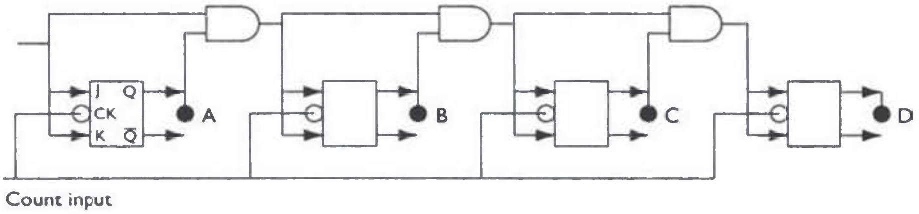


Figure 2.39 Four-bit synchronous up-counter

Definition

Bit: 1 bit
Nibble: 4 bits
Byte: 8 bits
kb: kilobit
kB: kilobyte (1024 bits)
Word: A defined number of bits.

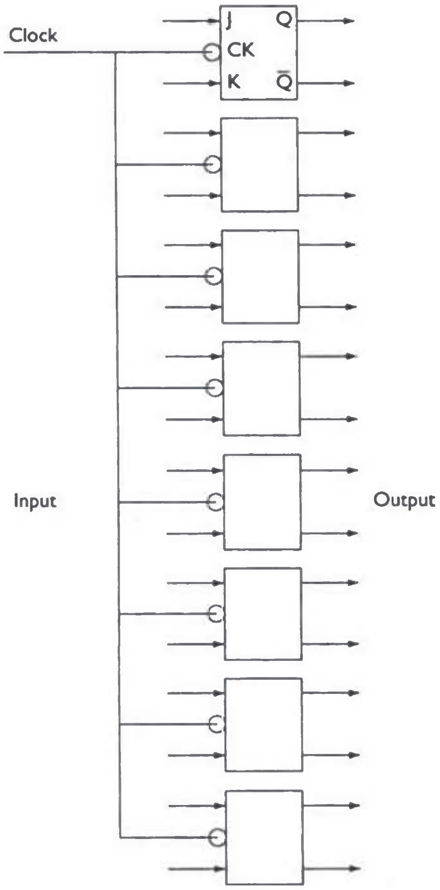


Figure 2.40 Eight-bit register using flip-flops

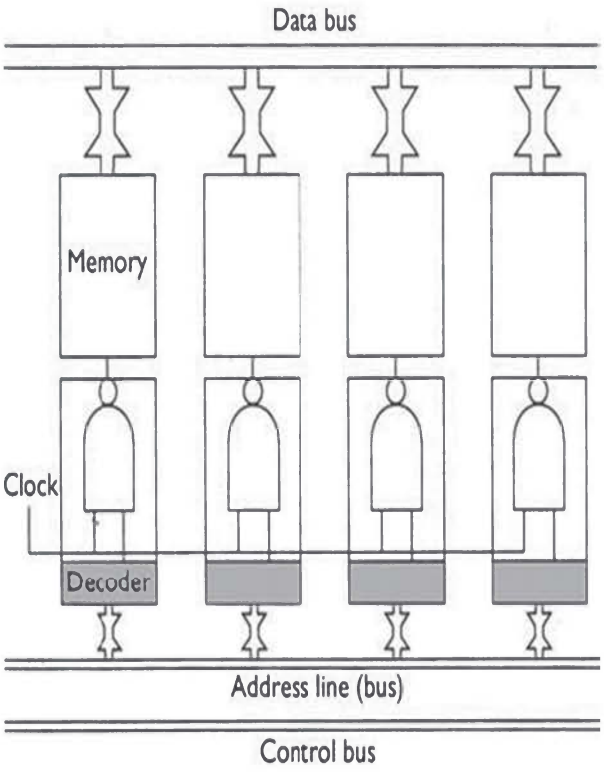


Figure 2.41 Four-byte memory card with address lines and decoders

This task is achieved by allocating an address to each register. The address bus in this example will only need two lines to select one of four memory locations using an address decoder.

The addresses will be binary; '00', '01', '10' and '11' such that if '11' is on the address bus the simple combinational logic (AND gate), will only operate one register, usually via a pin marked CS or chip select. Once a register has been selected, a signal from the control bus will 'tell' the register whether to read from or write to, the data bus. A clock pulse will ensure all operations are synchronized.

This example may appear to be a complicated way of accessing just four bytes of data. In fact, it is the principle of this technique, which is important, as the same method can be applied to access memory chips containing vast quantities of data. Note that with an address bus of two lines, 4 bytes could be accessed (22 4). If the number of address lines was increased to eight, then 256 bytes would be available (28 256). Ten address lines will address one kilobyte of data and so on.

The memory, which has just been described, together with the techniques used to access the data are typical of most computer systems. The type of memory is known as random access memory (RAM). Data can be written to and read from this type of memory but note that the memory is volatile, in other words it will 'forget' all its information when the power is switched off!

Another type of memory that can be 'read from' but not 'written to' is known as read only memory (ROM). This type of memory has data permanently stored and is not lost when power is switched off. There are many types of ROM, which hold permanent data, but one other is worthy of a mention, that is EPROM. This stands for erasable, programmable, read only memory. Its data can be changed with special equipment (some are erased with ultraviolet light), but for all other purposes its memory is permanent. In an engine management electronic control unit (ECU), operating data and a controlling program are stored in ROM, whereas instantaneous data (engine speed, load, temperature etc.) are stored in RAM.

2.4.7 Clock or astable circuits

Control circuits made of logic gates and flip-flops usually require an oscillator circuit to act as a clock. Figure 2.42 shows a very popular device, the 555-timer chip.

The external resistors and capacitor will set the frequency of the output due to the charge time of the capacitor. Comparators inside the chip cause the output to set and reset the memory (a flip-flop) as the capacitor is charged and discharged alternately to 1/3 and 2/3 of the supply voltage. The output of the chip is in the form of a square wave signal. The chip also has a reset pin to stop or start the output.

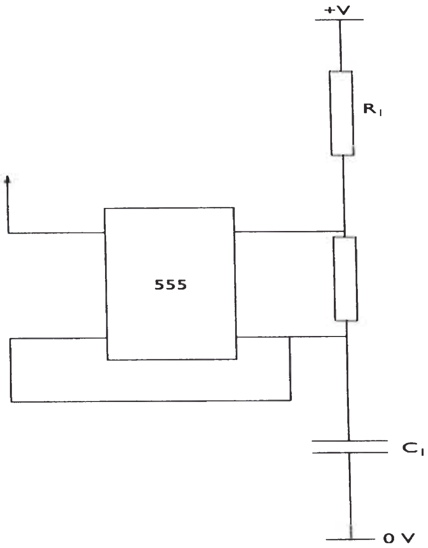


Figure 2.42 Astable circuit using a 555 IC

2.5 Microprocessor systems

2.5.1 Introduction

The advent of the microprocessor has made it possible for tremendous advances in all areas of electronic control, not least of these in the motor vehicle. Designers have found that the control of vehicle systems – which is now required to meet the customers' needs and the demands of regulations – has made it necessary to use computer control. Figure 2.43 shows a block diagram of a microcomputer containing the four major parts. These are the input and output ports, some form of memory and the CPU or central processing unit (microprocessor). It is likely that some systems will incorporate more memory chips and other specialized components. Three buses carrying data, addresses and control signals link each of the parts shown. If all the main elements as introduced above are constructed on one chip, it is referred to as a microcontroller.

2.5.2 Ports

The input port of a microcomputer system receives signals from peripherals or external components. In the case of a personal computer system, a keyboard is

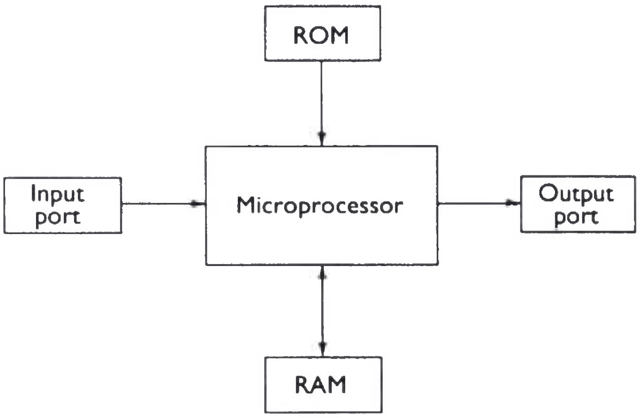


Figure 2.43 Basic microcomputer block diagram

one provider of information to the input port. A motor vehicle application could be the signal from a temperature sensor, which has been analogue to digital converted. These signals must be in digital form and usually between 0 and 5 V. A computer system, whether a PC or used on a vehicle, will have several input ports.

The output port is used to send binary signals to external peripherals. A personal computer may require output to a monitor and printer, and a vehicle computer may, for example, output to a circuit that will control the switching of the ignition coil.

2.5.3 Central processing unit (CPU)

The central processing unit or microprocessor is the heart of any computer system. It is able to carry out calculations, make decisions and be in control of the rest of the system. The microprocessor works at a rate controlled by a system clock, which generates a square wave signal usually produced by a crystal oscillator. Modern microprocessor controlled systems can work at clock speeds in excess of 300 MHz. The microprocessor is the device that controls the computer via the address, data and control buses. Many vehicle systems use microcontrollers and these are discussed later in this section.

2.5.4 Memory

Key fact

Memory is used to store both the instructions and data for the microprocessor.

The way in which memory actually works was discussed briefly in an earlier section. We will now look at how it is used in a microprocessor controlled system. Memory is the part of the system that stores both the instructions for the microprocessor (the program) and any data that the microprocessor will need to execute the instructions.

It is convenient to think of memory as a series of pigeon-holes, which are each able to store data. Each of the pigeon-holes must have an address, simply to distinguish them from each other and so that the microprocessor will ‘know’ where a particular piece of information is stored. Information stored in memory, whether it is data or part of the program, is usually stored sequentially. It is worth noting that the microprocessor reads the program instructions from sequential memory addresses and then carries out the required actions. In modern PC systems, memories can be of 128 megabytes or more! Vehicle microprocessor controlled systems do not require as much memory but mobile multimedia systems will.

2.5.5 Buses

A computer system requires three buses to communicate with or control its operations. The three buses are the data bus, address bus and the control bus. Each one of these has a particular function within the system.

Key fact

The data bus is used to carry information from one part of the computer to another.

The data bus is used to carry information from one part of the computer to another. It is known as a bi-directional bus as information can be carried in any direction. The data bus is generally 4, 8, 16 or 32 bits wide. It is important to note that only one piece of information at a time may be on the data bus. Typically, it is used to carry data from memory or an input port to the microprocessor or from the microprocessor to an output port. The address bus must first address the data that is accessed.

The address bus starts in the microprocessor and is a unidirectional bus. Each part of a computer system, whether memory or a port, has a unique address in binary format. Each of these locations can be addressed by the microprocessor and the held data placed on the data bus. The address bus, in effect, tells the computer which part of its system is to be used at any one moment.

Finally, the control bus, as the name suggests, allows the microprocessor, in the main, to control the rest of the system. The control bus may have up to 20 lines but has four main control signals. These are read, write, input/output request and memory request. The address bus will indicate which part of the computer system is to operate at any given time and the control bus will indicate how that part should operate. For example, if the microprocessor requires information from a memory location, the address of the particular location is placed on the address bus. The control bus will contain two signals, one memory request and one read signal. This will cause the contents of the memory at one particular address to be placed on the data bus. These data may then be used by the microprocessor to carry out another instruction.

2.5.6 Fetch–execute sequence

A microprocessor operates at very high speed, controlled by the system clock. Broadly speaking, the microprocessor has a simple task. It has to fetch an instruction from memory, decode the instruction and then carry out or execute the instruction. This cycle, which is carried out relentlessly (even if the instruction is to do nothing), is known as the fetch–execute sequence. Earlier in this section it was mentioned that most instructions are stored in consecutive memory locations such that the microprocessor, when carrying out the fetch–execute cycle, is accessing one instruction after another from sequential memory locations.

Key fact

One operating cycle in a microprocessor is known as a fetch–execute sequence.

Key fact

A microprocessor operates at very high speed, controlled by the system clock.

The full sequence of events may be very much as follows.

- The microprocessor places the address of the next memory location on the address bus.
- At the same time a memory read signal is placed on the control bus.
- The data from the addressed memory location are placed on the data bus.
- The data from the data bus are temporarily stored in the microprocessor.
- The instruction is decoded in the microprocessor internal logic circuits.
- The ‘execute’ phase is now carried out. This can be as simple as adding two numbers inside the microprocessor or it may require data to be output to a port. If the latter is the case, then the address of the port will be placed on the address bus and a control bus ‘write’ signal is generated.

The fetch and decode phase will take the same time for all instructions, but the execute phase will vary depending on the particular instruction. The actual time taken depends on the complexity of the instructions and the speed of the clock frequency to the microprocessor.

2.5.7 A typical microprocessor

Figure 2.44 shows the architecture of a simplified microprocessor, which contains five registers, a control unit and the arithmetic logic unit (ALU).

The operation code register (OCR) is used to hold the op-code of the instruction currently being executed. The control unit uses the contents of the OCR to determine the actions required. The temporary address register (TAR) is used to hold the operand of the instruction if it is to be treated as an address. It outputs to the address bus. The temporary data register (TDR) is used to hold data, which are to be operated on by the ALU, its output is therefore to an input of the ALU.

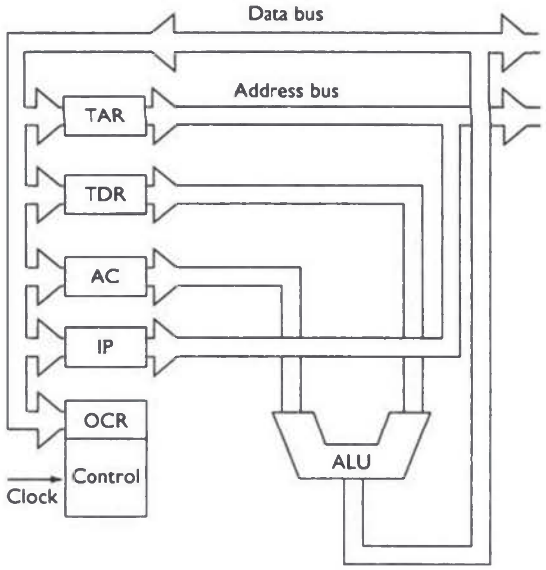


Figure 2.44 Simplified microprocessor with five registers, a control unit and the ALU or arithmetic logic unit

The ALU carries out additions and logic operations on data held in the TDR and the accumulator. The accumulator (AC) is a register, which is accessible to the programmer and is used to keep such data as a running total.

The instruction pointer (IP) outputs to the address bus so that its contents can be used to locate instructions in the main memory. It is an incremental register, meaning that its contents can be incremented by one directly by a signal from the control unit.

Execution of instructions in a microprocessor proceeds on a step by step basis, controlled by signals from the control unit via the internal control bus. The control unit issues signals as it receives clock pulses.

The process of instruction execution is as follows:

1. Control unit receives the clock pulse.
2. Control unit sends out control signals.
3. Action is initiated by the appropriate components.
4. Control unit receives the clock pulse.
5. Control unit sends out control signals.
6. Action is initiated by the appropriate components. And so on.

A typical sequence of instructions to add a number to the one already in the accumulator is as follows:

1. IP contents placed on the address bus.
2. Main memory is read and contents placed on the data bus.
3. Data on the data bus are copied into OCR.
4. IP contents incremented by one.
5. IP contents placed on the address bus.
6. Main memory is read and contents placed on the data bus.
7. Data on the data bus are copied into TDR.
8. ALU adds TDR and AC and places result on the data bus.
9. Data on the data bus are copied into AC.
10. IP contents incremented by one.

The accumulator now holds the running total. Steps 1 to 4 are the fetch sequence and steps 5 to 10 the execute sequence. If the full fetch–execute sequence above was carried out, say, nine times this would be the equivalent of multiplying the number in the accumulator by 10! This gives an indication as to just how basic the level of operation is within a computer.

Now to take a giant step forwards. It is possible to see how the microprocessor in an engine management ECU can compare a value held in a RAM location with one held in a ROM location. The result of this comparison of, say, instantaneous engine speed in RAM and a pre-programmed figure in ROM, could be to set the ignition timing to another pre-programmed figure.

2.5.8 Microcontrollers

As integration technology advanced it became possible to build a complete computer on a single chip. This is known as a microcontroller. The microcontroller must contain a microprocessor, memory (RAM and/or ROM), input ports and output ports. A clock is included in some cases.

A typical family of microcontrollers is the 'Intel' 8051 series. These were first introduced in 1980 but are still a popular choice for designers. A more up-to-date member of this family is the 87C528 microcontroller which has 32K EPROM, 512 bytes of RAM, three (16 bit) timers, four I/O ports and a built in serial interface.

Microcontrollers are available such that a pre-programmed ROM may be included. These are usually made to order and are only supplied to the original customer. Figure 2.45 shows a simplified block diagram of the 8051 microcontroller.

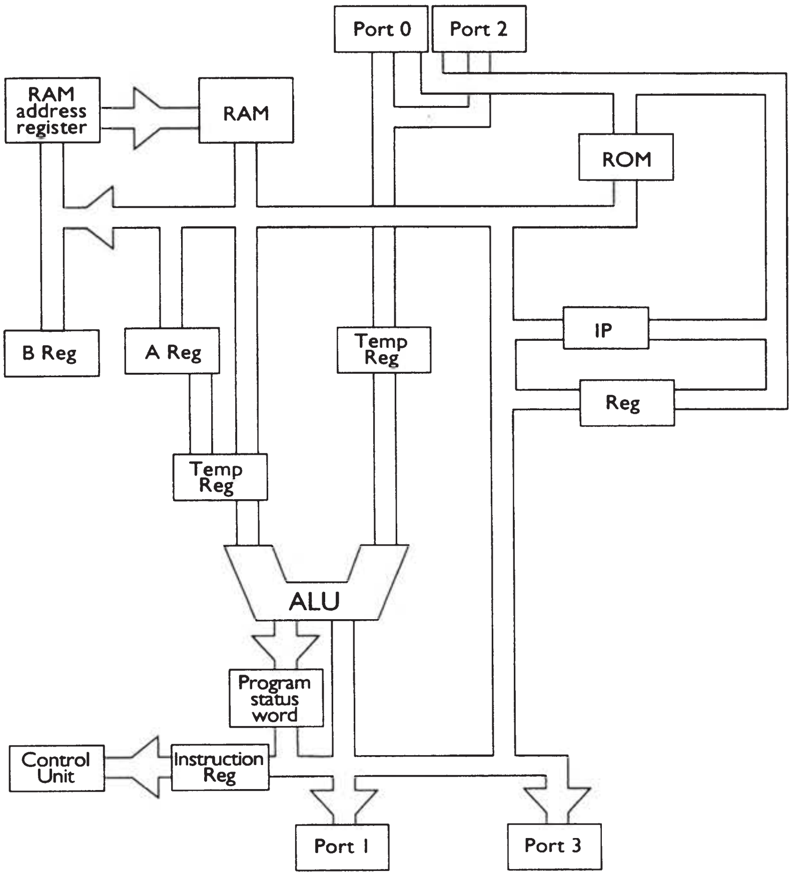


Figure 2.45 Simplified block diagram of the 8051 microcontroller

2.5.9 Testing microcontroller systems

If a microcontroller system is to be constructed with the program (set of instructions) permanently held in ROM, considerable testing of the program is required. This is because, once the microcontroller goes into production, tens if not hundreds of thousands of units will be made. A hundred thousand microcontrollers with a hard-wired bug in the program would be a very expensive error!

There are two main ways in which software for a microcontroller can be tested. The first, which is used in the early stages of program development, is by a simulator. A simulator is a program that is executed on a general purpose computer and which simulates the instruction set of the microcontroller. This method does not test the input or output devices.

The most useful aid for testing and debugging is an in-circuit emulator. The emulator is fitted in the circuit in place of the microcontroller and is, in turn, connected to a general purpose computer. The microcontroller program can then be tested in conjunction with the rest of the hardware with which it is designed to work. The PC controls the system and allows different procedures to be tested. Changes to the program can easily be made at this stage of the development.

2.5.10 Programming

To produce a program for a computer, whether it is for a PC or a microcontroller-based system is generally a six-stage process.

1 Requirement analysis

This seeks to establish whether in fact a computer-based approach is in fact the best option. It is, in effect, a feasibility study.

2 Task definition

The next step is to produce a concise and unambiguous description of what is to be done. The outcome of this stage is to produce the functional specifications of the program.

3 Program design

The best approach here is to split the overall task into a number of smaller tasks. Each of which can be split again and so on if required. Each of the smaller tasks can then become a module of the final program. A flow chart like the one shown in Figure 2.46 is often the result of this stage, as such charts show the way subtasks interrelate.

4 Coding

This is the representation of each program module in a computer language. The programs are often written in a high-level language such as Turbo C, Pascal or even Basic. Turbo C and C are popular as they work well in program modules and produce a faster working program than many of the other languages. When the source code has been produced in the high-level language, individual modules are linked and then compiled into machine language – in other words a language consisting of just ‘1s’ and ‘0s’ and in the correct order for the microprocessor to understand.

5 Validation and debugging

Once the coding is completed it must be tested extensively. This was touched upon in the previous section but it is important to note that the program must be tested under the most extreme conditions. Overall, the tests must show that, for

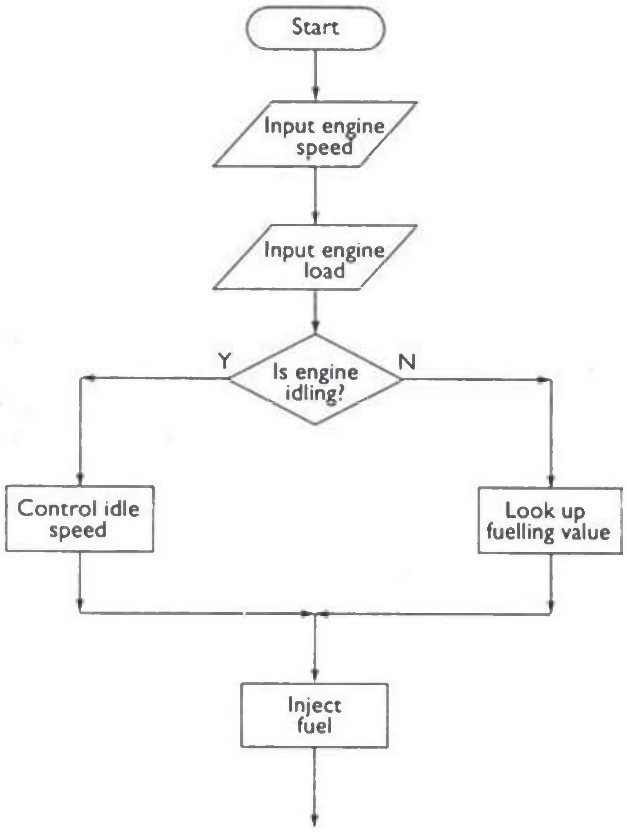


Figure 2.46 Computer program flowchart

an extensive range of inputs, the program must produce the required outputs. In fact, it must prove that it can do what it was intended to do! A technique known as single stepping where the program is run one step at a time, is a useful aid for debugging.

6 Operation and maintenance

Finally, the program runs and works but, in some cases, problems may not show up for years and some maintenance of the program may be required for new production; the Millennium bug, for example!

The six steps above should not be seen in isolation, as often the production of a program is iterative and steps may need to be repeated several times.

Some example programs and source code examples can be downloaded from my web site (the URL address is given in the preface).

2.6 Measurement

2.6.1 What is measurement

Measurement is the act of measuring physical quantities to obtain data that are transmitted to recording/display devices and/or to control devices. The term 'instrumentation' is often used in this context to describe the science and technology of the measurement system.

The first task of any measurement system is to translate the physical value to be measured, known as the measurand, into another physical variable, which can be used to operate the display or control device. In the motor vehicle system, the majority of measurands are converted into electrical signals. The sensors that carry out this conversion are often called transducers.

2.6.2 A measurement system

A complete measurement system will vary depending on many factors but many vehicle systems will consist of the following stages.

1. Physical variable.
2. Transduction.
3. Electrical variable.
4. Signal processing.
5. A/D conversion.
6. Signal processing.
7. Display or use by a control device.

Some systems may not require Steps 5 and 6. As an example, consider a temperature measurement system with a digital display. This will help to illustrate the above seven-step process.

1. Engine water temperature.
2. Thermistor.
3. Resistance decreases with temperature increase.
4. Linearization.
5. A/D conversion.
6. Conversion to drive a digital display.
7. Digital read-out as a number or a bar graph.

Figure 2.47 shows a complete measurement system as a block diagram.

2.6.3 Sources of error in measurement

An important question to ask when designing an instrumentation or measurement system is: What effect will the measurement system have on the variable being measured?

Consider the water temperature measurement example discussed in the previous section. If the transducer is immersed in a liquid, which is at a higher temperature than the surroundings, then the transducer will conduct away some of the heat and lower the temperature of the liquid. This effect is likely to be negligible in this example, but in others, it may not be so small. However, even in this case it is possible that, due to the fitting of the transducer, the water temperature surrounding the sensor will be lower than the rest of the system. This is known as an invasive measurement. A better example may be that if a device is fitted into a petrol pipe to measure flow rate, then it is likely that the device itself will restrict the flow in some way. Returning to the previous example of the temperature transducer it is also possible that the very small current passing through the transducer will have a heating effect.

Key fact

Invasive measurement: The act of measuring changes the value.

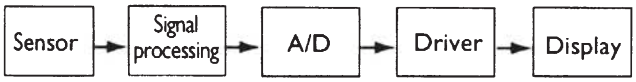


Figure 2.47 Measurement system block diagram

Errors in a measurement system affect the overall accuracy. Errors are also not just due to invasion of the system. There are many terms associated with performance characteristics of transducers and measurement systems. Some of these terms are considered below.

Key fact

Errors in a measurement system affect the overall accuracy.

Accuracy

A descriptive term meaning how close the measured value of a quantity is to its actual value. Accuracy is expressed usually as a maximum error. For example, if a length of about 30 cm is measured with an ordinary wooden ruler then the error may be up to 1 mm too high or too low. This is quoted as an accuracy of 1 mm. This may also be expressed as a percentage which in this case would be 0.33%. An electrical meter is often quoted as the maximum error being a percentage of full-scale deflection. The maximum error or accuracy is contributed to by a number of factors explained below.

Resolution

The 'fineness' with which a measurement can be made. This must be distinguished from accuracy. If a quality steel ruler were made to a very high standard but only had markings or graduations of one per centimetre it would have a low resolution even though the graduations were very accurate.

Hysteresis

For a given value of the measurand, the output of the system depends on whether the measurand has acquired its value by increasing or decreasing from its previous value. You can prove this next time you weigh yourself on some scales. If you step on gently you will 'weigh less' than if you jump on and the scales overshoot and then settle.

Repeatability

The closeness of agreement of the readings when a number of consecutive measurements are taken of a chosen value during full range traverses of the measurand. If a 5 kg set of weighing scales was increased from zero to 5 kg in 1 kg steps a number of times, then the spread of readings is the repeatability. It is often expressed as a percentage of full scale.

Zero error or zero shift

The displacement of a reading from zero when no reading should be apparent. An analogue electrical test meter, for example, often has some form of adjustment to zero the needle.

Linearity

The response of a transducer is often non-linear (see the response of a thermistor in the next section). Where possible, a transducer is used in its linear region. Non-linearity is usually quoted as a percentage over the range in which the device is designed to work.

Sensitivity or scale factor

A measure of the incremental change in output for a given change in the input quantity. Sensitivity is quoted effectively as the slope of a graph in the linear region. A figure of 0.1 V/°C for example, would indicate that a system would increase its output by 0.1 V for every 1 °C increase in temperature of the input.

Response time

The time taken by the output of a system to respond to a change in the input. A system measuring engine oil pressure needs a faster response time than a fuel tank quantity system. Errors in the output will be apparent if the measurement is taken quicker than the response time.

Looking again at the seven steps involved in a measurement system will highlight the potential sources of error.

1. Invasive measurement error.
2. Non-linearity of the transducer.
3. Noise in the transmission path.
4. Errors in amplifiers and other components.
5. Quantization errors when digital conversion takes place.
6. Display driver resolution.
7. Reading error of the final display.

Many good textbooks are available for further study, devoted solely to the subject of measurement and instrumentation. This section is intended to provide the reader with a basic grounding in the subject.

2.7 Sensors

2.7.1 Thermistors

Thermistors are the most common device used for temperature measurement on a motor vehicle. The principle of measurement is that a change in temperature will cause a change in resistance of the thermistor, and hence an electrical signal proportional to the measured can be obtained.

Most thermistors in common use are of the negative temperature coefficient (NTC) type (but not all). The actual response of the thermistors can vary but typical values for those used in motor vehicles will vary from several kilo ohms (kΩ) at 0 °C to a few hundred ohms at 100 °C. The large change in resistance for a small change in temperature makes the thermistor ideal for most vehicles' uses. It can also be easily tested with simple equipment.

Key fact

Most thermistors in common use are of the negative temperature coefficient (NTC) type.

Thermistors are constructed of semiconductor materials such as cobalt or nickel oxides. The change in resistance with a change in temperature is due to the electrons being able to break free from the covalent bonds more easily at higher temperatures; this is shown in Figure 2.49(i). A thermistor temperature measuring system can be very sensitive due to large changes in resistance with a relatively small change in temperature. A simple circuit to provide a varying voltage signal proportional to temperature is shown in Figure 2.49(ii). Note the supply must be constant and the current flowing must not significantly heat the thermistor. These could both be sources of error.

The temperature of a typical thermistor will increase by 1 °C for each 1.3 mW of power dissipated. Figure 2.49(iii) shows the resistance against temperature curve for a thermistor. This highlights the main problem with a thermistor, its non-linear response. Using a suitable bridge circuit, it is possible to produce non-linearity that will partially compensate for the thermistor's non-linearity. This is represented by Figure 2.49(iv). The combination of these two responses is also shown. The optimum linearity is achieved when the mid points of the temperature and the voltage ranges lie on the curve. Figure 2.49(v) shows a bridge circuit for this purpose. It is possible to work out suitable values for R1, R2 and R3. This then gives the more linear output as represented by Figure 2.49(vi). The voltage signal can now be A/D converted if necessary, for further use.

Figure 2.48 Temperature sensor

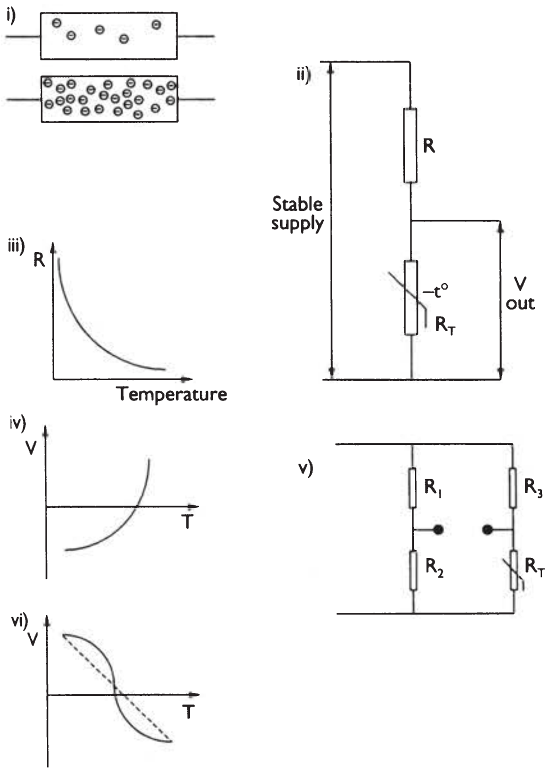


Figure 2.49 (i) How a thermistor changes resistance; (ii) circuit to provide a varying voltage signal proportional to temperature; (iii) resistance against temperature curve for a thermistor; (iv) non-linearity to compensate partially for the thermistor's non-linearity; (v) bridge circuit to achieve maximum linearity; (vi) final output signal

The resistance Rt of a thermistor decreases non-linearly with temperature according to the relationship:

$$R_t = Ae^{(B/T)}$$

where: R_t = resistance of the thermistor, T = absolute temperature, B = characteristic temperature of the thermistor (typical value 3000 K), A = constant of the thermistor.

For the bridge configuration as shown V_o is given by:

$$V_o = V_s\left(\frac{R_2}{R_2 + R_1} - \frac{R_1}{R_1 + R_3}\right)$$

By choosing suitable resistor values the output of the bridge will be as shown in Figure 2.49 (vi). This is achieved by substituting the known values of R_t at three temperatures and deciding that, for example, V_o = 0 at 0 °C, V_o = 0.5 V at 50 °C and V_o = 1 V at 100 °C.

2.7.2 Thermocouples

If two different metals are joined together at two junctions, the thermoelectric effect known as the Seebeck effect takes place. If one junction is at a higher temperature than the other junction, then this will be registered on the meter. This is the basis for the sensor known as the thermocouple. Figure 2.50 shows the thermocouple principle and appropriate circuits. Notice that the thermocouple measures a difference in temperature that is $T_1 - T_2$. To make the system of any practical benefit then T_1 must be kept at a known temperature.

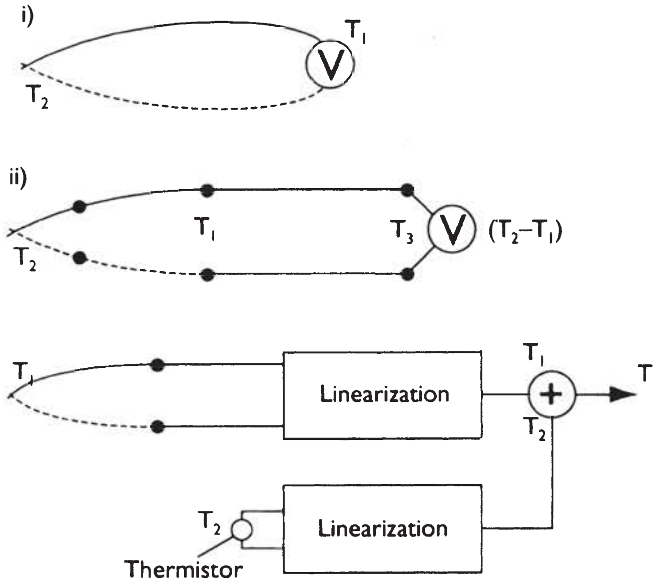


Figure 2.50 Thermocouple principle and circuits

Definition

Seebeck effect: The development of a voltage due to differences in temperature between two junctions of dissimilar metals.

Key fact

Thermocouples are in general used for measuring high temperatures.

The lower figure shows a practical circuit in which, if the connections to the meter are at the same temperature, the two voltages produced at these junctions will cancel out.

Cold junction compensation circuits can be made to compensate for changes in temperature of T_1. These often involve the use of a thermistor circuit.

Thermocouples are in general used for measuring high temperatures. A thermocouple combination of a 70% platinum and 30% rhodium alloy in a junction with a 94% platinum and 6% rhodium alloy, is known as a type B thermocouple and has a useful range of 0–1500 °C. Vehicle applications are in areas such as exhaust gas and turbo charger temperature measurement.

2.7.3 Inductive sensors

Inductive-type sensors are used mostly for measuring speed and position of a rotating component. They work on the very basic principle of electrical induction (a changing magnetic flux will induce an electromotive force in a winding). Figure 2.51 shows the inductive sensor principle and a typical device used as a crankshaft speed and position sensor.

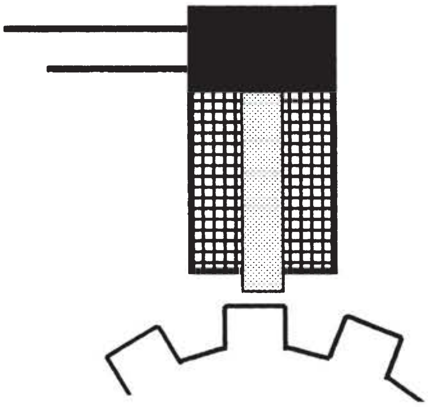


Figure 2.51 Inductive sensor

The output voltage of most inductive-type sensors approximates to a sine wave. The amplitude of this signal depends on the rate of change of flux. This is determined mostly by the original design: by the number of turns, magnet strength and the gap between the sensor and the rotating component. Once in use though, the output voltage increases with the speed of rotation. In the majority of applications, it is the frequency of the signal that is used. The most common way of converting the output of an inductive sensor to a useful signal is to pass it through a Schmitt trigger circuit. This produces constant amplitude but a variable frequency square wave.

In some cases the output of the sensor is used to switch an oscillator on and off or quench the oscillations. A circuit for this is shown in Figure 2.52. The oscillator produces a very high frequency of about 4 MHz and this when switched on and off by the sensor signal and then filtered, produces a square wave. This system has a good resistance to interference.

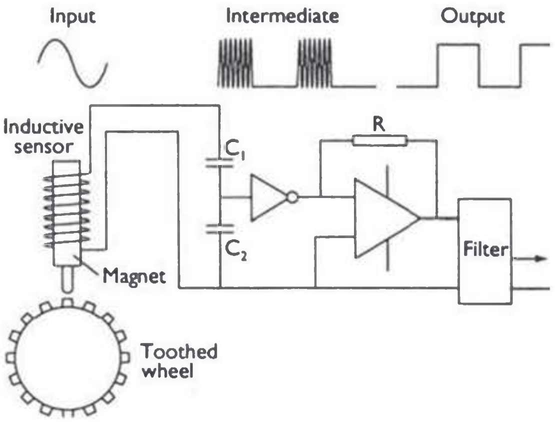


Figure 2.52 Inductive sensor and quenched oscillator circuit

Figure 2.53 Inductive sensor on an engine

2.7.4 Hall Effect

The Hall Effect was first noted by a Dr E.H. Hall: it is a simple principle, as shown in Figure 2.54. If a certain type of crystal is carrying a current in a transverse magnetic field then a voltage will be produced at right angles to the supply current. The magnitude of the voltage is proportional to the supply current and to the magnetic field strength. Figures 2.55 and 2.56 show part of a Bosch distributor, the principle of which is to 'switch' the magnetic field on and off using

Definition

Hall Effect: The production of a voltage difference across an electrical conductor, transverse to an electric current in the conductor and a magnetic field perpendicular to the current. This effect was discovered by Edwin Hall in 1879.

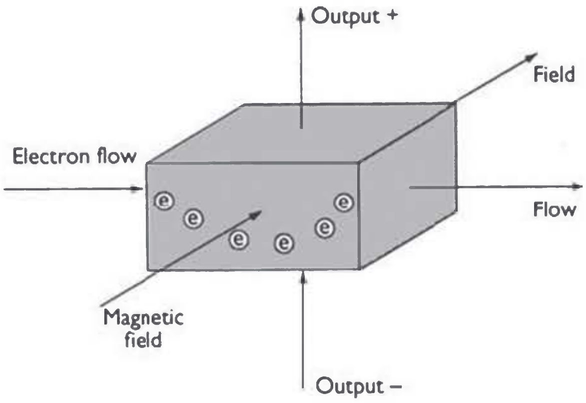


Figure 2.54 Hall Effect principle

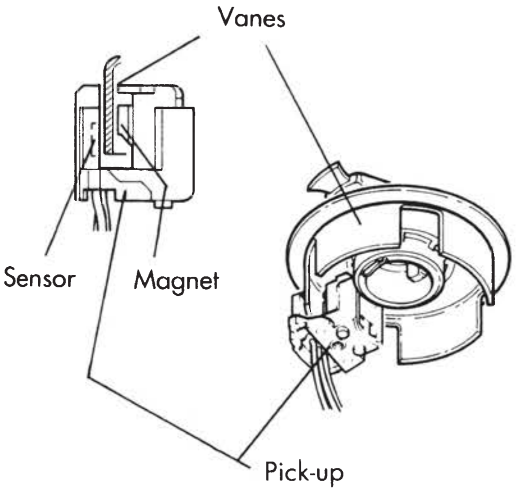


Figure 2.55 Hall Effect sensor used in a distributor

Figure 2.56 Hall Effect distributor

a chopper plate. The output of this sensor is almost a square wave with constant amplitude.

The Hall Effect can also be used to detect current flowing in a cable. The magnetic field produced around the cable is proportional to the current flowing.

Hall Effect sensors are very popular. This is partly due to their reliability but also the fact that they directly produce a constant amplitude square wave in speed measurement applications and a varying DC voltage for either position sensing or current sensing.

A number of new sensors are becoming available. Hall Effect sensors are now often used in place of inductive sensors for applications such as engine speed and wheel speed. The two main advantages are that measurement of lower (or even zero) speed is possible and the voltage output of the sensors is independent of speed. Figure 2.57 shows a Hall Effect sensor used to sense wheel speed.

Figure 2.57 Hall Effect wheel speed sensor

2.7.5 Strain gauges

Figure 2.58 shows a simple strain gauge together with a bridge and amplifier circuit used to convert its change in resistance into a voltage signal. The second strain gauge is fitted on the device under test but in a non-strain position to

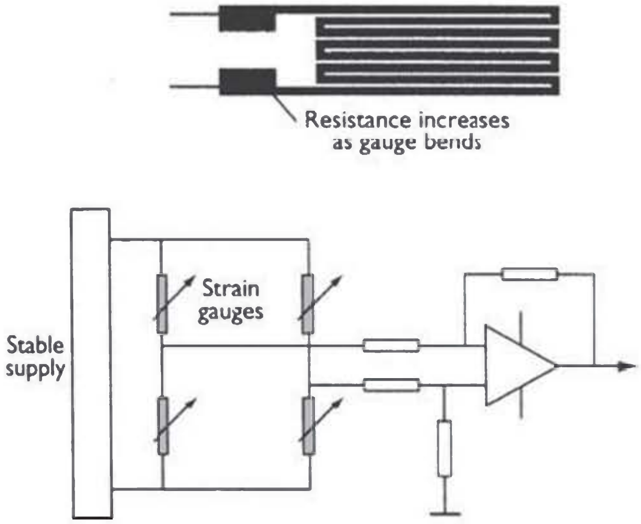


Figure 2.58 Strain gauge and a bridge circuit

compensate for temperature changes. Quite simply, when a strain gauge is stretched its resistance will increase, and when it is compressed its resistance decreases.

Most strain gauges consist of a thin layer of film that is fixed to a flexible backing sheet, usually paper. This, in turn, is bonded to the part where strain is to be measured. The sensitivity of a strain gauge is defined by its 'gauge factor'.

$$K = \frac{\Delta R / R}{E}$$

where: K *= gauge factor;* R *= original resistance;* ΔR *= change in resistance;* E *= strain (change in length/original length, l/l).*

Most resistance strain gauges have a resistance of about 100 and a gauge factor of about 2.

Strain gauges can be used indirectly to measure engine manifold pressure. Figure 2.59 shows an arrangement of four strain gauges on a diaphragm forming part of an aneroid chamber used to measure pressure. When changes in manifold pressure act on the diaphragm the gauges detect the strain. The output of the circuit is via a differential amplifier as shown, which must have a very high input resistance so as not to affect the bridge balance. The actual size of this sensor may be only a few millimetres in diameter. Changes in temperature are compensated for, as all four gauges would be affected in a similar way, thus the bridge balance would remain constant.

Key fact

Strain gauges can be used indirectly to measure engine manifold pressure.

2.7.6 Variable capacitance

The value of a capacitor is determined by the surface area of its plates, the distance between the plates and the nature of the dielectric. Sensors can be constructed to take advantage of these properties. Three sensors using the variable capacitance technique are shown in Figure 2.60. These are as follows:

1. Liquid level sensor. The change in liquid level changes the dielectric value.
2. Pressure sensor. Similar to the strain gauge pressure sensor but this time the distance between capacitor plates changes.
3. Position sensor. Detects changes in the area of the plates.

Definition

Dielectric: Having the property of transmitting electric force without conduction; insulating.

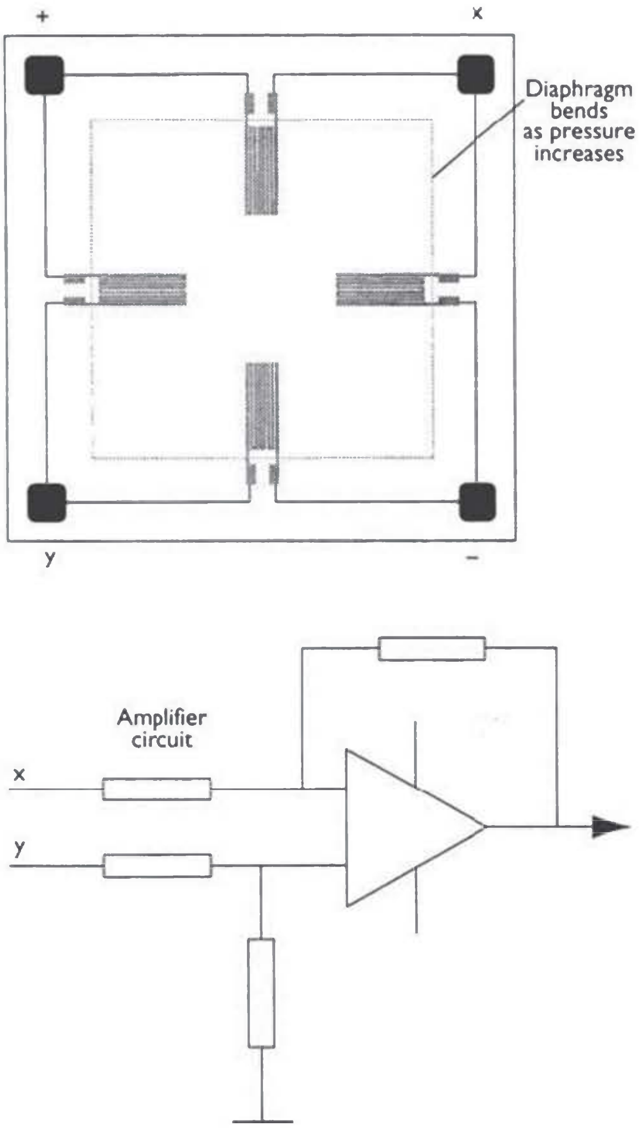


Figure 2.59 Strain gauge pressure sensor, bridge circuit and amplifier

2.7.7 Variable resistance

The two best examples of vehicle applications for variable resistance sensors are the throttle position sensor and the flap-type air flow sensor. Whereas variable capacitance sensors are used to measure small changes, variable resistance sensors generally measure larger changes in position. This is due to a lack of sensitivity inherent in the construction of the resistive track.

The throttle position sensor, as shown in Figure 2.61, is a potentiometer in which, when supplied with a stable voltage (often 5 V) the voltage from the wiper contact will be proportional to the throttle position. In many cases now, the throttle potentiometer is used to indicate the rate of change of throttle position. This information is used when implementing acceleration enrichment or, inversely, over-run fuel cut-off.

The output voltage of a rotary potentiometer can be calculated:

$$V_o = V_s \left(\frac{a_i}{a_t} \right)$$

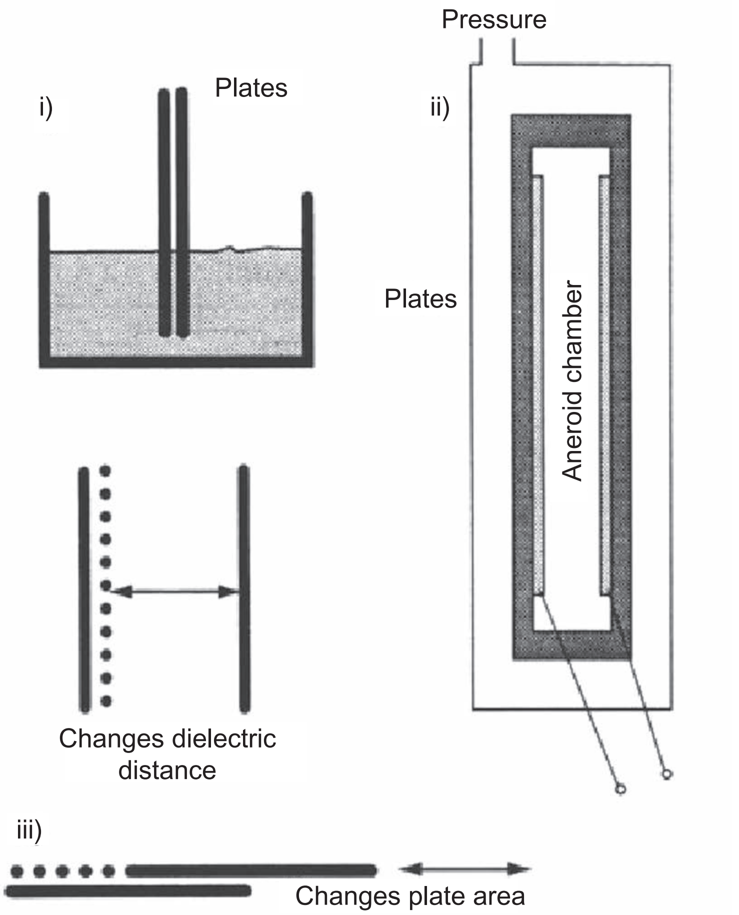


Figure 2.60 Variable capacitance sensors (i) liquid level; (ii) pressure (iii) position

where: V_o *= voltage out;* V_s *= voltage supply;* a_i *= angle moved;* a_t *= total angle possible.*

The air flow sensor shown, as Figure 2.62, works on the principle of measuring the force exerted on the flap by the air passing through it. A calibrated coil spring exerts a counter force on the flap such that the movement of the flap is proportional to the volume of air passing through the sensor. To reduce the

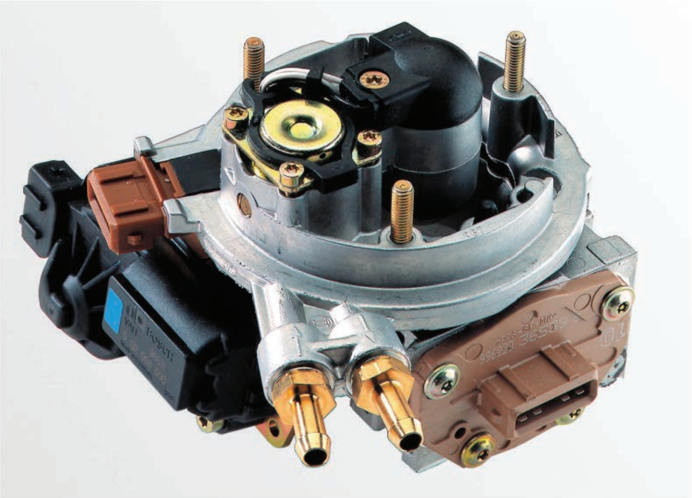


Figure 2.61 Throttle potentiometer mounted on a throttle body

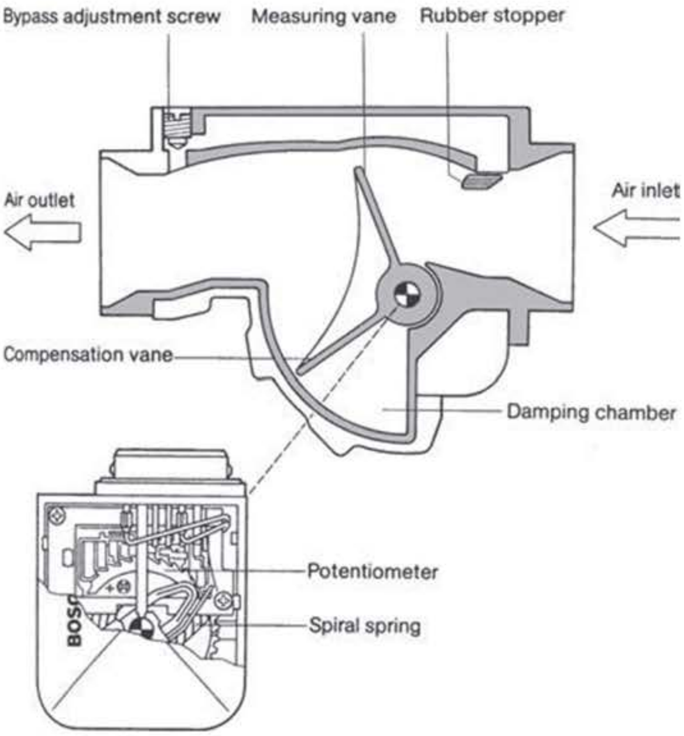


Figure 2.62 Air flow meter (vane type)

fluctuations caused by individual induction strokes a compensation flap is connected to the sensor flap. The fluctuations therefore affect both flaps and are cancelled out. Any damage due to back firing is also minimized due to this design. The resistive material used for the track is a ceramic metal mixture, which is burnt into a ceramic plate at a very high temperature. The slider potentiometer is calibrated such that the output voltage is proportional to the quantity of inducted air.

2.7.8 Accelerometer (knock sensors)

A piezoelectric accelerometer is a seismic mass accelerometer using a piezoelectric crystal to convert the force on the mass due to acceleration into an electrical output signal. The crystal not only acts as the transducer but as the

Figure 2.63 Vane type air flow meter

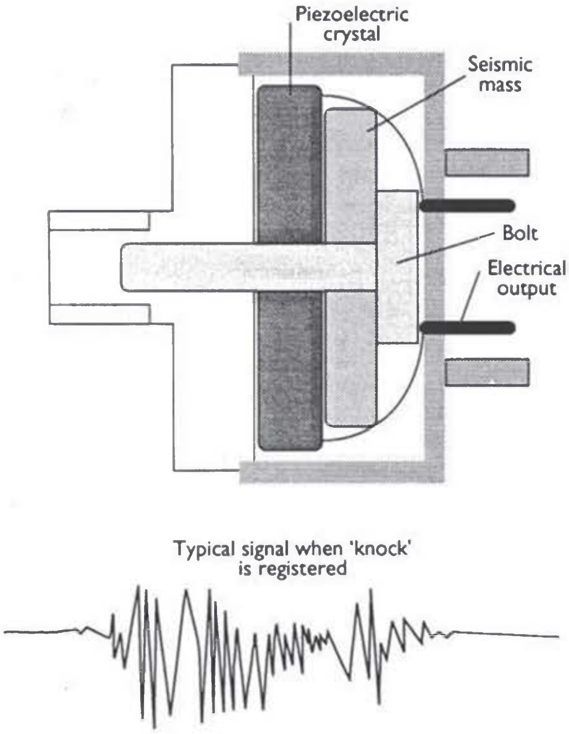


Figure 2.64 Piezoelectric accelerometer or knock sensor

suspension spring for the mass. Figure 2.64 shows a typical accelerometer (or knock sensor) for vehicle use.

The crystal is sandwiched between the body of the sensor and the seismic mass and is kept under compression by the bolt. Acceleration forces acting on the seismic mass cause variations in the amount of crystal compression and hence generate the piezoelectric voltage. The oscillations of the mass are not damped except by the stiffness of the crystal. This means that the sensor will have a very strong resonant frequency but will also be at a very high frequency (in excess of 50 kHz), giving a flat response curve in its working range up to about 15 kHz.

The natural or resonant frequency of a spring mass system is given by:

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

where: f = resonant frequency; k = spring constant (very high in this case); m = mass of the seismic mass (very low in this case).

When used as an engine knock sensor, the sensor will also detect other engine vibrations. These are kept to a minimum by only looking for 'knock' a few degrees before and after top dead centre (TDC). Unwanted signals are also filtered out electrically. A charge amplifier is used to detect the signal from this type of sensor. The sensitivity of a vehicle knock sensor is about 20 mV/g (g 9.81 m/s).

Knock sensing on petrol/gasoline engine vehicles has been used since the mid-1980s to improve performance, reduce emissions and improve economy. These sensors give a good 'flat' response over the 2–20 kHz range. The diesel knock sensor shown in Figure 2.65 works between 7 and 20 kHz. With suitable control electronics, the engine can be run near the detonation border line (DBL). This improves economy, performance and emissions.

Definition

DBL: Detonation border line (more information in Chapter 8).

Figure 2.65 Knock sensor

2.7.9 Linear variable differential transformer (LVDT)

This sensor is used for measuring displacement in a straight line (hence linear). Devices are available to measure distances of less than 0.5 mm and over 0.5 m, either side of a central position. Figure 2.66 shows the principle of the linear variable differential transducer.

The device has a primary winding and two secondary windings. The primary winding is supplied with an AC voltage and AC voltages are induced in the secondary windings by transformer action. The secondary windings are connected in series opposition so that the output of the device is the difference between their outputs. When the ferromagnetic armature is in the central position the output is zero. As the armature now moves one way or the other, the output is increased in one winding and decreased in the other, producing a voltage which, within the working range, is proportional to the displacement.

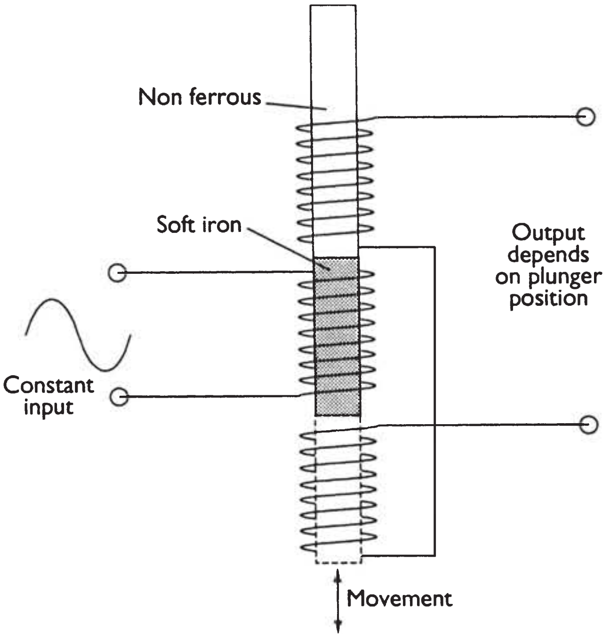


Figure 2.66 Principle of the linear variable with differential transducer

A phase sensitive detector can be used to convert the movement into a DC voltage, often 5 V. For a device moving 12 mm this gives a sensitivity of 0.42 V/mm.

LVDTs are used in some manifold pressure sensors where a diaphragm transforms changes in pressure to linear movement.

Definition

Sensitivity: A measure of the smallest signal the sensor can produce or an instrument can measure.

2.7.10 Hot wire air flow sensor

The distinct advantage of a hot wire air flow sensor is that it measures air mass flow. The basic principle is that, as air passes over a hot wire it tries to cool the wire down. If a circuit is created such as to increase the current through the wire in order to keep the temperature constant, then this current will be proportional to the air flow. A resistor is also incorporated to compensate for temperature variations. The 'hot wire' is made of platinum, is only a few millimetres long and about 70 μm thick. Because of its small size the time constant of the sensor is very short – in fact in the order of a few milliseconds. This is a great advantage as any pulsations of the air flow will be detected and reacted to in a control unit accordingly. The output of the circuit involved with the hot wire sensor is a voltage across a precision resistor. Figure 2.67 shows a Bosch hot wire air mass sensor.

Key fact

LVDTs are used in some manifold pressure sensors where a diaphragm transforms changes in pressure to linear movement.

The resistance of the hot wire and the precision resistor are such that the current to heat the wire varies between 0.5 A and 1.2 A with different air mass flow rates. High resistance resistors are used in the other arm of the bridge and so current flow is very small. The temperature compensating resistor has a resistance of about 500 Ω which must remain constant other than by way of temperature change. A platinum film resistor is used for these reasons. The compensation resistor can cause the system to react to temperature changes within about 3 s.

The output of this device can change if the hot wire becomes dirty. Heating the wire to a very high temperature for 1 s every time the engine is switched off prevents this by burning off any contamination. In some air mass sensors a variable resistor is provided to set the idle mixture.

Figure 2.67 Hot wire mass air flow meter
(Source: Bosch Media)

Figure 2.68 Hot film air mass flow meter

2.7.11 Thin film air flow sensor

The thin film air flow sensor is similar to the hot wire system. Instead of a hot platinum wire a thin film of nickel is used. The response time of this system is even shorter than the hot wire. Figure 2.68 shows this sensor in more detail.

2.7.12 Vortex flow sensor

Figure 2.69 shows the principle of a vortex flow sensor. It has a bluff body, which partially obstructs the flow. Vortices form at the down-stream edges of the bluff body at a frequency that is linearly dependent on the flow velocity. Detection of the vortices provides an output signal whose frequency is proportional to flow velocity. Detection of the vortices can be by an ultrasonic transmitter and receiver that will produce a proportional square wave output. The main advantage of this device is the lack of any moving parts, thus eliminating problems with wear.

For a vortex flow sensor to work properly, the flow must be great enough to be turbulent, but not so high as to cause bubbles when measuring fluid flow. As a rough guide, the flow should not exceed 50 m/s.

When used as an engine air flow sensor, this system will produce an output frequency of about 50 Hz at idle speed and in excess of 1 kHz at full load.

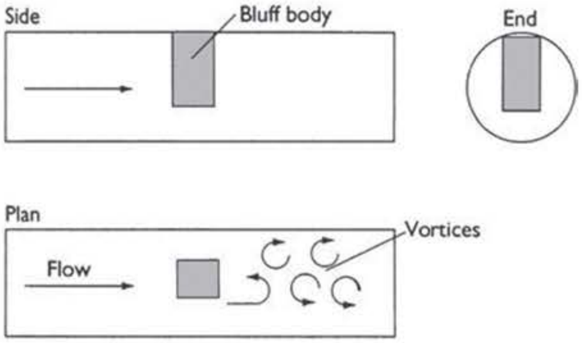


Figure 2.69 Principle of a vortex flow sensor

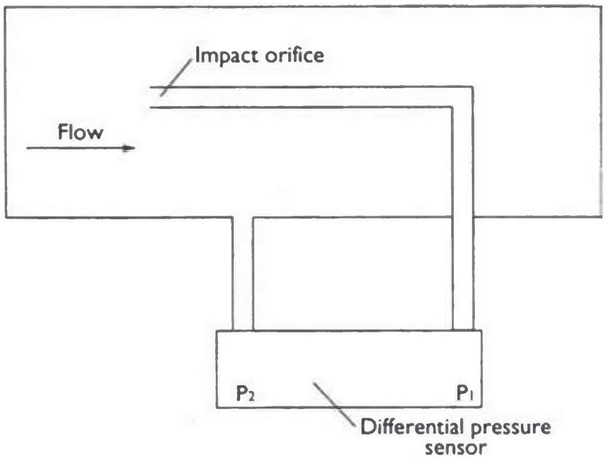


Figure 2.70 Pitot tube and differential pressure sensor for air flow sensing

2.7.13 Pitot tube

A Pitot tube air flow sensor is a very simple device. It consists of a small tube open to the air flow such that the impact of the air will cause an increase in pressure in the tube compared with the pressure outside the tube. This same system is applied to aircraft to sense air speed when in flight. The two tubes are connected to a differential pressure transducer such as a variable capacitance device. P_1 and P_2 are known as the impact and static pressures, respectively. Figure 2.70 shows a Pitot tube and differential pressure sensor used for air flow sensing.

2.7.14 Turbine fluid flow sensor

Using a turbine to measure fluid flow is an invasive form of measurement. The act of placing a device in the fluid will affect the flow rate. This technique however is still used as, with careful design, the invasion can be kept to a minimum. Figure 2.71 shows a typical turbine flow sensor.

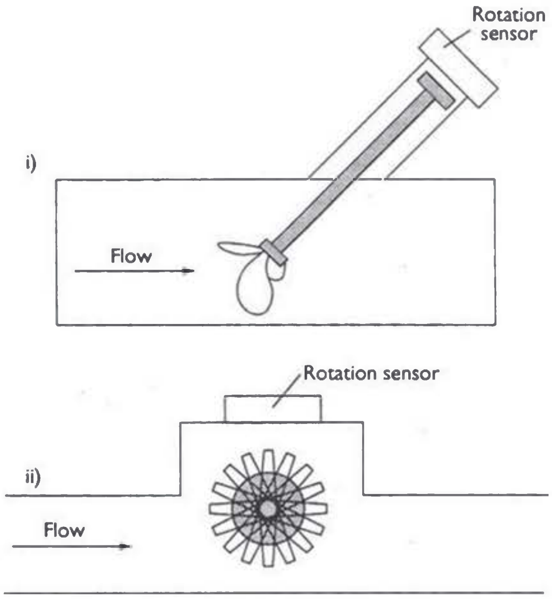


Figure 2.71 Turbine flow centre

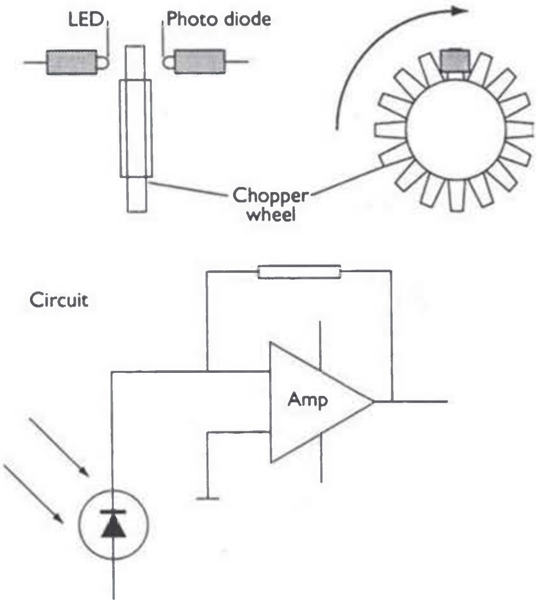


Figure 2.72 Optical sensor

The output of the turbine, rotational speed proportional to flow rate, can be converted to an electrical signal in a number of ways. Often an optical sensor is used as described under the next heading.

2.7.15 Optical sensors

An optical sensor for rotational position is a relatively simple device. The optical rotation sensor and circuit shown in Figure 2.72 consist of a photo-transistor as a detector and a light emitting diode light source. If the light is focused to a very narrow beam then the output of the circuit shown will be a square wave with frequency proportional to speed.

2.7.16 Oxygen sensors

The vehicle application for an oxygen sensor is to provide a closed loop feedback system for engine management control of the air–fuel ratio. The amount of oxygen sensed in the exhaust is directly related to the mixture strength, or air–fuel ratio. The ideal air–fuel ratio of 14.7 : 1 by mass is known as a lambda (λ) value of one. Exhaust gas oxygen (EGO) sensors are placed in the exhaust pipe near to the manifold to ensure adequate heating. The sensors operate reliably at temperatures over 300 °C. In some cases, a heating element is incorporated to ensure this temperature is reached quickly. This type of sensor is known as a heated exhaust gas oxygen sensor, or HEGO for short. The heating element (which consumes about 10 W) does not operate all the time, which ensures that the sensor does not exceed 850 °C – the temperature at which damage may occur to the sensor. It is for this reason that the sensors are not often fitted directly in the exhaust manifold. Figure 2.73 shows a zirconia type exhaust gas oxygen sensor.

The main active component of most types of oxygen sensors is zirconium dioxide (ZrO_2). This ceramic is housed in gas permeable electrodes of platinum. A further ceramic coating is applied to the side of the sensor exposed to the exhaust gas as a protection against residue from the combustion process. The principle of

Figure 2.73 Lambda sensor

operation is that, at temperatures in excess of 300 °C, the zirconium dioxide will conduct the oxygen ions. The sensor is designed to be responsive very close to a lambda value of one. As one electrode of the sensor is open to a reference value of atmospheric air, a greater quantity of oxygen ions will be present on this side. Due to electrolytic action these ions permeate the electrode and migrate through the electrolyte (ZrO_2). This builds up a charge rather like a battery. The size of the charge is dependent on the oxygen percentage in the exhaust. A voltage of 400 mV is the normal figure produced at a lambda value of one.

The closely monitored closed loop feedback of a system using lambda sensing allows very accurate control of engine fuelling. Close control of emissions is therefore possible.

Key fact

The main active component of most types of oxygen sensors is zirconium dioxide (ZrO_2).

2.7.17 Light sensors

A circuit employing a light sensitive resistor is shown in Figure 2.74. The circuit can be configured to switch on or off in response to an increase or decrease in

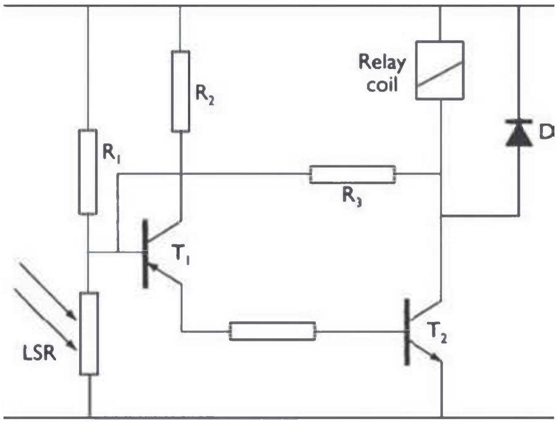


Figure 2.74 Light sensitive resistor circuit

light. Applications are possible for self-dipping headlights, a self-dipping interior mirror, or parking lights that will automatically switch on at dusk.

2.7.18 Thick-film air temperature sensor

The advantage which makes a nickel thick-film thermistor ideal for inlet air temperature sensing is its very short time constant. In other words its resistance varies very quickly with a change in air temperature. The response of a thick film sensor is almost linear. It has a sensitivity of about 2 ohms/°C and, as with most metals, it has a positive temperature coefficient (PTC) characteristic.

2.7.19 Methanol sensor

In the move towards cleaner exhausts, one idea is to use mixed fuels. Methanol is one potential fuel that can be mixed with petrol. The problem is that petrol has a different stoichiometric air requirement to methanol.

An engine management system can be set for either fuel or a mixture of the fuels. However, the problem with mixing is that the ratio will vary. A special sensor is needed to determine the proportion of methanol, and once fitted this sensor will make it possible to operate the vehicle on any mixture of petrol and methanol.

The methanol sensor (Figure 2.75) is based on the dielectric principle. The measuring cell is a capacitor filled with fuel and the methanol content is calculated from its capacitance. Two further measurements are taken – the temperature of the fuel and its conductance. These correction factors ensure cross-sensitivity (a kind of double checking) and the measurement error is therefore very low. The sensor can be fitted to the fuel line so the data it provides to the ECU are current and reliable. The control unit can then adapt the fuelling strategy to the fuel mix currently in use. Some further development is taking place but this sensor looks set to play a major part in allowing the use of alternative fuels in the near future.

2.7.20 Rain sensor

Rain sensors are used to switch on wipers automatically. Most work on the principle of reflected light. The device is fitted inside the windscreen and light from an LED is reflected back from the outer surface of the glass. The amount of light reflected changes if the screen is wet, even with a few drops of rain. Figure 2.76 shows the principle of operation and Figure 2.77 shows a typical sensor.

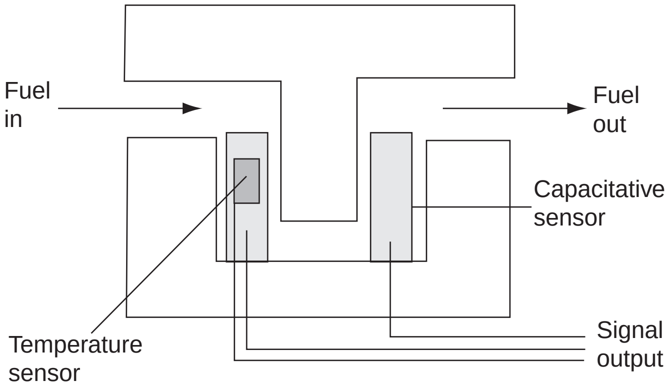


Figure 2.75 Methanol sensor

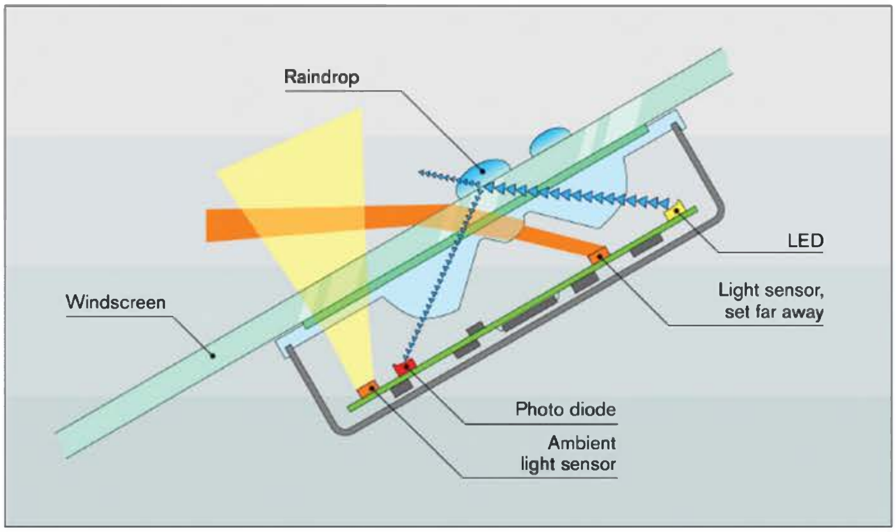


Figure 2.76 Rain sensor (Source: Bosch Media)

Figure 2.77 Rain sensor package

2.7.21 Oil sensor

An interesting sensor used to monitor oil quality is now available; the type shown in Figure 2.78 works by monitoring changes in the dielectric constant of the oil. The dielectric constant increases as antioxidant additives in the oil deplete. The value also rapidly increases if coolant contaminates the oil. The sensor output increases as the dielectric constant increases.

Figure 2.78 Oil quality sensor

2.7.22 Dynamic vehicle position sensors

These sensors are used for systems such as active suspension, stability control and general systems where the movement of the vehicle is involved. Most involve the basic principle of an accelerometer; that is, a ball hanging on a string or a seismic mass acting on a sensor.

An accelerometer is available that senses a number of directions. The sensor shown in Figure 2.79 can be constructed as part of an ECU. This sensor is used for ride control systems such as ESP.

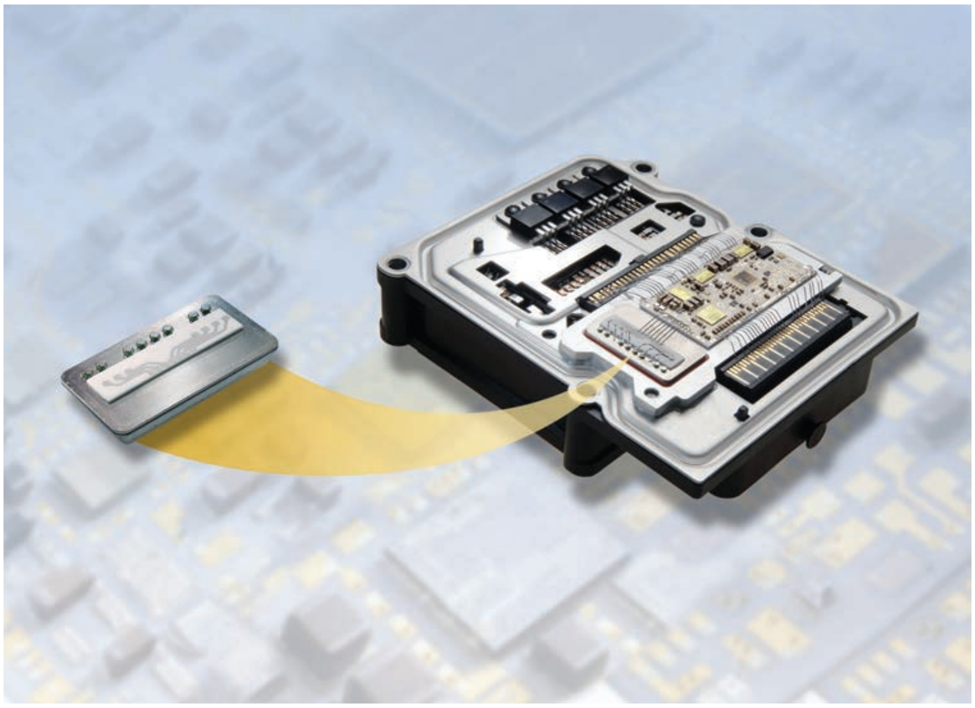


Figure 2.79 Sensors built in to an ECU

2.7.23 Summary

The above brief look at various sensors hardly scratches the surface of the number of types, and the range of sensors available for specific tasks. The subject of instrumentation is now a science in its own right. The overall intention of this section has been to highlight some of the problems and solutions to the measurement of variables associated with vehicle technology.

Sensors used by motor vehicle systems are following a trend towards greater integration of processing power in the actual sensor. Four techniques are considered, starting with the conventional system. Figure 2.80 shows each level of sensor integration in a block diagram form.

Key fact

Sensors used by motor vehicle systems are following a trend towards greater integration of processing power in the actual sensor.

Conventional

An analogue sensor in which the signal is transmitted to the ECU via a simple wire circuit. This technique is very susceptible to interference.

Integration level 1

Analogue signal processing is now added to the sensor, this improves the resistance to interference.

Integration level 2

At the second level of integration, analogue to-digital conversion is also included in the sensor. This signal is made bus compatible (CAN for example) and hence becomes interference proof.

Integration level 3

The final level of integration is to include ‘intelligence’ in the form of a microcomputer as part of the sensor. The digital output will be interference proof. This level of integration will also allow built in monitoring and diagnostic ability.

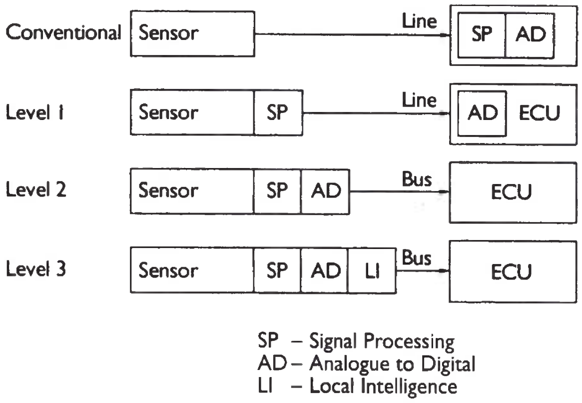


Figure 2.80 Block diagram of four types of sensor and their different levels of integration

These types of sensor are very expensive at the time of writing but the price is falling and will continue to do so as more use is made of the 'intelligent sensor'.

2.8 Actuators

2.8.1 Introduction

There are many ways of providing control over variables in and around the vehicle. 'Actuators' is a general term used here to describe a control mechanism. When controlled electrically actuators will work either by the thermal or magnetic effect. In this section, the term actuator will be used to mean a device that converts electrical signals into mechanical movement. This section is not written with the intention of describing all available types of actuator. Its intention is to describe some of the principles and techniques used in controlling a wide range of vehicle systems.

Definition

Actuator: A mechanical device for moving or controlling a mechanism or system. It is usually operated by electricity but other sources of energy may be used.

2.8.2 Solenoid actuators

The basic operation of solenoid actuators is very simple. The term 'solenoid' means: 'many coils of wire wound onto a hollow tube'. However, the term is often misused, but has become so entrenched that terms like 'starter solenoid' – when really it is starter relay – are in common use.

A good example of a solenoid actuator is a fuel injector. Figures 2.81 and 2.82 show two typical examples. When the windings are energized the armature is attracted due to magnetism and compresses the spring. In the case of a fuel injector, the movement is restricted to about 0.1 mm. The period that an injector remains open is very small – under various operating conditions, between 1.5 and 10 ms is typical. The time it takes an injector to open and close is also critical for accurate fuel metering. Further details about injection systems are discussed in Chapters 9 and 10.

The reaction time for a solenoid-operated device, such as a fuel injector, depends very much on the inductance of the winding. Figure 2.83 shows a graph of solenoid-operated actuator variables.

A suitable formula to show the relationship between some of the variables is as follows:

$$i = \frac{V}{R}\left(1 - e^{-Rt/L}\right)$$

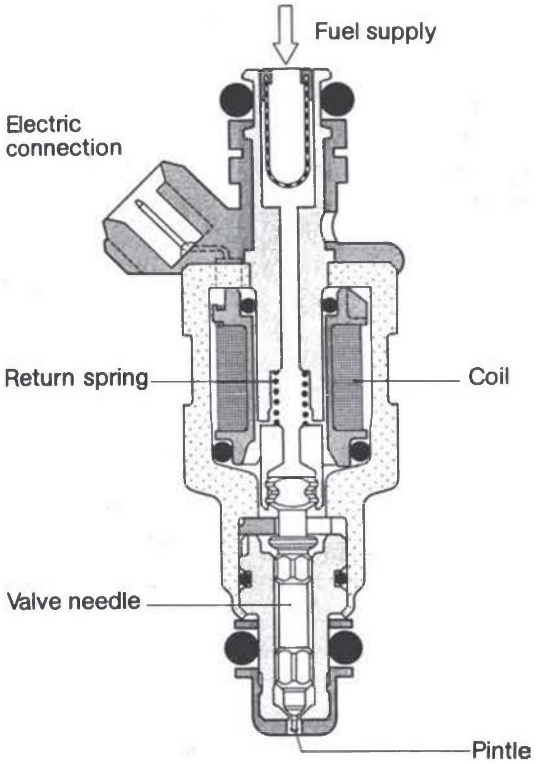


Figure 2.81 Fuel Injector components

Figure 2.82 Direct injection injector (Source: Bosch Media)

where: i = instantaneous current in the winding, V = supply voltage, R = total circuit resistance, L = inductance of the injector winding, t = time current has been flowing, e = base of natural logs.

The resistance of commonly used manifold injectors is about 16 Ω, other types vary but a few ohms is typical. Some systems use ballast resistors in series with the fuel injectors. This allows lower inductance and resistance operating windings to be used, thus speeding up reaction time. Other types of solenoid actuators, for example door lock actuators, have less critical reaction times. However, the basic principle remains the same.

2.8.3 EGR valve

One interesting in actuator technology is the rotary electric exhaust gas recirculation (EEGR) valve for use in diesel engine applications. This device is

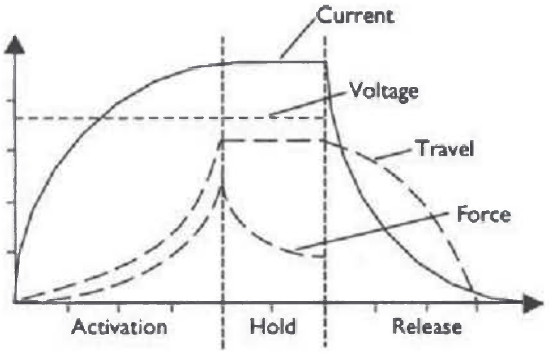


Figure 2.83 Solenoid operated actuator variables

shown in Figure 2.84. The main claims for this valve are its self-cleaning action, accurate gas flow control and its reaction speed.

Figure 2.84 Rotary electric exhaust gas recirculation valve (Source: Delphi Media)

2.8.4 Motorized actuators

Permanent magnet electric motors are used in many applications and are very versatile. The output of a motor is, of course, rotation, but this can be used in many ways. If the motor drives a rotating 'nut' through which a plunger is fitted, and on which there is a screw thread, the rotary action can easily be converted to linear movement. In most vehicle applications the output of the motor has to be geared down, this is to reduce speed and increase torque. Permanent magnet motors are almost universally used now in place of older and less practical motors with field windings. Some typical examples of where these motors are used are:

- windscreen wipers
- windscreen washers
- headlight lift
- electric windows
- electric sun roof
- electric aerial operation
- seat adjustment
- mirror adjustment
- headlight washers
- headlight wipers
- fuel pumps
- ventilation fans.

Key fact

In most vehicle applications the output of a motor has to be geared down to reduce speed and increase torque.

One disadvantage of simple motor actuators is that no direct feedback of position is possible. This is not required in many applications; however, in some cases, such as seat adjustment when a 'memory' of the position may be needed, a variable resistor type sensor can be fitted to provide feedback. Three typical motor actuators are shown in Figure 2.85.

A rotary idle actuator is shown in Figures 2.86 and 2.87. This device is used to control idle speed by controlling air bypass. There are two basic types in common

Figure 2.85 Seat adjustment motor

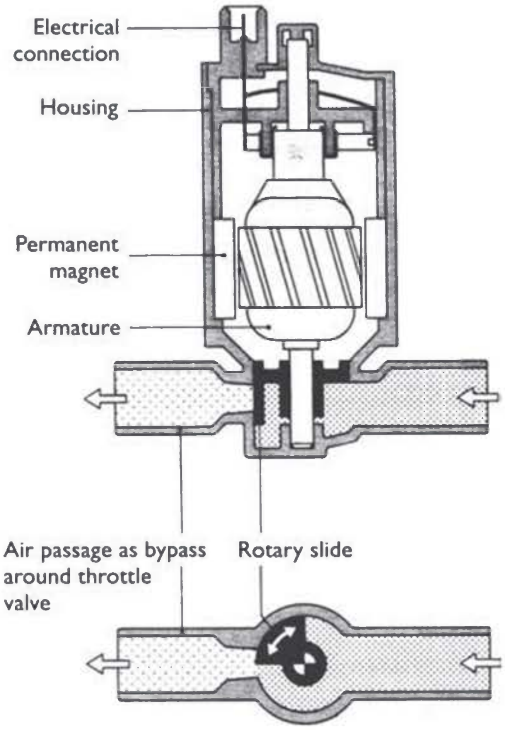


Figure 2.86 Rotary idle actuator components

Figure 2.87 Rotary idle actuator

Key fact

The on/off ratio or duty cycle of the square wave supply will determine the average valve open time and hence idle speed.

use. These are single winding types, which have two terminals, and double winding types, which have three terminals. Under ECU control, the motor is caused to open and close a shutter, thus controlling air bypass. These actuators only rotate about 90° to open and close the valve. As these are permanent magnet motors, the term ‘single or double windings’ refers to the armature.

The single winding type is fed with a square wave signal causing it to open against a spring and then close again, under spring tension. The on/off ratio or duty cycle of the square wave will determine the average valve open time and hence idle speed.

With the double winding type the same square wave signal is sent to one winding but the inverse signal is sent to the other. As the windings are wound in opposition to each other if the duty cycle is 50% then no movement will take place. Altering the ratio will now cause the shutter to move in one direction or the other.

2.8.5 Stepper motors

Stepper motors are becoming increasingly popular as actuators in motor vehicles and in many other applications. This is mainly because of the ease with which they can be controlled by electronic systems. Stepper motors fall into three distinct groups (Figure 2.88):

1. variable reluctance motors
2. permanent magnet motors
3. hybrid motors.

Variable reluctance motors rely on the physical principle of maximum flux. A number of windings are set in a circle on a toothed stator. The rotor also has teeth and is made of a permeable material. Note in this example that the rotor

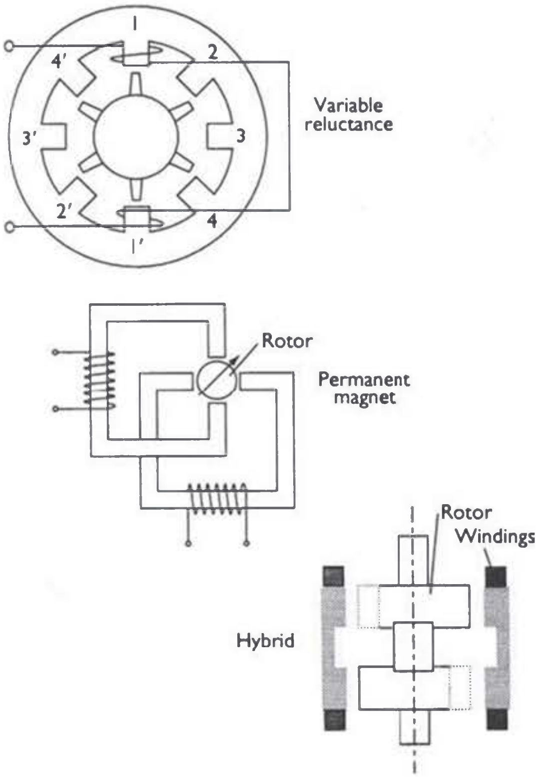


Figure 2.88 Basic principle of variable reluctance, permanent magnet and hybrid stepper motors

has two teeth less than the stator. When current is supplied to a pair of windings of one phase, the rotor will line up with its teeth positioned such as to achieve maximum flux. It is now simply a matter of energizing the windings in a suitable order to move the rotor. For example, if phase four is energized, the motor will 'step' once in a clockwise direction. If phase two is energized the step would be anti-clockwise.

These motors do not have a very high operating torque and have no torque in the non-excited state. They can, however, operate at relatively high frequencies. The step angles are usually 15°, 7.5°, 1.8° or 0.45°.

Permanent magnet stepper motors have a much higher starting torque and also have a holding torque when not energized. The rotor is now, in effect, a permanent magnet. In a variable reluctance motor the direction of current in the windings does not change; however, it is the change in direction of current that causes the permanent magnet motor to step. Permanent magnet stepper motors have step angles of 45°, 18°, 15° or 7.5°. Because of their better torque and holding properties, permanent magnet motors are becoming increasingly popular. For this reason, this type of motor will be explained in greater detail.

The hybrid stepper motor as shown in Figure 2.89 is, as the name suggests, a combination of the previous two motors. These motors were developed to try and combine the high speed operation and good resolution of the variable reluctance type with the better torque properties of the permanent magnet motor. A pair of toothed wheels is positioned on either side of the magnet. The teeth on the 'North' and 'South' wheels are offset such as to take advantage of the variable reluctance principle but without losing all the torque benefits. Step angles of these motors are very small: 1.8°, 0.75° or 0.36°.

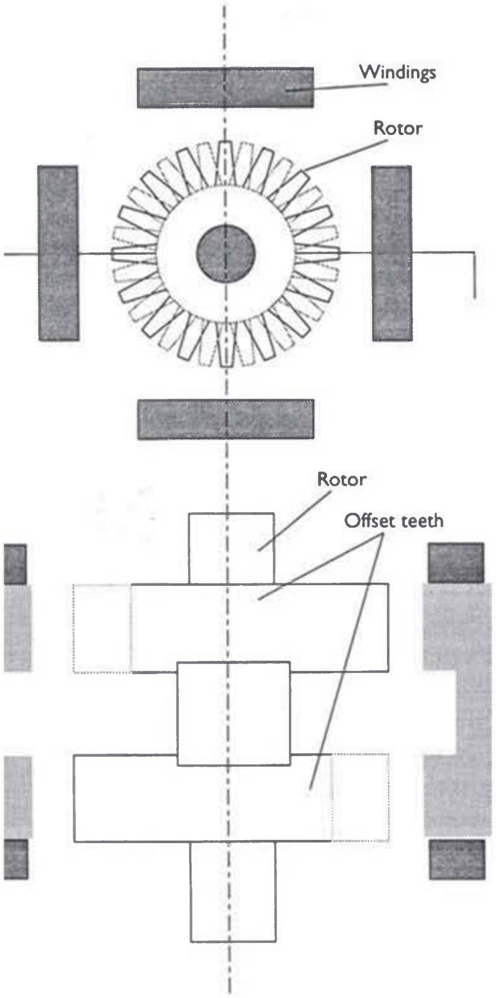


Figure 2.89 Stepper motor

All of the above-mentioned types of motor have been, and are being, used in various vehicle applications. These applications range from idle speed air bypass and carburettor choke control to speedometer display drivers.

Let's look now in more detail at the operation and construction of the permanent magnet stepper motor. The most basic design for this type of motor comprises two double stators displaced by one pole pitch. The rotor is often made of barium-ferrite in the form of a sintered annular magnet. As the windings shown in Figure 2.90 are energized first in one direction then the other, the motor will rotate in 90° steps. The step angle is simply 360° divided by the number of stator poles. Half steps can be achieved by switching off a winding before it is reversed. This will cause the rotor to line up with the remaining stator poles and implement a half step of 45°. The direction of rotation is determined by the order in which the windings are switched on, off or reversed. Figure 2.90 shows a four-phase stepper motor and circuit.

Impulse sequence graphs for two phase stepper motors are shown in Figure 2.91. The first graph is for full steps, and the second graph for implementing half steps.

The main advantage of a stepper motor is that feedback of position is not required. This is because the motor can be indexed to a known starting point and then a calculated number of steps will move the motor to any suitable position.

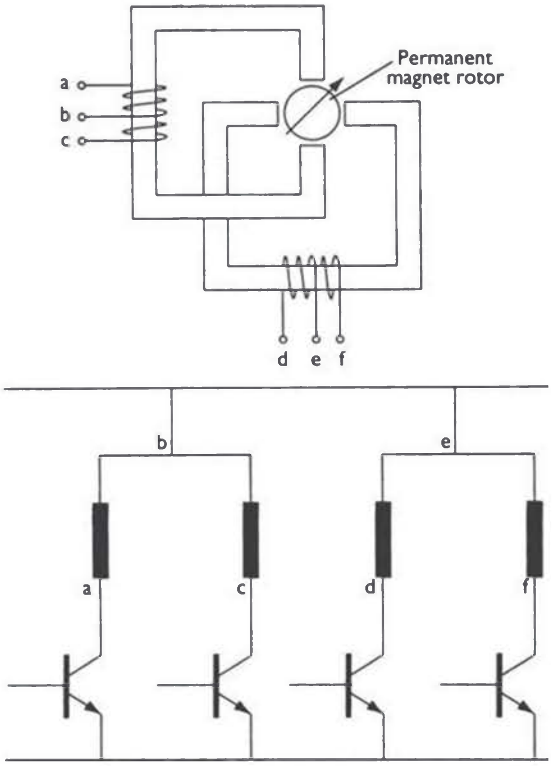


Figure 2.90 Four-phase stepper motor and circuit

The calculations often required for stepper applications are listed below:

$\alpha = 360/z$

$z = 360/\alpha$

$fz= (n.z)/60$

$n = (fz.60)/z$

$w = (fz.2\pi)/z$

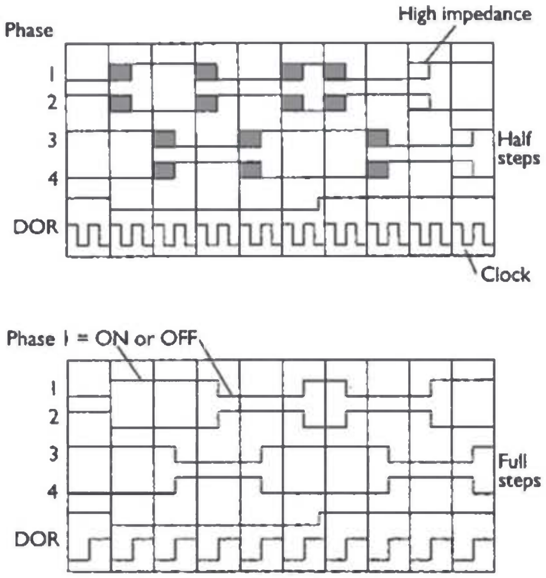


Figure 2.91 Impulse sequence graphs for two-phase stepper motors: the first graph is for half steps, the second for implementing full steps

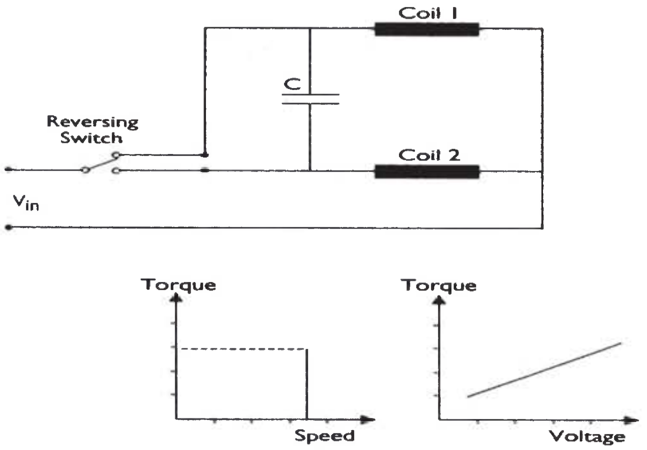


Figure 2.92 Reversing synchronous motor and circuit and its speed torque characteristic

where: α *= step angle, n = revolutions per minute, w = angular velocity, fz= step frequency, z = steps per revolution.*

2.8.6 Synchronous motors

Synchronous motors are used when a drive is required that must be time synchronized. They always rotate at a constant speed, which is determined by the system frequency and the number of pole pairs in the motor.

$$n = (f \times 60)/p$$

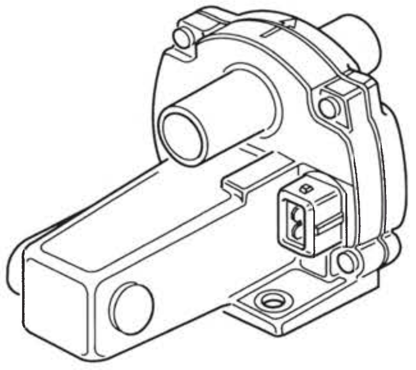
where: n = rpm, f = frequency, p = number of pole pairs.

Figure 2.92 shows a reversing synchronous motor and its circuit together with the speed torque characteristic. This shows a constant speed and a break off at maximum torque. Maximum torque is determined by supply voltage.

2.8.7 Thermal actuators

An example of a thermal actuator is the movement of a traditional-type fuel or temperature gauge needle (see Chapter 13). A further example is an auxiliary air device used on many earlier fuel injection systems. The principle of this device is shown in Figure 2.93. When current is supplied to the terminals, a heating element operates and causes a bimetallic strip to bend, which closes a simple valve.

The main advantage of this type of actuator, apart from its simplicity, is that if placed in a suitable position its reaction time will vary with the temperature of its surroundings. This is ideal for applications such as fast idle on cold starting control, where once the engine is hot no action is required from the actuator.



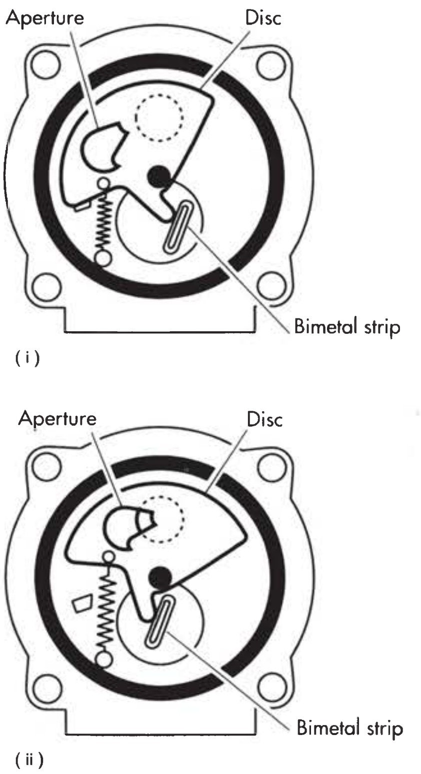


Figure 2.93 Diagram showing operation of an earlier type of extra air valve: (i) bypass channel closed; (ii) bypass channel partially open

2.9 Testing electronic components, sensors and actuators

2.9.1 Introduction

Individual electronic components can be tested in a number of ways but a digital multimeter or an oscilloscope is normally the best option. Table 2.3 suggests some methods of testing components removed from the circuit.

Table 2.3 Electronic component testing

Component	Test method
Resistor	Measure the resistance value with an ohmmeter and compare this to the value written or colour coded on the component.
Capacitor	A capacitor can be difficult to test without specialist equipment but try this: Charge the capacitor up to 12 V and connect it to a digital voltmeter. As most digital meters have an internal resistance of about 10 MΩ, calculate the expected discharge time ($T = 5CR$) and see if the device complies! A capacitor from a contact breaker ignition system should take about 5 seconds to discharge in this way.
Inductor	An inductor is a coil of wire so a resistance check is the best method to test for continuity.
Diode	Many multimeters have a diode test function. If so the device should read open circuit in one direction, and about 0.4 to 0.6 V in the other direction. This is its switch on voltage. If no meter is available with this function then wire the diode to a battery via a small bulb, it should light with the diode one way and not the other.
LED	Most LED's can be tested by connecting them to a 1.5 V battery. Note the polarity though, the longest leg or the flat side of the case is negative.
Transistor (bipolar)	Some multimeters even have transistor testing connections but if not available the transistor can be connected into a simple circuit as in Figure 2.94 and voltage tests carried out as shown. This also illustrates a method of testing electronic circuits in general. It is fair to point out that without specific data it is difficult for the non-specialist to test unfamiliar circuit boards. It's always worth checking for obvious breaks and dry joints though!
Digital components	A logic probe can be used. This is a device with a very high internal resistance so it does not affect the circuit under test. Two different coloured lights are used, one glows for a 'logic 1' and the other for 'logic 0'. Specific data is required in most cases but basic tests can be carried out.

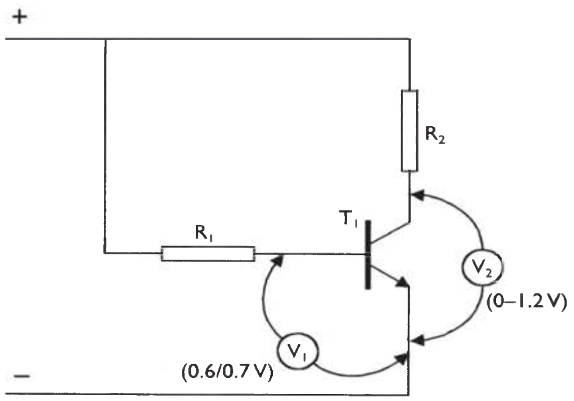


Figure 2.94 Transistor test

2.9.2 Testing sensors

Testing sensors to diagnose faults is usually a matter of measuring their output signal. In some cases the sensor will produce this on its own (an inductive sensor for example). In other cases, it will be necessary to supply the correct voltage to the device to make it work (Hall sensor for example). In this case, it is normal to check the vehicle circuit is supplying the voltage before proceeding to test the sensor.

The following table lists some common sensors together with suggested test methods (correct voltage supply is assumed):

Figure 2.95 Lambda sensors performance can be tested with a voltmeter or ideally a scope

Table 2.4 Sensor testing

Sensor	Test method
Inductive (reluctance)	A simple resistance test is good. Values vary from about 800 to 1200 Ω. The 'sine wave' output can be viewed on a 'scope' or measured with an AC voltmeter.
Hall effect	The square wave output can be seen on a scope or the voltage output measured with a DC voltmeter. This varies between O to 8 V for a Hall sensor used in a distributor as the chip is magnetised or not.
Thermistor	Most thermistors have a negative temperature coefficient (NTC). This means the resistance falls as temperature rises. A resistance check with an ohmmeter should give readings broadly as follows: 0 °C = 4500 Ω, 20 °C = 1200 Ω and 100 °C = 200 Ω.
Flap air flow	The main part of this sensor is a variable resistor. If the supply is left connected then check the output on a DC voltmeter. The voltage should change smoothly from about 0 to the supply voltage (often 5 V).
Hot wire air flow	This sensor includes some electronic circuits to condition the signal from the hot wire. The normal supply is either 5 or 12 V. The output should change between about 0 and 5 V as the air flow changes.
Throttle potentiometer	This sensor is a variable resistor. If the supply is left connected then check the output on a DC voltmeter. The voltage should change smoothly from about 0 to the supply voltage (often 5 V). If no supply then check the resistance, again it should change smoothly.
Oxygen (lambda)	The lambda sensor produces its own voltage a bit like a battery. This can be measured with the sensor connected to the system. The voltage output should vary smoothly between 0.2 and 0.8 V as the mixture is controlled by the ECU.
Pressure	The normal supply to an externally mounted manifold absolute pressure (MAP) sensor is 5 V. The output should change between about 0 and 5 V as the manifold pressure changes. As a rough guide 2.5 V at idle speed.

2.9.3 Testing actuators

Testing actuators is simple, as many are operated by windings. The resistance can be measured with an ohmmeter. A good tip is that where an actuator has more than one winding (a stepper motor for example), the resistance of each should be about the same. Even if the expected value is not known, it is likely that if the windings all read the same then the device is in order.

With some actuators, it is possible to power them up from the vehicle battery. A fuel injector should click for example, and a rotary air bypass device should rotate about half a turn. Be careful with this method as some actuators could be damaged.

CHAPTER 3

Tools and equipment

3.1 Basic equipment

3.1.1 Introduction

Diagnostic techniques are very much linked to the use of test equipment. In other words you must be able to interpret the results of tests. In most cases this involves comparing the result of a test to the reading given in a data book or other source of information. By way of an introduction, the following table lists some of the basic words and descriptions relating to tools and equipment.

3.1.2 Basic hand tools

You will not learn how to use tools by reading a book; it is clearly a very practical skill. However, you can follow the recommendations made here and by the manufacturers. Even the range of basic hand tools, is now quite daunting and very expensive. It is worth repeating the general advice given by Snap-on for the use of hand tools:

- Only use a tool for its intended purpose

Figure 3.1 Snap-on tool kit

Table 3.1 Tools and equipment

Hand tools	Spanners and hammers and screwdrivers and all the other basic bits!
Special tools	A collective term for items not held as part of a normal tool kit. Or items required for just one specific job.
Test equipment	In general, this means measuring equipment. Most tests involve measuring something and comparing the result of that measurement to data. The devices can range from a simple ruler to an engine analyser.
Dedicated test equipment	Some equipment will only test one specific type of system. The large manufacturers supply equipment dedicated to their vehicles. For example, a diagnostic device which plugs in to a certain type of fuel injection ECU.
Accuracy	Careful and exact, free from mistakes or errors and adhering closely to a standard.
Calibration	Checking the accuracy of a measuring instrument.
Serial port	A connection to an electronic control unit, a diagnostic tester or computer for example. Serial means the information is passed in a 'digital' string like pushing black and white balls through a pipe in a certain order.
Code reader or scanner	This device reads the 'black and white balls' mentioned above or the on-off electrical signals, and converts them in to language we can understand.
Combined diagnostic and information system	Usually now PC based these systems can be used to carry out tests on vehicle systems and they also contain an electronic workshop manual. Test sequences guided by the computer can also be carried out.
Oscilloscope	The main part of 'scope' is the display, which is like a TV or computer screen. A scope is a voltmeter but instead of readings in numbers it shows the voltage levels by a trace or mark on the screen. The marks on the screen can move and change very fast allowing us to see the way voltages change.

- Always use the correct size tool for the job you are doing
- Pull a wrench rather than pushing whenever possible
- Do not use a file or similar, without a handle
- Keep all tools clean and replace them in a suitable box or cabinet
- Do not use a screwdriver as a pry bar
- Always follow manufacturers recommendations (you cannot remember everything)
- Look after your tools and they will look after you!

3.1.3 Accuracy of test equipment

Accuracy can be described in a number of slightly different ways:

- Careful and exact
- Free from mistakes or errors
- Precise
- Adhering closely to a standard.

Consider measuring a length of wire with a steel rule. How accurately could you measure it? To the nearest 0.5 mm perhaps? This raises a number of issues: Firstly you could make an error reading the ruler. Secondly, why do we need to know the length of a bit of wire to the nearest 0.5 mm? Thirdly the ruler may have stretched and not give the correct reading!

The first and second of these issues can be dispensed with by knowing how to read the test equipment correctly and also knowing the appropriate level of accuracy required. A micrometer for a plug gap? A ruler for valve clearances? I think you get the idea. The accuracy of the equipment itself is another issue.

Accuracy is a term meaning how close the measured value of something is, to its actual value. For example, if a length of about 30 cm is measured with an ordinary wooden ruler, then the error may be up to 1 mm too high or too low. This is quoted as an accuracy of ±1 mm. This may also be given as a percentage which in this case would be 0.33%.

Resolution or in other words the 'fineness', with which a measurement can be made, is related to accuracy. If a steel ruler was made to a very high standard but only had markings of one per centimetre it would have a very low resolution even though the graduations were very accurate. In other words the equipment is accurate but your reading will not be!

To ensure instruments are, and remain accurate, there are just two simple guidelines:

1. Look after the equipment, a micrometer thrown on the floor will not be accurate.
2. Ensure instruments are calibrated regularly – this means being checked against known good equipment.

Here is a summary of the steps to ensure a measurement is accurate:

Definition

Accuracy: How close the measured value of something is to the actual value.

Definition

Resolution: The 'fineness' with which a measurement can be made.

Table 3.2 Accurate measurement process

Step	Example
Decide on the level of accuracy required.	Do we need to know that the battery voltage is 12.6 V or 12.635 V?
Choose the correct instrument for the job.	A micrometer to measure the thickness of a shim.
Ensure the instrument has been looked after and calibrated when necessary.	Most instruments will go out of adjustment after a time. You should arrange for adjustment at regular intervals. Most tool suppliers will offer the service or in some cases you can compare older equipment to new stock.
Study the instructions for the instrument in use and take the reading with care. Ask yourself if the reading is about what you expected.	Is the piston diameter 70.75 mm or 170.75 mm?
Make a note if you are taking several readings.	Don't take a chance write it down.

3.1.4 Multimeters

An essential tool for working on vehicle electrical and electronic systems is a good digital multimeter (often referred to as a DMM). Digital meters are most suitable for accuracy of reading as well as available facilities.

The following list of functions broadly in order, starting from essential to desirable should be considered:

Definition

DMM: Digital multimeter.

Table 3.3 Multimeter functions

Function	Range	Accuracy
DC Voltage	500 V	0.3%
DC Current	10 A	1.0%
Resistance	0 to 10 MΩ	0.5%
AC Voltage	500 V	2.5%
AC Current	10 A	2.5%
Dwell	3,4,5,6,8 cylinders	2.0%
rpm	10,000 rpm	0.2%
Duty cycle	% on/off	0.2%/kHz
Frequency	over 100 kHz	0.01%
Temperature	> 9000 C	0.3% +30 C
High current clamp	1000 A (DC)	Depends on conditions
Pressure	3 bar	10.0% of standard scale

A way of determining the quality of a digital multimeter as well as by the facilities provided, is to consider the following

- accuracy
- loading effect of the meter
- protection circuits.

The loading effect is a consideration for any form of measurement. With a multimeter this relates to the internal resistance of the meter. It is recommended

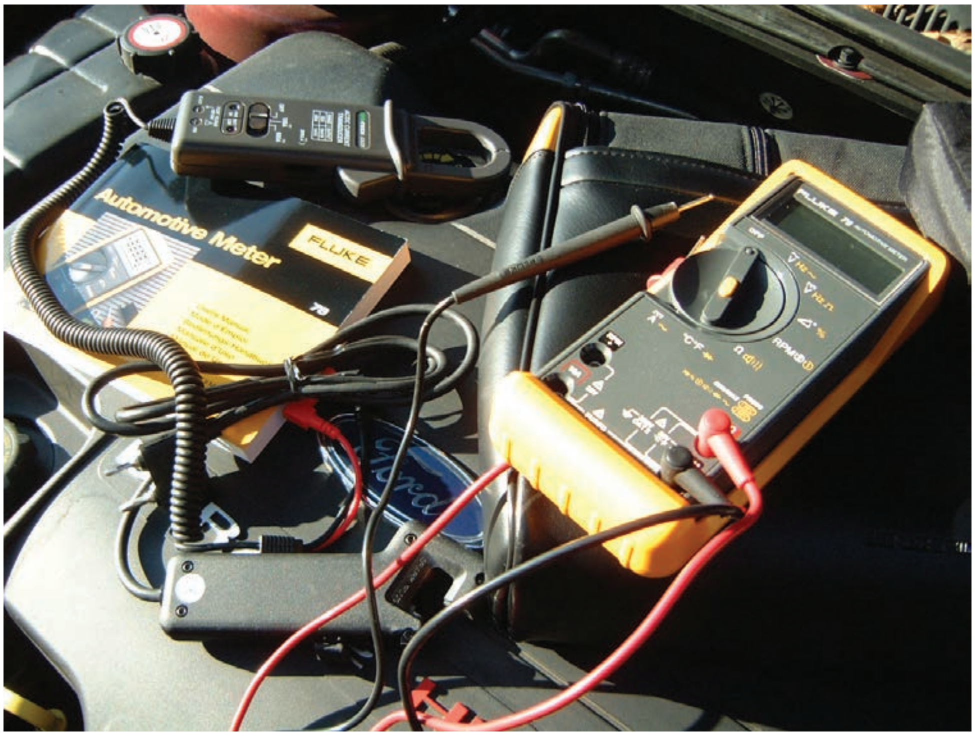


Figure 3.2 Multimeter and accessories

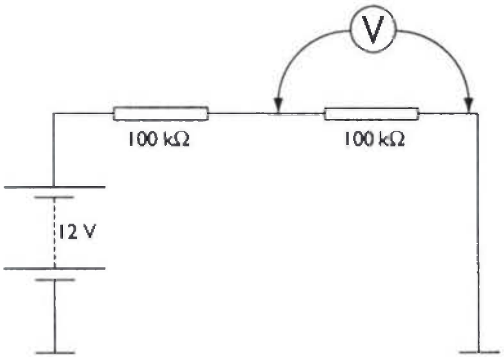


Figure 3.3 Loading effect of a mete

that the internal resistance of a meter should be a minimum of 10 MΩ. This not only ensures greater accuracy but also prevents the meter damaging sensitive circuits.

Figure 3.3 shows two equal resistors connected in series across a 12 V supply. The voltage across each resistor should be 6 V. However, the internal resistance of the meter will affect the circuit conditions and change the voltage reading. If the resistor values were 100 kΩ the effect of meter internal resistance would be as follows:

Meter resistance 1 MΩ

The parallel combined value of 1 MΩ and 100 kΩ = 91 kΩ. The voltage drop in the circuit across this would be:

$$91/(100 + 91) \times 12 = 5.71 \text{ V}$$

This is an error of about 5%.

Meter resistance 10 MΩ

The parallel combined value of 10 MΩ and 100 KΩ = 99 KΩ. The voltage drop in the circuit across this would be:

$$99/(100 + 99) \times 12 = 5.97 \text{ V}$$

This is an error of about 0.5%.

Of course understanding accuracy is important, but there are two further skills that are important when using a multimeter: where to put the probes, and what the reading you get actually means!

Figure 3.4 shows a block diagram of a digital voltmeter. This is very similar to other types of digital instrumentation systems.

Key fact

An 'invasive measurement' error is in addition to the basic accuracy of the meter.

Key fact

A voltmeter connects in parallel across a circuit.

An ammeter connects in series.

An ohmmeter connects across a component – but the circuit must be isolated.

3.1.5 Logic probe

This device is a useful way of testing logic circuits but it is also useful for testing some types of sensor. Figure 3.5 shows a typical logic probe. Most types consist of two power supply wires and a metal 'probe'. The display consists of three LEDs labelled 'high', 'low' and 'pulse'. These LEDs light up together with an audible signal in some cases, when the probe touches either a high, low or pulsing voltage. Above or below 2.5 V is often used to determine high or low on a 5 V circuit.

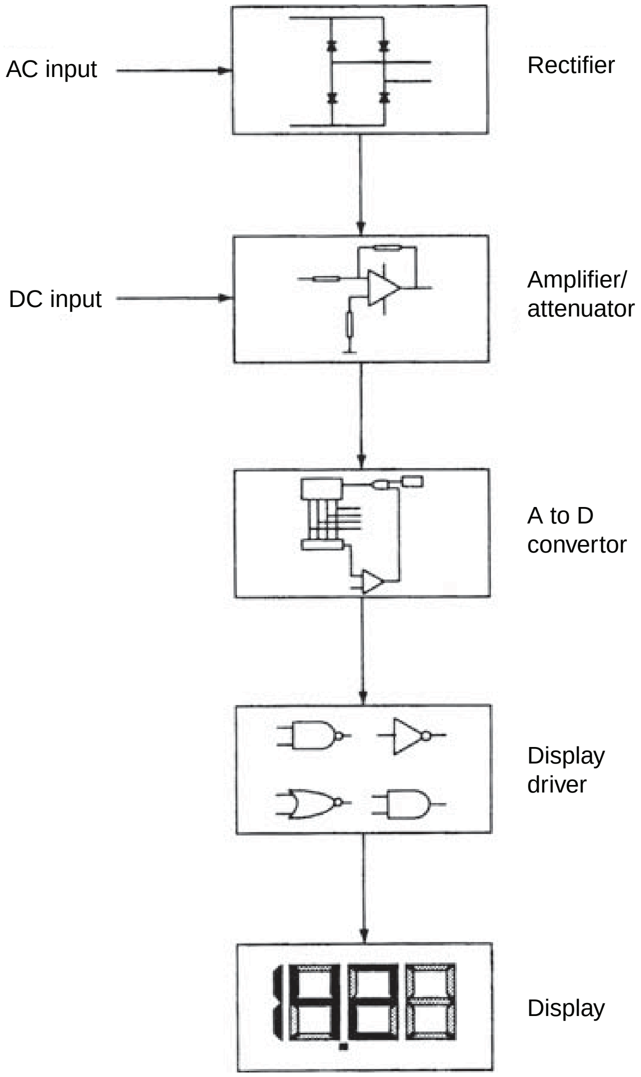


Figure 3.4 Digital voltmeter block diagram

Figure 3.5 Logic probe (Source: Maplin)

3.2 Oscilloscopes

3.2.1 Introduction

There were traditionally two types of oscilloscope; analogue or digital. However, the digital scope is now universal. An oscilloscope draws a graph of voltage (the vertical scale or Y axis) against time (the horizontal scale or X axis).

The trace is made to move across the screen from left to right and then to 'fly back' and start again. The frequency at which the trace moves across the screen is known as the time base, which can be adjusted either automatically or manually.

The signal from the item under test can either be amplified or attenuated (reduced), much like changing the scale on a voltmeter.

The trigger, which is what starts the trace moving across the screen starts, can be caused internally or externally. When looking at signals such as ignition voltages, triggering is often external, each time an individual spark fires or each time number one spark plug fires for example.

The voltage signal under test is A/D converted and the time base is a simple timer or counter circuit. Because the signal is plotted digitally on a screen from data in memory, the 'picture can be saved, frozen or printed. The speed of data conversion and the sampling rate as well as the resolution of the screen are very important to ensure accurate results.

The highly recommended Pico automotive diagnostics kit (Figure 3.6) turns a laptop or desktop PC into a powerful automotive diagnostic tool for fault finding sensors, actuators and electronic circuits.

The high resolution PC oscilloscope connects to a USB port on a PC and can take up to 32 million samples per trace, making it possible to capture complex

Key fact

An oscilloscope draws a graph of voltage against time.

Definition

USB: Universal serial bus.

Figure 3.6 Automotive oscilloscope kit (Source: PicoTech)

automotive waveforms – including CAN bus and FlexRay signals (more on this later) – and then zoom in on areas of interest. Being PC-based these waveforms can then be saved for future reference, printed or emailed.

The scope can be used to measure and test virtually all of the electrical and electronic components and circuits in any modern vehicle including:

- Ignition (primary and secondary)
- Injectors and fuel pumps
- Starter and charging circuits
- Batteries, alternators and starter motors
- Lambda, Airflow, knock and MAP sensors
- Glow plugs / timer relays
- CAN bus, LIN bus and FlexRay.

This powerful and flexible automotive diagnostic tool has been designed for ease of use so is equally suitable for both novice and expert users. It is powered directly from the USB port, eliminating the need for power leads or battery packs, and making it suitable for use in the workshop or on the road.

Excellent software is included which means that the user can simply select the sensor or circuit to be tested and the software will automatically load the required settings. It will also give full details of how to connect the scope, along with advice on what the waveform should look like and general technical information on the component being tested.

All the waveforms shown in this book were captured using this piece of equipment. Visit: www.picoauto.com for more information.

3.2.2 Waveforms

You will find the words 'waveform', 'pattern' and 'trace' are used in books and workshop manuals but they mean the same thing. I will try to stick to waveform.

When you look at a waveform on a screen it is important to remember that the height of the scale represents voltage and the width represents time. Both of these axes can have their scales changed. They are called axis because the 'scope' is drawing a graph of the voltage at the test points over a period of time. The time scale can vary from a few µs to several seconds. The voltage scale can vary from a few mV to several kV. For most test measurements only two connections are needed just like a voltmeter. The time scale will operate at intervals pre-set by the user. It is also possible to connect a 'trigger' wire so that for example the time scale starts moving across the screen each time the ignition coil fires. This keeps the display in time with the speed of the engine. Figure 3.7 shows an example waveform.

Most of the waveforms shown in various parts of this book are from a correctly operating vehicle but some incorrect ones are also presented for comparison. The skill you will learn by practice is to note when your own measurements vary from the ideal – and how to interpret them.

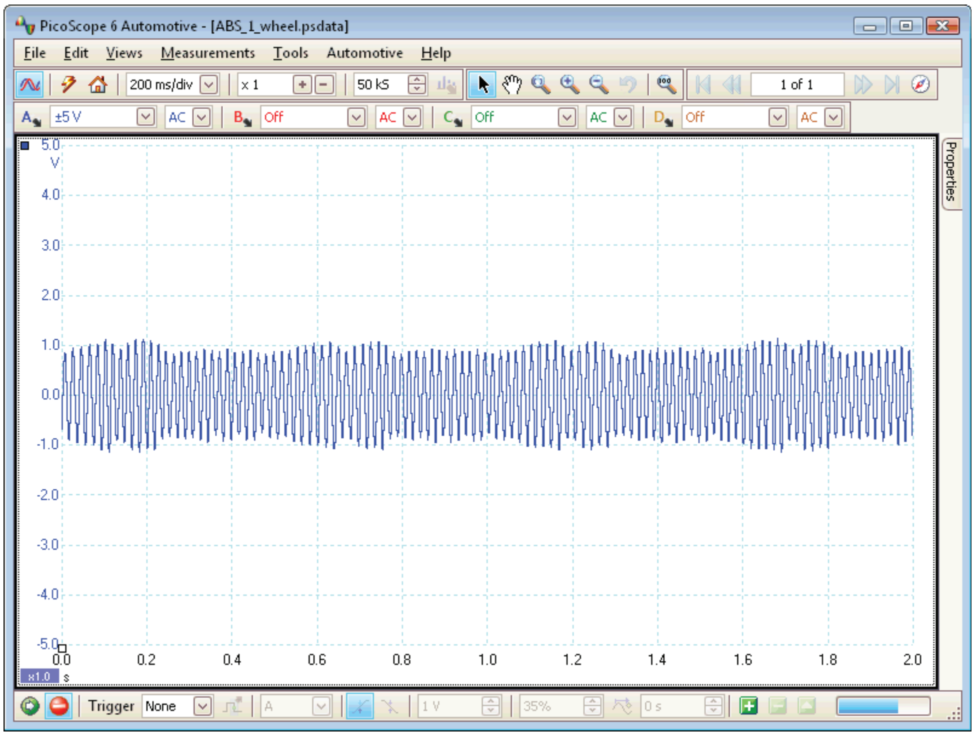


Figure 3.7 ABS waveform captured on a PicoScope

3.3 Scanners/Fault code readers and analysers

Definition

OBD: On-board diagnostics.

Note: Please refer to Chapter 5 for more detail about OBD systems

3.3.1 On-board diagnostics introduction

On-board diagnostics (OBD) is a generic term referring to a vehicle's self-diagnostic and reporting system. OBD systems give the vehicle owner or a technician access to information for various vehicle systems.

The amount of diagnostic information available via OBD has varied considerably since its introduction in the early 1980s. Early versions of OBD would simply illuminate a malfunction indicator light (MIL) if a problem was detected, but did not provide any information about the problem. Modern OBD systems use a standardized digital communications port to provide real-time data in addition to a standardized series of diagnostic trouble codes (DTCs), which allow a technician to identify and remedy faults on the vehicle. The current versions are OBD2 and in European EOBD2. The standard OBD2 and EOBD2 are quite similar.

3.3.2 Serial port communications

Most modern vehicle systems now have ECUs that contain self-diagnosis circuits. The information produced is read via a serial link using a scanner.

A special interface, stipulated by one of a number of standards (see next section), is required to read the data. The standards are designed to work with a single or two wire port allowing many vehicle electronic systems to be connected to a central diagnostic plug. The sequence of events to extract DTCs from the ECU is as follows:

1. Test unit transmits a code word.
2. ECU responds by transmitting a baud rate recognition word.
3. Test unit adopts the appropriate setting.
4. ECU transmits fault codes.

The test unit converts the DTCs to suitable output text. Further functions are possible which may include:

- Identification of ECU and system to ensure the test data is appropriate to the system currently under investigation.
- Read out of current live values from sensors. Spurious figures can be easily recognised. Information such as engine speed, temperature, airflow and so on can be displayed and checked against test data.
- System function stimulation to allow actuators to be tested by moving them and watching for suitable response.
- Programming of system changes such as basic idle CO or changes in basic timing can be programmed into the system.

3.3.3 OBD2 signal protocols

Five different signalling protocols that are permitted with the OBD2 interface. Most vehicles implement only one of them. It is often possible to deduce the protocol used based on which pins are present on the J1962 connector (Figure 3.8).

Some details of the different protocols are presented here for interest. No need to memorise them!

SAE J1850 PWM (pulse-width modulation): A standard of Ford Motor Company

- pin 2: Bus+
- pin 10: Bus–

Definition

Protocol: A set of rules which is used to allow computers to communicate with each other.

Figure 3.8 Diagnostic data link connector (DLC)

- High voltage is +5 V
- Message length is restricted to 12 bytes, including CRC
- Employs a multi-master arbitration scheme called 'Carrier Sense Multiple Access with Non-Destructive Arbitration' (CSMA/NDA).

SAE J1850 VPW (variable pulse width): A standard of General Motors

- pin 2: Bus+
- Bus idles low
- High voltage is +7 V
- Decision point is +3.5 V
- Message length is restricted to 12 bytes, including CRC
- Employs CSMA/NDA.

ISO 9141-2: Primarily used by Chrysler, European, and Asian vehicles

- pin 7: K-line
- pin 15: L-line (optional)
- UART signalling
- K-line idles high, with a 510 ohm resistor to Vbatt
- The active/dominant state is driven low with an open-collector driver
- Message length is restricted to 12 bytes, including CRC.

ISO 14230 KWP2000 (Keyword Protocol 2000)

- pin 7: K-line
- pin 15: L-line (optional)
- Physical layer identical to ISO 9141-2
- Message may contain up to 255 bytes in the data field.

ISO 15765 CAN: The CAN protocol was developed by Bosch for automotive and industrial control. Since 2008 all vehicles sold in the US (and most others) are required to implement CAN as one of their signalling protocols.

- pin 6: CAN High
- pin 14: CAN Low.

All OBD2 pin-outs use the same connector but different pins are utilized with the exception of pin 4 (battery ground) and pin 16 (battery positive).

Definition

ISO: International standards organization.

Definition

CAN: Controller area network.

3.3.4 AutoTap OBD scanner

Author's Note: This section outlines the use and features of the AutoTap scanner. I have chosen this particular tool as a case study because it provides some very advanced features at a very competitive price. The scanner is designed specifically to work with OBD2 systems. However, it worked fine on all the petrol engined EOBD systems I have used it on so far. For more information: www.autotap.com.

Like any professional scanner or code reader, the AutoTap scan tool connects the special OBD2 data link connector, which is always accessible from the driver's seat (often on or under the dash). A USB cable makes the scanner connection to a computer. The scanner translates the signals from the vehicle's computer controlled sensors to easy to read visual displays. It also reads out any diagnostic trouble codes (DTCs).

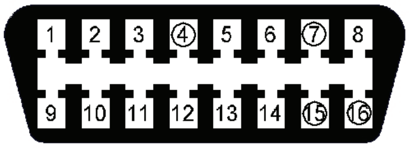
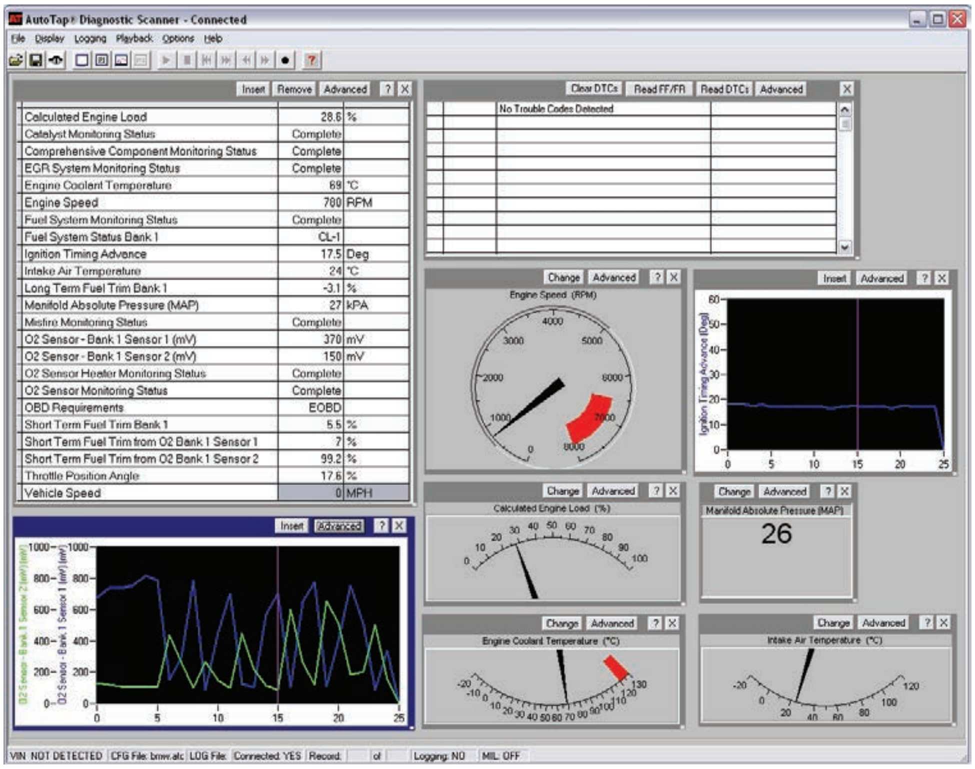


Figure 3.9 Connector pin-out: 4-battery ground/earth, 7-K line, 15-L line, 16-battery positive

Figure 3.10 AutoTap scanner and extension cable

The software also allows the technician (or hobbyist) to choose which parameters or signals they want to see, and whether they are to be viewed as tables, graphs, meters or gauges.

It is possible to set the ranges and alarms and pick display colours. Once a screen configuration is created it can be saved for future use. Different screen configurations are useful for different vehicles, or perhaps one for major maintenance, one for tuning, one for quick checks at a race track.



Calculated Engine Load	28.6	%
Catalyst Monitoring Status	Complete	
Comprehensive Component Monitoring Status	Complete	
EGR System Monitoring Status	Complete	
Engine Coolant Temperature	69	°C
Engine Speed	780	RPM
Fuel System Monitoring Status	Complete	
Fuel System Status Bank 1	CL-1	
Ignition Timing Advance	17.5	Deg
Intake Air Temperature	24	°C
Long Term Fuel Trim Bank 1	-3.1	%
Manifold Absolute Pressure (MAP)	27	kPA
Misfire Monitoring Status	Complete	
O2 Sensor - Bank 1 Sensor 1 (mV)	370	mV
O2 Sensor - Bank 1 Sensor 2 (mV)	150	mV
O2 Sensor Heater Monitoring Status	Complete	
O2 Sensor Monitoring Status	Complete	
OBD Requirements	EOBD	
Short Term Fuel Trim Bank 1	5.5	%
Short Term Fuel Trim from O2 Bank 1 Sensor 1	7	%
Short Term Fuel Trim from O2 Bank 1 Sensor 2	99.2	%
Throttle Position Angle	17.6	%
Vehicle Speed	0	MPH

No Trouble Codes Detected

VIN: NOT DETECTED | CFG File: bmw.alc | LOG File: | Connected: YES | Record: | of | Logging: NO | MIL: OFF

Figure 3.11 Screen grab from the AutoTap software showing tables, gauges and graphs

Lots of data is provided in easy-to-read views with multiple parameters. Graphs can be used to show short-term logs, and gauges for instant readings.

DTCs can be checked immediately on connecting the scanner and starting up the software. This gives the critical info needed in the shortest time possible. When repairs are completed the tool can be used to turn off the malfunction indicator light (MIL). This light is also described as the check engine light.

The software will also log data, for example, during a road test. This is particularly useful for diagnosing intermittent faults. The data can be played back after a road or dynamometer test. It can also be exported to a spreadsheet file for later analysis.

Overall, to read live data and get access to powertrain (engine related) system DTCs, this is an excellent piece of equipment (see Figure 3.9).

Definition

MIL: Malfunction indicator light.

3.3.5 Bosch KTS diagnostic equipment

Author's Note: This section will outline the use and features of the Bosch KTS 650 diagnostic system. I have chosen this particular tool as a case study because it provides everything that a technician needs to diagnose faults, but at a professional price. The system is a combination of a scanner, multimeter, oscilloscope and information system (when used with Esitronic). For more information: www.bosch.com.

Modern vehicles are being fitted with more and more electronics. That complicates diagnosis and repair, especially as the individual systems are often interlinked. The work of service and repair workshops is being fundamentally changed. Automotive engineers have to continually update their knowledge of vehicle electronics. But this is no longer sufficient on its own. The ever-growing number of electrical and electronic vehicle components is no longer manageable without modern diagnostic technology – such as the latest range of KTS control unit diagnostic testers from Bosch. In addition, more and more of the previously purely mechanical interventions on vehicles now require the use of electronic control units – such as the oil change, for example.

Figure 3.12 Diagnostic system in use (Source: Bosch Media)

Figure 3.13 Adapter and cable kit (Source: Bosch Media)

Vehicle workshops operate in a very competitive environment and have to be able to carry out demanding repair work efficiently, to a high standard and at a competitive price on a wide range of vehicle makes and models. The Bosch KTS control-unit diagnostic testers, used in conjunction with the comprehensive Esitronic workshop software, offers the best possible basis for efficient diagnosis and repair of electrical and electronic components. The testers are available in different versions, suited to the individual requirements of the particular workshop.

The portable KTS 650 with built-in computer and touch-screen can be used anywhere. It has a 20GB hard drive, a touch-screen and a DVD drive. When being used away from the workshop, the power supply of the KTS 650 comes from the vehicle battery or from rechargeable batteries with one to two hours' service life. For use in the workshop, there is a tough wheeled trolley with a built-in charger unit. As well as having all the necessary adapter cables, the trolley can also carry an inkjet printer and an external keyboard, which can be connected to the KTS 650 via the usual PC interfaces.

The Esitronic software package accounts for the in-depth diagnostic capacity of the KTS diagnostic testers. With the new common rail diesel systems, for example, even special functions such as quantitative comparison and compression testing can be carried out. This allows for reliable diagnosis of the faulty part and avoids unnecessary dismantling and re-assembly or the removal and replacement of non-faulty parts.

Modern diagnostic equipment is also indispensable when workshops have to deal with braking systems with electronic control systems such as ABS, ASR and ESP. Nowadays, the diagnostic tester may even be needed for bleeding a brake system.

In addition, KTS and Esitronic allow independent workshops to reset the service interval warning, for example after an oil change or a routine service, or perhaps find the correct default position for the headlamps after one or both of these have been replaced.

As well as ISO norms for European vehicles and SAE norms for American and Japanese vehicles, the KTS testers can also deal with CAN norms for checking modern CAN bus systems, which are coming into use more and more frequently in new vehicles. The testers are connected directly to the diagnostics socket via a serial diagnostics interface by means of an adapter cable.

The system automatically detects the control unit and reads out the actual values, the error memory and other controller-specific data. Thanks to a built-in multiplexer, it is even easier for the user to diagnose the various systems in the vehicle. The multiplexer determines the connection in the diagnostics socket so that communication is established correctly with the selected control unit.

3.3.6 Engine analysers

Some form of engine analyser has become an almost essential tool for fault finding modern vehicle engine systems. The latest machines are now generally based around a personal computer. This allows more facilities that can be added to by simply changing the software. However, the latest more portable systems such as the Pico Automotive kit will now do as many tests as the engine analyser, currently with the exception of exhaust emissions.

Whilst engine analysers are designed to work specifically with the motor vehicle, it is worth remembering that the machine consists basically of three parts.

- Multimeter

Figure 3.14 Engine analysers

- Gas analyser
- Oscilloscope.

This is not intended to imply that other tests available such as cylinder balance are less valid, but to show that the analyser is not magic; it is just able to present results of electrical tests in a convenient way to allow diagnosis of faults. The key component of any engine analyser is the oscilloscope facility, which allows the user to 'see' the signal under test.

The trend with engine analysers seems to be to allow both guided test procedures with pass/fail recommendations for the less skilled technician, and freedom to test any electrical device using the facilities available in any reasonable way. This is more appropriate for the highly skilled technician. Some of the routines available on modern engine analysers are listed below.

Tune-up: A full prompted sequence that assesses each component in turn with results and diagnosis displayed at the end of each component test. Stored data allows pass/fail diagnosis by automatically comparing results of tests with data on the disk. Printouts can be taken to show work completed.

Symptom analysis: This allows direct access to specific tests relating to reported driveability problems.

Waveforms: A comprehensive range of digitised waveforms can be displayed with colour highlights. The display can be frozen or recalled to look for intermittent faults. A standard lab scope mode is available to allow examination of EFI or ABS traces for example. Printouts can be made from any display. An interesting feature on some systems is 'transient capture' which ensures even the fastest spikes and intermittent signals are captured and displayed for detailed examination.

Adjustments: Selecting specific components from a menu can make simple quick adjustments. Live readings are displayed appropriate to the selection.

UK MOT Emissions test: Full MOT procedure tests are integrated and displayed on the screen with pass fail diagnosis to the department of transport specifications for both gas analysis and diesel smoke if appropriate options are fitted. The test results include engine rpm and oil temperature as well as the gas readings. These can all be printed for garage or customer use.

Engine analyser connections to the vehicle are similar for most equipment manufacturers.

Table 3.4 Typical waveforms that can be displayed on most analysers and automotive oscilloscopes

Primary	Secondary	Diagnostic	Cylinder test
Primary waveform	Secondary waveform	Voltage waveform	Vacuum waveform
Primary parade waveform	Secondary parade waveform	Lab scope waveform	Power balance waveform
Dwell bar graph	kV Histogram	Fuel injector waveform	Cylinder time balance bar graph
Duty cycle/Dwell bar graph	kV bar graph	Alternator waveform	Cylinder shorting even/odd bar graph
Duty cycle/Voltage bar graph	Burn time bar graph		Cranking amps bar graph

Table 3.5 Analyser connections

Connection	Purpose or one example of use
Battery positive	Battery and charging voltages
Battery negative	A common earth connection
Coil positive	To check supply voltage to coil
Coil negative (adapters are available for DIS)	To look at dwell, rpm and primary waveforms
Coil HT lead clamp (adapters are available for DIS)	Secondary waveforms
Number one cylinder plug lead clamp	Timing light and sequence of waveforms
Battery cable amp clamp	Charging and starting current
Oil temperature probe (dip stick hole)	Oil temperature
Vacuum connection	Engine load
Exhaust pipe	Emissions testing

3.4 Emission testing

3.4.1 Introduction

Checking the exhaust emissions of a vehicle has three main purposes:

1. Ensure optimum performance.
2. Compliance with regulations and limits.
3. Diagnostic information.

There are many different exhaust testing systems available.

3.4.2 Exhaust gas measurement

It has now become standard to measure four of the main exhaust gases namely:

- Carbon monoxide (CO).
- Carbon dioxide (CO_2).
- Hydrocarbons (HC).
- Oxygen (O_2).

On many analysers, lambda value and the air fuel ratio are calculated and displayed in addition to the four gasses. The Greek symbol lambda (λ) is used to represent the ideal air fuel ratio (AFR) of 14.7:1 by mass. In other words just the right amount of air to burn up all the fuel. Table 3.7 lists gas, lambda and AFR readings for a closed loop lambda control system, before (or without) and after the catalytic converter. These are for a modern engine in excellent condition and are a guide only – always check current data for the vehicle you are working on.

The composition of exhaust gas is now a critical measurement and hence a certain degree of accuracy is required. To this end the infrared measurement

Key fact

The Greek symbol lambda (λ) represents the ideal air fuel ratio (AFR) of 14.7:1 by mass.

Table 3.6 Exhaust examples

Reading:	CO%	HC ppm	CO_2%	O_2%	Lambda (λ)	AFR
Before catalyst	0.6	120	14.7	0.7	1.0	14.7
After catalyst	0.2	12	15.3	0.1	1.0	14.7

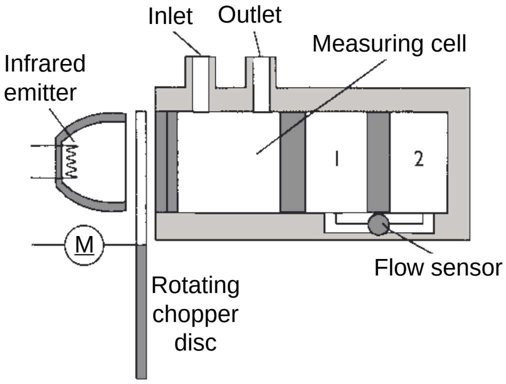


Figure 3.15 Carbon monoxide measurement technique

technique has become the most suitable for CO, CO_2 and HC. Each individual gas absorbs infrared radiation at a specific rate. Oxygen is measured by electro-chemical means in much the same way as the on vehicle lambda sensor.

CO can be measured as shown in Figure 3.15. The emitter is heated to about 700 °C which, by using a suitable reflector, produces a beam of infrared light. This beam is passed via a chopper disc, through a measuring cell to a receiver chamber. This sealed chamber contains a gas with a defined content of CO (in this case). This gas absorbs some of the CO specific radiation and its temperature increases. This causes expansion and therefore a gas flow from chamber 1 to chamber 2. This flow is detected by a flow sensor, which produces an AC output signal. This is converted and calibrated to a zero CO reading. The AC signal is produced due to the action of the chopper disk. If the chopper disc was not used then the flow from chamber 1 to chamber 2 would only take place when the machine was switched on or off.

If the gas to be measured is now pumped through the measuring cell, some of the infrared radiation will be absorbed before it reaches the receiver chamber. This varies the heating effect on the CO specific gas and hence the measured flow between chambers 1 and 2 will change. The flow meter will produce a change in its AC signal, which is converted, and then output to a suitable display. A similar technique is used for the measurement of CO_2 and HC. At present it is not possible to measure nitrogen oxides (NOx) without the most sophisticated laboratory equipment. Research is being carried out in this area.

Accurate measurement of exhaust gas is not only required for annual tests but is essential to ensure an engine is correctly tuned. Table 3.6 lists typical values measured from a car exhaust. Note the toxic HC and CO emissions whilst small, are none-the-less dangerous.

3.4.3 Exhaust analyser

The facilities of an exhaust analyser produced by Bosch are outlined here.

Figure 3.16 Exhaust gas measuring components (Source: Bosch Media)

The measuring system shown in Figures 3.16 and 3.17 can be used for petrol/gasoline, diesel and natural gas vehicles (a statutory requirement in Germany). It is designed for quick and mobile use in workshops and is a robust design. It measures the usual four-gases, weighs less than 15 kg and can be ready for operation in just a few minutes. The system is controlled by software, which takes users through the test sequence. The device can be serviced by users themselves every six months.

The system measures the HC, CO, CO_2 and O_2 exhaust components for petrol/gasoline engines. It can also be expanded to measure NO if necessary. It records

Figure 3.17 Exhaust gas measuring system in use (Source: Bosch Media)

Definition

Bluetooth: A proprietary open wireless protocol for exchanging data over short distances from fixed and mobile devices, creating personal area networks (PANs).

engine speed and temperature. Adding a smoke opacity measuring device means exhaust gas analyses can be carried out on diesel vehicles. Linking with the KTS (see section 3.3.5) allows important OBD engine and transmission control unit data to be read as well as the gases. The laptop and KTS can be connected via a cable or Bluetooth.

3.4.4 Emission limits

Limits and regulations relating to exhaust emissions vary in different countries and in different situations. For example, in the UK certain limits have to be met during the annual test. The current test default limits (for vehicles since September 2002 fitted with a catalytic converter) are:

At a minimum oil temperature 60 °C:

Fast idle (2500 to 3000 rpm)

- CO <= 0.2%
- HC <= 200 ppm
- Lambda 0.97 to 1.03.

Idle (450 to 1500 rpm)

- CO <= 0.3%.

Manufacturers however, have to meet stringent regulations when producing new vehicles. In Europe the emission standards are defined in a series of EU directives staging the progressive introduction of increasingly stringent standards (see Table 3.7).

In the USA, what are known as Tier II standards are divided into several numbered ‘bins’. Eleven bins were initially defined, with bin 1 being the cleanest (Zero Emission Vehicle) and 11 the dirtiest. However, bins 9, 10 and 11 are temporary. Only the first ten bins were used for light-duty vehicles below 8500 pounds GVWR, but medium-duty passenger vehicles up to 10 000 pounds (4536 kg) GVWR and to all 11 bins. Manufacturers can make vehicles which fit into any of the available bins, but still must meet average targets for their entire fleets.

The two least-restrictive bins for passenger cars, 9 and 10, were phased out at the end of 2006. However, bins 9 and 10 were available for classifying a restricted number of light-duty trucks until the end of 2008, when they were removed along with bin 11 for medium-duty vehicles. As of 2009, light-duty trucks must meet the same emissions standards as passenger cars.

Table 3.7 European past and future emission limits

Emissions Standard	Particulate matters (PM)/ (mg/km)		Oxides of nitrogen (NOx) (mg/km)		Hydrocarbons (HC) (mg/km)	
	Diesel	Petrol	Diesel	Petrol	Diesel	Petrol
Euro 2 (1996)	80-100	-	-	-	-	-
Euro 3 (2000)	50	-	500	150	-	200
Euro 4 (2005)	25	-	250	80	-	100
Euro 5 (2009)	5	5	180	70	-	100
Euro 6 (2014)	5	5	80	70	-	100

Table 3.8 Tier 2 exhaust emission standards (USA)

	Standard	Emission Limits at 50 000 miles					Emission Limits at Full Useful Life (120 000 miles)[a]				
		NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)	NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)
Federal	Bin 1	-	-	-	-	-	0	0	0	0	0
	Bin 2	-	-	-	-	-	0.02	0.01	2.1	0.01	0.004
	Bin 3	-	-	-	-	-	0.03	0.055	2.1	0.01	0.011
	Bin 4	-	-	-	-	-	0.04	0.07	2.1	0.01	0.011
	Bin 5	0.05	0.075	3.4	-	0.015	0.07	0.09	4.2	0.01	0.018
	Bin 6	0.08	0.075	3.4	-	0.015	0.1	0.09	4.2	0.01	0.018
	Bin 7	0.11	0.075	3.4	-	0.015	0.15	0.09	4.2	0.02	0.018
	Bin 8	0.14	0.100/0.125[c]	3.4	-	0.015	0.2	0.125/0.156	4.2	0.02	0.018
	Bin 9[b]	0.2	0.075/0.140	3.4	-	0.015	0.3	0.090/0.180	4.2	0.06	0.018
	Bin 10[b]	0.4	0.125/0.160	3.4/4.4	-	0.015/0.018	0.6	0.156/0.230[c]	4.2/6.4	0.08	0.018/0.027
	Bin 11[b]	0.6	0.195	5	-	0.022	0.9	0.28	7.3	0.12	0.032

[a] In lieu of intermediate useful life standards (50 000 miles) or to gain additional nitrogen oxides credit, manufacturers may optionally certify to the Tier 2 exhaust emission standards with a useful life of 150 000 miles.

[b] Bins 9-11 expire in 2006 for light-duty vehicles and light light-duty trucks and 2008 for heavy light-duty trucks and medium-duty passenger vehicles.

[c] Pollutants with two numbers have a separate certification standard (1st number) and in-use standard (2nd number).

Phase 2 was 2004 to 2009 and now even more stringent standards are coming into use. Also, the California Air Resources Board (CARB) may also adopt and enforce its own emissions standards. However, regardless of whether a manufacturer receives CARB approval, all new motor vehicles and engines must still receive certification from the environmental protection agency (EPA) before a vehicle is introduced.

3.5 Pressure testing

3.5.1 Introduction

Measuring the fuel pressure on fuel injection engine is of great value when fault finding. Many types of pressure tester are available and they often come as part of a kit consisting of various adapters and connections (Figure 3.18). The principle of mechanical gauges is that they contain a very small tube wound in a spiral. As fluid or gas under pressure is forced into the spiral tube, it unwinds causing a needle to move over a graduated scale.

Measuring engine cylinder compression or leakage is a useful test. Figure 3.19 shows an engine compression tester. This device is used to compare cylinder compressions as well as to measure actual values.

Figure 3.18 Fuel pressure gauge kit (Source: Sealey)

Figure 3.19 Compression tester (Source: Sealey)

3.5.2 Automotive pressure oscilloscope transducer

Definition

Transducer: A device that converts a physical quantity (e.g. force, torque, pressure, rotation) to an electrical signal.

PicoTech have developed an accurate pressure transducer that can be used for pressure analysis of many automotive systems.

Some of the key features are:

- range accurate from 0.07 psi (5 mbar) to 500 psi (34.5 bar);
- 100 μs response time;
- zoom function for enhanced analysis;
- temperature compensation.

These result in an accurate representation of rapidly changing signals that span across a broad pressure range.

The three pressure ranges of the device allow for accurate measurement and analysis of many automotive pressures including:

- cylinder compression;
- fuel pressure;
- intake manifold vacuum;
- pulses from the exhaust.

Figure 3.20 Automotive pressure transducer (Source: PicoTech)

The first range gives high resolution and accuracy for high-pressure tests such as cranking and running cylinder compression or fuel pressure testing (Figure 3.21).

The second range measures from −15 to 50 psi (approximately −1 to 3.45 bar). This range is ideal for vacuum tests and fuel system tests. The zoom function is especially useful on these tests as it makes it easy to analyse the valves operating with the vacuum waveform, or the injectors through the fuel waveform.

With the third range you can measure −5 to 5 psi (approximately −0.34 to 0.34 bar). This setting is sensitive enough to allow analysis of small pressures or pulses such as from the exhaust. This is an excellent way of checking for even running cylinders.

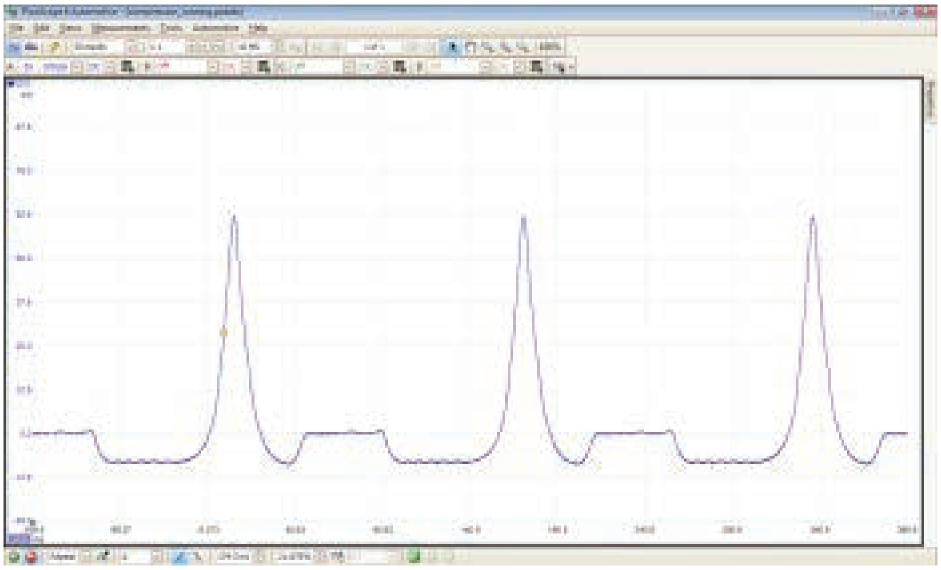


Figure 3.21 Running compression waveform

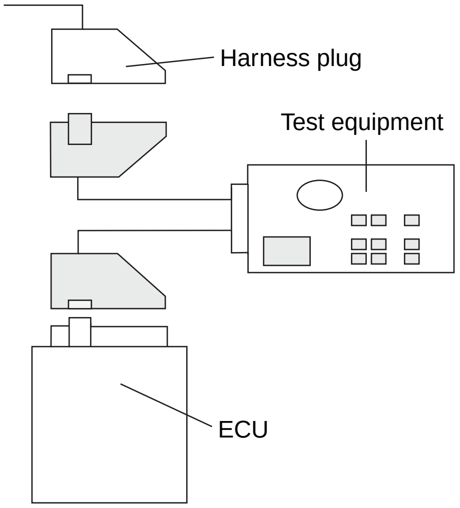


Figure 3.22 Breakout box test equipment

3.5.3 Breakout boxes

Manufacturers have used a system similar to Figure 3.22 for many years, known simply as a breakout box. A multimeter takes the readings between predetermined test points on the box which are connected to the ECU wiring. A variation on this system is a digitally controlled tester that will run very quickly through a series of tests and display the results. These can be compared with stored data allowing a pass/fail output.

3.6 Diagnostic procedures

3.6.1 Introduction

Finding the problem when complex automotive systems go wrong – is easy. Well, it is easy if you have the necessary knowledge. This knowledge is in two parts:

1. an understanding of the system in which the problem exists,
2. the ability to apply a logical diagnostic routine.

It is also important to be clear about two definitions:

- Symptom(s) – what the user of the system (vehicle or whatever) notices.
- Fault – the error in the system that causes the symptom(s).

The knowledge requirement and use of diagnostic skills can be illustrated with the simple example in the next section.

3.6.2 The 'theory' of diagnostics

One theory of diagnostics can be illustrated by the following example: After connecting a hosepipe to the tap and turning on the tap, no water comes out of the end. Your knowledge of this system tells you that water should come out providing the tap is on, because the pressure from a tap pushes water through the pipe, and so on. This is where diagnostic skills become essential. The following stages are now required.

1. Confirm that no water is coming out by looking down the end of the pipe!
2. Does water come out of the other taps, or did it come out of this tap before you connected the hose?
3. Consider what this information tells you, for example, the hose must be blocked or kinked.
4. Walk the length of the pipe looking for a kink.
5. Straighten out the hose.
6. Check that water now comes out and that no other problems have been created.

The procedure just followed made the hose work but it is also guaranteed to find a fault in any system. It is easy to see how it works in connection with a hosepipe, but I'm sure anybody could have found that fault! The skill is to be able to apply the same logical routine to more complex situations. The routine can be summarized by the six-stages of diagnostics (Figure 3.23).

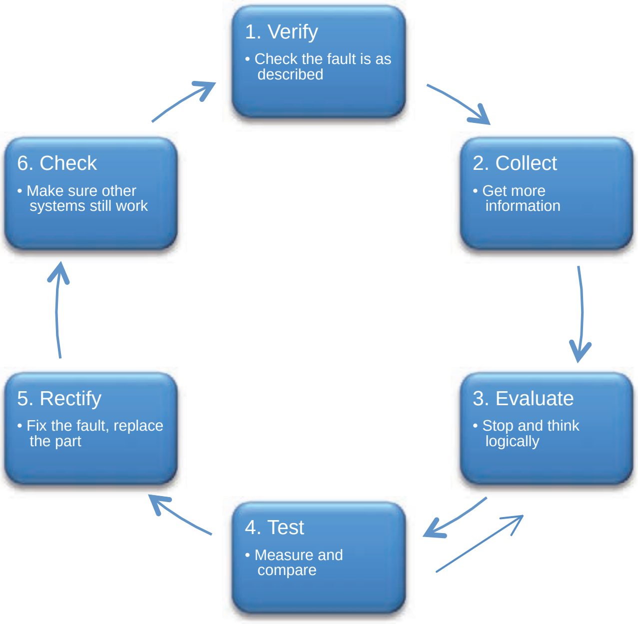


Figure 3.23 Six-stage diagnostic process

Steps 3 and 4 form a loop, within the larger loop, until the fault is located. Remember that using a logical process will not only ensure you find the fault, it will also save time and effort.

CHAPTER 4

Electrical systems and circuits

4.1 The systems approach

4.1.1 What is a system?

System is a word used to describe a collection of related components, which interact as a whole. A motorway system, the education system or computer systems are three varied examples. A large system is often made up of many smaller systems which in turn can each be made up of smaller systems and so on. Figure 4.1 shows how this can be represented in a visual form. One further definition: 'A group of devices serving a common purpose'.

Using the systems approach helps to split extremely complex technical entities into more manageable parts. It is important to note however, that the links between the smaller parts and the boundaries around them are also very important. System boundaries will overlap in many cases.

The modern motor vehicle is a complex system and in itself forms just a small part of a larger transport system. It is the ability for the motor vehicle to be split into systems on many levels, which aids both in its design and construction. The systems approach helps in particular with understanding how something works and further how to go about repairing it when it doesn't.

Definition

System: From the Latin systēma, in turn from Greek σύστημσ systēma, is a set of interacting or interdependent system components forming an integrated whole.

4.1.2 Vehicle systems

Splitting the vehicle into systems is not an easy task because it can be done in many different ways. A split between mechanical systems and electrical

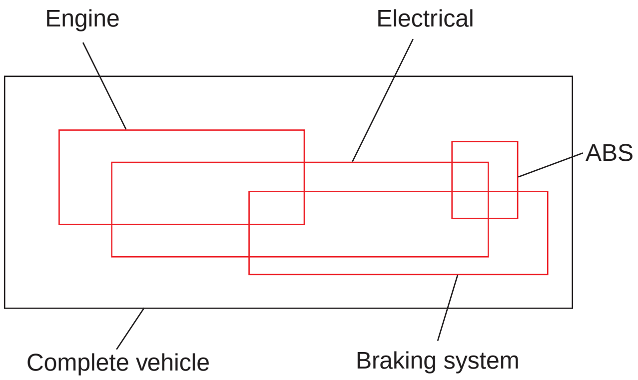


Figure 4.1 Systems in systems representation

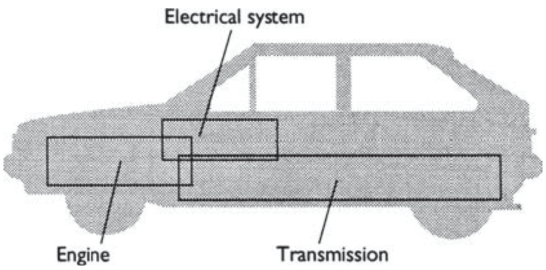


Figure 4.2 Vehicle systems representation

systems would seem a good start. However, this division can cause as many problems as it solves. For example, in which half do we put anti-lock brakes, mechanical or electrical? The answer is of course both. None-the-less, it still makes it easier to be able to just consider one area of the vehicle and not have to try to comprehend the whole.

Once a complex set of interacting parts such as a motor vehicle has been 'systemized', the function or performance of each part can be examined in more detail. In other words, looking at what each part of the system should do in turn then helps to determine how each part actually works. It is again important to stress that the links and interactions between various sub-systems are a very important consideration. Examples of this would be how the power demands of the vehicle lighting system will have an effect on the charging system operation, or in the case of a fault, how an air leak from a brake servo could cause a weak air/fuel ratio.

To further analyse a system whatever way it has been sub divided from the whole, consideration should be given to the inputs and the outputs of the system. Many of the complex electronic systems on a vehicle lend themselves to this form of analysis. Considering the electronic control unit (ECU) of the system as the control element and looking at its inputs and outputs is the recommended approach.

4.1.3 Open loop systems

An open loop system is designed to give the required output whenever a given input is applied. A good example of an open loop vehicle system would be the headlights. With the given input is the switch being operated, the output required is that the headlights will be illuminated.

Figure 4.3 Open loop system

This can be taken further by saying that an input is also required from the battery and a further input from, say, the dip switch. The feature, which determines that a system is open loop, is that no feedback is required for it to operate.

4.1.4 Closed loop systems

A closed loop system is identified by a feedback loop. It can be described as a system where there is a possibility of applying corrective measures if the output is not quite what is wanted. A good example of this in a vehicle is an automatic

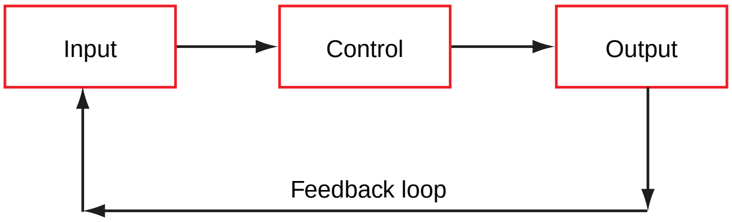


Figure 4.4 Closed loop system

temperature control system (Figure 4.5). The interior temperature of the vehicle is determined by the output from the heater which is switched on or off in response to a signal from a temperature sensor inside the cabin. The feedback loop is the fact that the output from the system, temperature, is also an input to the system.

The feedback loop in any closed loop system can be in many forms. The driver of a car with a conventional heating system can form a feedback loop by turning the heater down when he is too hot and turning it back up when cold. The feedback on an ABS system is a signal that the wheel is locking, where the system reacts by reducing the braking force – until it stops locking, when braking force can be increased again – and so on to maintain a steady state.

Key fact

A closed loop system always has a feedback loop that may be negative or positive.

4.1.5 Summary

Many complex vehicle systems are represented in this book as block diagrams. In this way several inputs can be shown supplying information to an ECU which, in turn, controls the system outputs.

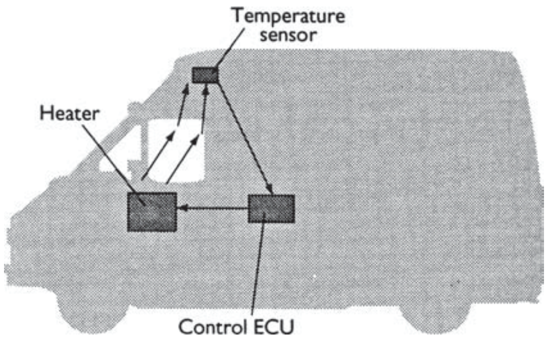


Figure 4.5 Closed loop heating system

4.2 Electrical wiring, terminals and switching

4.2.1 Cables

Cables used for motor vehicle applications are almost always copper strands insulated with PVC. Copper, beside its very low resistivity of about 1.7×10^{-8} Ωm, has ideal properties such as ductility and malleability. This makes it the natural choice for most electrical conductors. PVC as the insulation is again ideal as it not only has very high resistance, in the order of 10^{15} Ωm, but is also very resistant to petrol, oil, water and other contaminants.

The choice of cable size depends on the current drawn by the consumer. The larger the cable used then the smaller the volt drop in the circuit but the cable will

Table 4.1 typical maximum volt drops

Circuit (12 V)	Load	Cable drop (V)	Maximum drop (V) including connections
Lighting circuit	<15 W	0.1	0.6
Lighting circuit	>15 W	0.3	0.6
Charging circuit	Nominal	0.5	0.5
Starter circuit	Maximum at 20 °C	0.5	0.5
Starter solenoid	Pull-in	1.5	1.9
Other circuits	Nominal	0.5	1.5

Key fact

The choice of cable size depends on the current drawn by the consumer.

be heavier. This means a trade-off must be sought between allowable volt drop and maximum cable size. Table 4.1 lists some typical maximum volt drops in a circuit.

In general the supply to a component must not be less than 90% of the system supply. Vehicles using a 24 V supply mean the figures in the previous table should be doubled. Volt drop in a cable can be calculated as follows:

Calculate the current $I = W/Vs$

Volt drop $Vd = I.\rho.l/A$

where: I *= current in amps,* W *= power rating of component in watts,* Vs *= system supply in volts,* Vd *= volt drop in volts,* ρ *= resistivity of copper in* Ω*m,* l *= length of the cable in m,* A *= cross sectional area in* m^2.

A transposition of this formula will allow the required cable cross section to be calculated.

$$A = I\rho l/Vd$$

where: I *= maximum current in amps,* Vd *= maximum allowable volt drop in volts.*

Cable is available in stock sizes and the following table lists some typical sizes and uses. The current rating is assuming that the cable length is not excessive and that operating temperature is within normal limits. Cables normally consist of multiple strands to provide greater flexibility.

4.2.2 Colour codes and terminal designations

As seems to be the case for any standardisation a number of colour code and terminal designation systems are in operation! For reference purposes I will just make mention of three. Firstly, the British Standard system (BS AU 7a: 1983). This system uses twelve colours to determine the main purpose of the cable and tracer colours to further refine its use. The main colour uses and some further examples are given in the following table.

A ‘European’ system used by a number of manufacturers is based broadly on the following table. Please note that there is no correlation between the ‘Euro’ system and the British standard colour codes. In particular note the use of the colour brown in each system! After some practice with the use of colour code systems the job of the technician is made a lot easier when fault finding an electrical circuit.

Table 4.2 Cables and their applications

Cable strand diameter (mm)	Cross sectional area (mm²)	Continuous current rating (A)	Example applications
9/0.30	0.6	5.75	Side lights etc.
14/0.25	0.7	6 A	Clock, radio
14/0.30	1.0	8.75	Ignition
28/0.30	2.0	17.5	Headlights, HRW
44/0.30	3.1	27.5	
65/0.30	4.6	35.0	Main supply
84/0.30	5.9	45.0	
97/0.30	6.9	50.0	Charging wires
120/0.30	8.5	60.0	
37/0.90	23.5	350.0	Starter supply
to		to	
61/0.90	39.0	700.0	

Table 4.3 British standard colour codes

Colour	Symbol	Destination/Use
Brown	N	Main battery feed
Blue	U	Headlight switch to dip switch
Blue/White	UW	Headlight main beam
Blue/Red	UR	Headlight dip beam
Red	R	Side light main feed
Red/Black	RB	Left hand side lights and no. plate
Red/Orange	RO	Right hand side lights
Purple	P	Constant fused supply
Green	G	Ignition controlled fused supply
Green/Red	GR	Left side indicators
Green/White	GW	Right side indicators
Light Green	LG	Instruments
White	W	Ignition to ballast resistor
White/Black	WB	Coil negative
Yellow	Y	Overdrive and fuel injection
Black	B	All earth connections
Slate	S	Electric windows
Orange	O	Wiper circuits (fused)
Pink/White	KW	Ballast resistor wire
Green/Brown	GN	Reverse
Green/Purple	GP	Stop lights
Blue/Yellow	UY	Rear Fog light

Table 4.4 European colour codes

Colour	Symbol	Destination/Use
Red	Rt	Main battery feed
White/Black	Ws/Sw	Headlight switch to dip switch
White	Ws	Headlight main beam
Yellow	Ge	Headlight dip beam
Grey	Gr	Side light main feed
Grey/Black	Gr/Sw	Left hand side lights
Grey/Red	Gr/Rt	Right hand side lights
Black/Yellow	Sw/Ge	Fuel injection
Black/Green	Sw/Gn	Ignition controlled supply
Black/White/Green	Sw/Ws/Gn	Indicator switch
Black/White	Sw/Ws	Left side indicators
Black/Green	Sw/Gn	Right side indicators
Light Green	LGn	Coil negative
Brown	Br	Earth
Brown/White	Br/Ws	Earth connections
Pink/White	KW	Ballast resistor wire
Black	Sw	Reverse
Black/Red	Sw/Rt	Stop lights
Green/Black	Gn/Sw	Rear Fog light

A system now in common use is the terminal designation system in accordance with DIN 72 552. This system is to enable easy and correct connections to be made on the vehicle, particularly in after sales repairs. It is important however to note that the designations are not to identify individual wires but are to define the terminals of a device. Table 4.5 lists some of the most popular numbers.

Ford motor company has used a circuit numbering and wire identification system. This is in use world-wide and is known as Function, System-Connection (FSC). The system was developed to assist in vehicle development and production processes. However, it is also very useful to help the technician with fault-finding. Many of the function codes are based on the DIN system. Note that earth wires are now black! The system works as follows:

31S-AC3A || 1.5 BK/RD

Function:
    31 = ground/earth
    S = additionally switched circuit
System:
    AC = headlamp levelling
Connection:
    3 = switch connection
    A = branch

Table 4.5 Terminal designation numbers (DIN 72 552)

1	Ignition coil negative
4	Ignition coil high tension
15	Switched positive (ignition switch output)
30	Input from battery positive
31	Earth connection
49	Input to flasher unit
49a	Output from flasher unit
50	Starter control (solenoid terminal)
53	Wiper motor input
54	Stop lamps
55	Fog lamps
56	Headlamps
56a	Main beam
56b	Dip beam
58L	Left side lights
58R	Right side lights
61	Charge warning light
85	Relay winding out
86	Relay winding input
87	Relay contact input (change over relay)
87a	Relay contact output (break)
87b	Relay contact output (make)
L	Left side indicators
R	Right side indicators
C	Indicator warning light (vehicle)

Size:
1.5 = 1.5 mm^2
Colour:
BK = Black (determined by function 31)
RD = Red stripe

As a final point to this section it must be noted that the colour codes and terminal designations given, are for illustration only. Further reference should always be made for specific details to manufacturer's information.

4.2.3 Harness design

The vehicle wiring harness has developed over the years from a loom containing just a few wires, to the looms used at present on top range vehicles containing well over 1000 separate wires. Modern vehicles tend to have wiring harnesses

Table 4.6 Ford system colour codes

Code	Colour
BK	Black
BN	Brown
BU	Blue
GN	Green
GY	Grey
LG	Light-Green
OG	Orange
PK	Pink
RD	Red
SR	Silver
VT	Violet
WH	White
YE	Yellow

Table 4.7 System codes

Letter	Main system	Examples
D	Distribution systems	DE = earth
A	Actuated systems	AK = wiper/washer
B	Basic systems	BA = charging BB = starting
C	Control systems	CE = power steering
G	Gauge systems	GA = level/pressure/temperature
H	Heated systems	HC = heated seats
L	Lighting systems	LE = headlights
M	Miscellaneous systems	MA = air bags
P	Powertrain control systems	PA = engine control
W	Indicator systems ('indications' not turn signals)	WC = bulb failure
X	Temporary for future features	XS = too much!

Key fact

The most popular wiring harness is a bundle of cables spirally wrapped in non-adhesive PVC or similar tape.

constructed in a number of ways. The most popular is still for the bundle of cables to be spirally wrapped in non-adhesive PVC tape. The tape is non-adhesive so as to allow the bundle of wires to retain some flexibility, as shown in Figures 4.6 and 4.7.

Another technique often used is to place the cables side by side and plastic weld them to a backing strip as shown in Figure 4.8. This method allows the loom to be run in narrow areas, for example behind the trim on the inner sill or under carpets.

Figure 4.6 PVC wound harness

Figure 4.7 Canvas tape harness

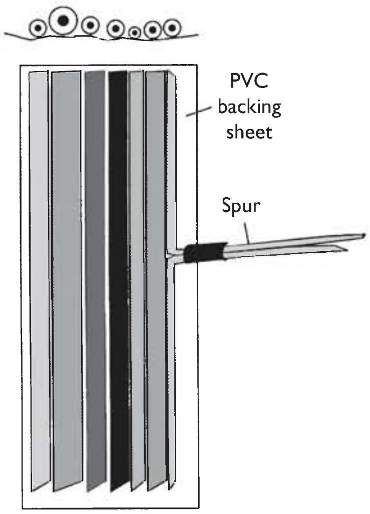


Figure 4.8 Cables side by side and plastic welded to a backing strip

A third way of grouping cables, as shown in Figure 4.9 is to place them inside PVC tubes. This has the advantage of being harder wearing and, if suitable sealing is arranged, can also be waterproof.

When deciding on the layout of a wiring loom within the vehicle, many issues must be considered. Some of these are as follows.

1. Cable runs must be as short as possible.
2. The loom must be protected against physical damage.
3. The number of connections should be kept to a minimum.
4. Modular design may be appropriate.
5. Accident damage areas to be considered.
6. Production line techniques should be considered.
7. Access must be possible to main components and sub-assemblies for repair purposes.

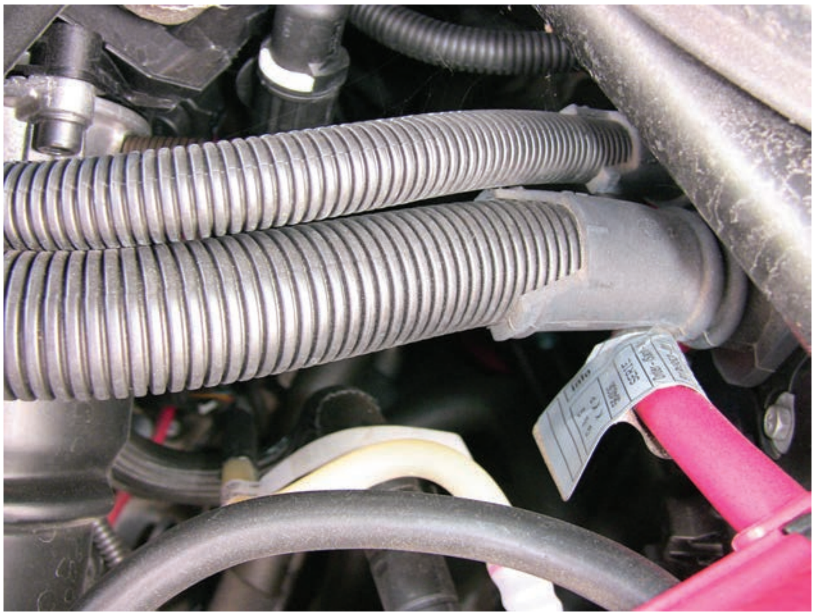


Figure 4.9 PVC tube and tape harness

Key fact

Keeping cable runs as short as possible will reduce volt drop problems and reduce the weight of the harness.

From the above list, which is by no means definitive, it can be seen that, as with most design problems, some of the main issues for consideration are at odds with each other. The more connections involved in a wiring loom, then the more areas there are for potential faults to develop. However, having a large multiplug assembly, which connects the entire engine wiring to the rest of the loom, can have considerable advantages. During production, the engine and all its ancillaries can be fitted as a complete unit if supplied ready wired, and in the after-sales repair market, engine replacement and repairs are easier to carry out.

Because wiring looms are now so large, it is often necessary to split them into more manageable sub-assemblies. This will involve more connection points. The main advantage of this is that individual sections of the loom can be replaced if damaged.

Keeping cable runs as short as possible will not only reduce volt drop problems but will allow thinner wire to be used, thus reducing the weight of the harness, which can now be quite considerable.

The overall layout of a loom on a vehicle will broadly follow one of two patterns; that is, an 'E' shape or an 'H' shape (Figure 4.10). The 'H' is the more common layout. It is becoming the norm to have one or two main junction points as part of the vehicle wiring with these points often being part of the fuse box and relay plate.

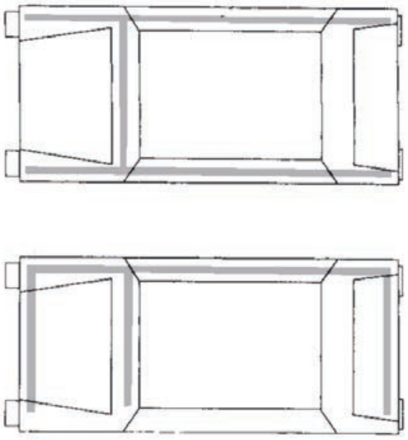


Figure 4.10 'H' and 'E' wiring layouts

Figure 4.11 shows a more realistic representation of the harness layout. This figure also serves to show the level of complexity and number of connection points involved. It is the aim of multiplexed systems (discussed later) to reduce these problems and provide extra 'communication' and diagnostic facilities.

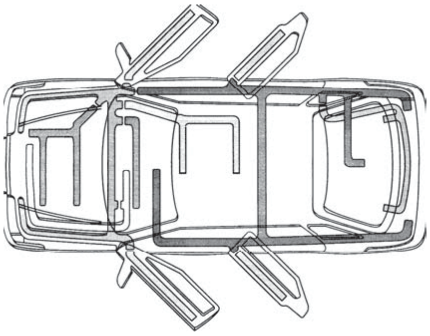


Figure 4.11 Typical wiring harness layout

4.2.4 Printed circuits

The printed circuit is often used in areas such as the rear of the instrument pack and other similar places. This allows these components to be supplied as complete units and also reduces the amount and complexity of the wiring in what are usually cramped areas.

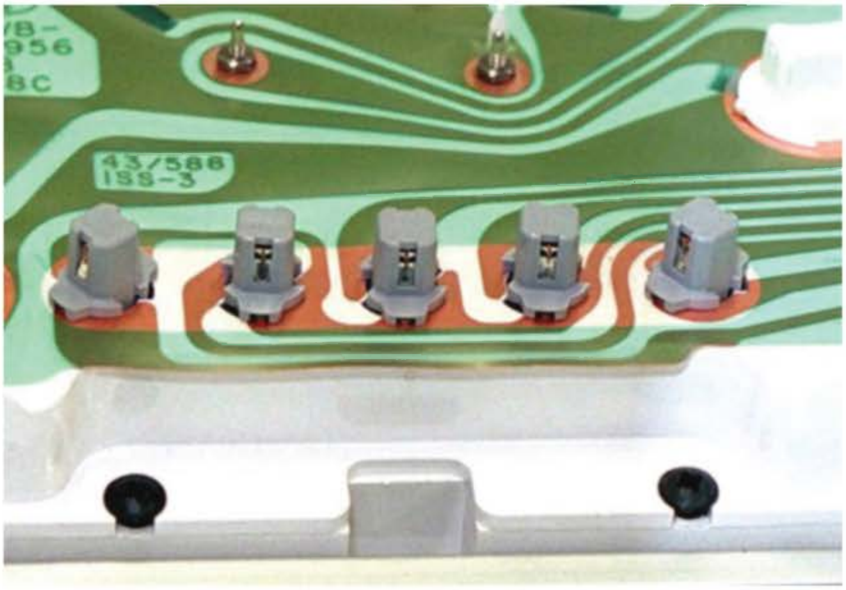


Figure 4.12 Instrument pack printed circuit

The printed circuits are constructed using a thin copper layer that is bonded to a plastic sheet – on both sides in some cases. The required circuit is then printed on to the copper using a material similar to wax. The unwanted copper can then be etched away with an acid wash. A further layer of thin plastic sheet can insulate the copper strips if required.

Figure 4.12 shows a picture of a typical printed circuit from an instrument panel and gives some indication as to how many wires would be required to do the same job. Connection to the main harness is by one or more multiplugs.

4.2.5 Fuses and circuit breakers

Some form of circuit protection is required to protect the electrical wiring of a vehicle against a short circuit and also to protect the electrical and electronic components. It is now common practice to protect almost all electrical circuits with a fuse. The simple definition of a fuse is that it is a deliberate weak link in the circuit. If an overload of current occurs then the fuse will melt and disconnect the circuit before any serious damage is caused. Automobile fuses are available in three types, glass cartridge, ceramic and blade type. The blade type is the most popular choice owing to its simple construction and reliability against premature failure due to vibration. Figure 4.13 shows different types of fuse and Figure 4.14 shows a selection of the common blade type.

Fuses are rated with a continuous and peak current value. The continuous value is the current that the fuse will carry without risk of failure, whereas the peak value is the current that the fuse will carry for a short time without failing. The peak value of a fuse is usually double the continuous value. Using a lighting circuit as an example, when the lights are first switched on a very high surge of current will flow due to the low (cold) resistance of the bulb filaments. When the filament resistance increases with temperature, the current will reduce, thus illustrating the need for a fuse to be able to carry a higher current for a short time.

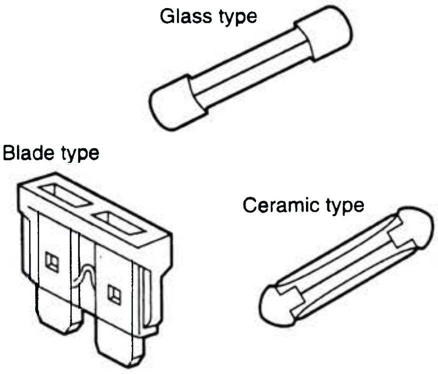


Figure 4.13 Different types of fuse

Key fact

It is common practice to protect almost all electrical circuits with a fuse.

Definition

Fuse ratings: The continuous value is the current that the fuse will carry without risk of failure, peak value is the current that the fuse will carry for a short time without failing.

Figure 4.14 Blade fuses

To calculate the required value for a fuse, the maximum possible continuous current should be worked out. It is then usual to choose the next highest rated fuse available. Blade fuses are available in a number of continuous rated values as listed in Table 4.8 together with their colour code.

Table 4.8 Fuse ratings and colours

Continuous current (A)	Colour
Blade type	
3	Violet
4	Pink
5	Clear/Beige
7.5	Brown
10	Red
15	Blue
20	Yellow
25	Neutral/White
30	Green
Ceramic type	
5	Yellow
8	White
16	Red
25	Blue

The chosen value of a fuse as calculated above must protect the consumer as well as the wiring. A good example of this is a fuse in a wiper motor circuit. If a value were used that is much too high, it would probably still protect against a severe short circuit. However, if the wiper blades froze to the screen, a large value fuse would not necessarily protect the motor from overload.

It is now common practice to use fusible links in the main output feeds from the battery as protection against major short circuits in the event of an accident or error in wiring connections. These links are simply heavy duty fuses and are rated in values such as 50, 100 or 150 A.

Occasionally, circuit breakers are used in place of fuses, this being more common on heavy vehicles. A circuit breaker has the same rating and function as a fuse but with the advantage that it can be reset. The disadvantage is the much higher cost. Circuit breakers use a bimetallic strip which, when subjected to excessive current, will bend and open a set of contacts. A latch mechanism prevents the contacts from closing again until a reset button is pressed.

Key fact

A circuit breaker has the same rating and function as a fuse but with the advantage that it can be reset.

4.2.6 Terminations

Many types of terminals are available and have developed from early bullet-type connectors into the high quality waterproof systems now in use. A popular choice for many years was the spade terminal. This is still a standard choice for connection to relays for example, but is now losing ground to the smaller blade or round terminals as shown in Figure 4.15. Circular multipin connectors are used in many cases, the pins varying in size from 1 mm to 5 mm. With any type of multipin connector, provision must always be made to prevent incorrect connection.

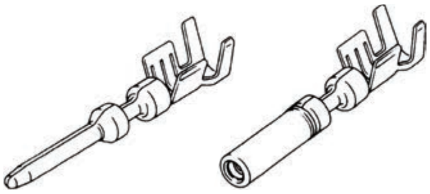


Figure 4.15 'Round' Crimp terminals

Protection against corrosion of the actual connector is provided in a number of ways. Earlier methods included applying suitable grease to the pins to repel water. It is now more usual to use rubber seals to protect the terminals, although a small amount of contact lubricant can still be used.

Many multiway connectors employ some kind of latch to prevent individual pins working loose, and also the complete plug and socket assembly is often latched. Figure 4.16 shows a common type of connector. Note the latch for security.

Figure 4.16 Terminals and wires

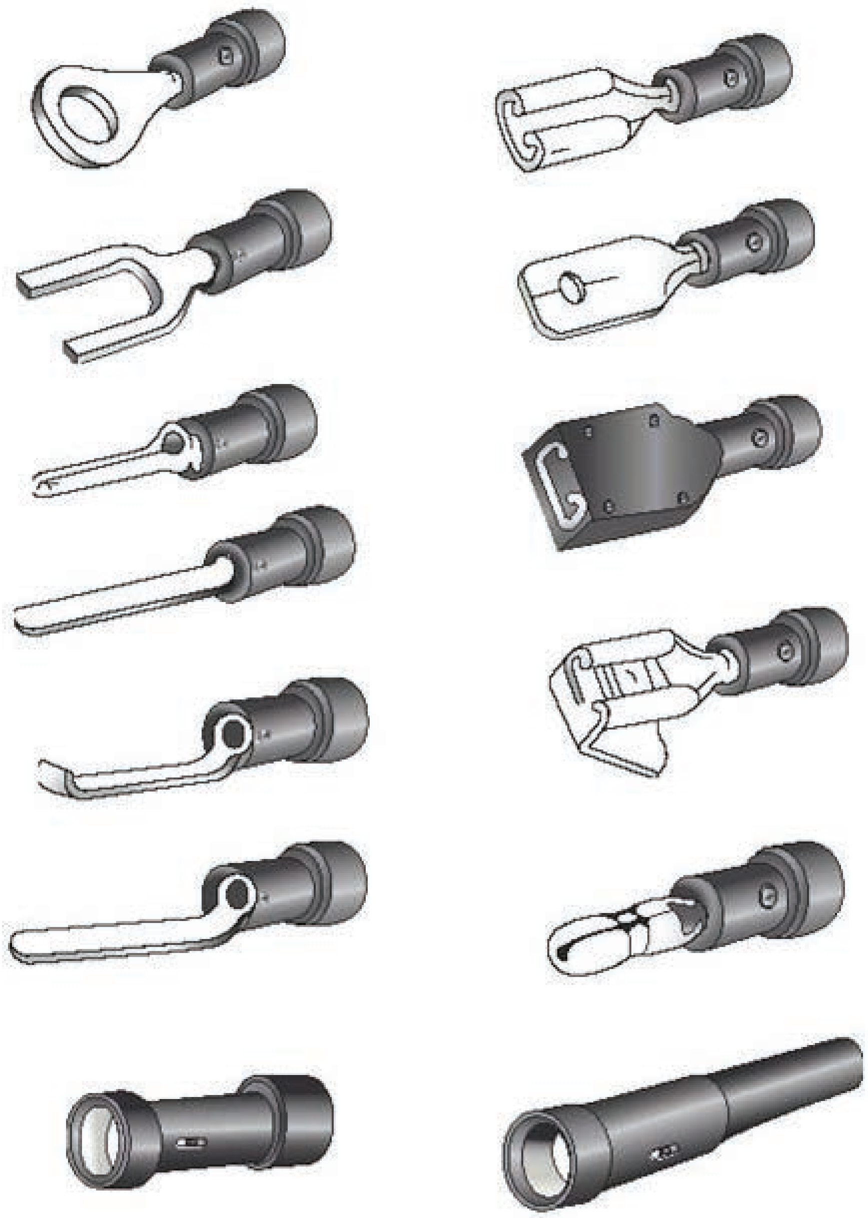


Figure 4.17 Crimp terminals for repair work

For high quality electrical connections, the contact resistance of a terminal must be kept to a minimum. This is achieved by ensuring a tight join with a large surface area in contact, and by using a precious metal coating often containing silver. It is worth noting that many connections are only designed to be removed a limited number of times before deterioration in effectiveness. This is to reduce the cost of manufacture but can cause problems on older vehicles.

Many forms of terminal are available for after-sales repair (Figure 4.17), some with more success than others. A good example is sealed terminals, which in some cases are specified by the manufacturers for repair purposes. These are pre-insulated polyamide terminations that provide a tough, environment resistant connection for most wire sizes used on motor vehicles. They simultaneously insulate, seal and protect the joint from abrasion and mechanical abuse. The stripped wire is inserted into the metallic barrel and crimped in the usual way. The tubing is then heated and adhesive flows under pressure from the tubing, filling any voids and providing an excellent seal with the cable. The seal prevents the ingress of water and other fluids, preventing electrolytic action. The connection is also resistant to temperature changes.

4.2.7 Switches

Developments in ergonomics and styling have made the simple switch into quite a complex issue. The method of operation of the switch must meet various criteria. The grouping of switches to minimize driver fatigue and distraction, access to a switch in an emergency and hazards from switch projections under impact conditions are just some of the problems facing the designer. It has now become the norm for the main function switches to be operated by levers mounted on the steering column. These functions usually include; lights, dip, flash, horn, washers and wipers. Other control switches are mounted within easy reach of the driver on or near the instrument fascia panel. As well as all the design constraints already mentioned, the reliability of the switch is important. Studies have shown that, for example, a headlamp dip switch may be operated in the region of 22 000 times during 80 000 km (50 000) miles of vehicle use (about 4 years). This places great mechanical and electrical stress on the switch.

A simple definition of a switch is ‘a device for breaking and making the conducting path for the current in a circuit’. This means that the switch can be considered in two parts; the contacts, which perform the electrical connection, and the mechanical arrangement, which moves the contacts. There are many forms of operating mechanisms, all of which make and break the contacts. Figure 4.18 shows just one common method of sliding contacts.

Definition

Switch: A device for breaking and making the conducting path for the current in a circuit.

The characteristics the contacts require are simple:

1. Resistance to mechanical and electrical wear.
2. Low contact resistance.
3. No build-up of surface films.
4. Low cost.

Materials often used for switch contacts include copper, phosphor bronze, brass, beryllium copper and in some cases silver or silver alloys. Gold is used for contacts in very special applications. The current that a switch will have to carry is the major consideration as arc erosion of the contacts is the largest problem. Silver is one of the best materials for switch contacts and one way of getting around the obvious problem of cost is to have only the contact tips made from silver, by resistance welding the silver to, for example, brass connections. It is common practice now to use switches to operate a relay that in turn will operate the main part of the circuit. This allows far greater freedom in the design of the switch due to very low current, but it may be necessary to suppress the inductive arc caused by the relay winding. It must also not be forgotten that the relay is also a switch, but as relays are not constrained by design issues the very fast and positive switching action allows higher currents to be controlled.

The electrical life of a switch is dependent on its frequency of operation, the on–off ratio of operation, the nature of the load, arc suppression and other circuit details, the amount of actuator travel used, ambient temperature and humidity and vibration levels, to name just a few factors.

The range of size and types of switches used on the motor vehicle is vast, from the contacts in the starter solenoid, to the contacts in a sunroof micro switch. Figure 4.19 shows one type of motor vehicle switch together with its specifications below.

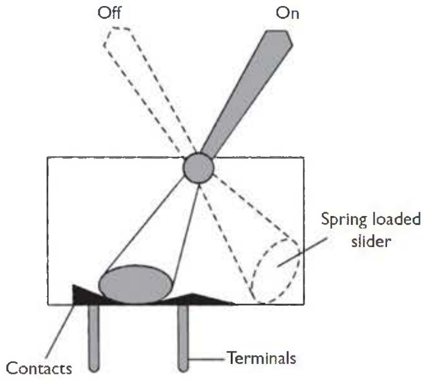


Figure 4.18 Switch with sliding contacts

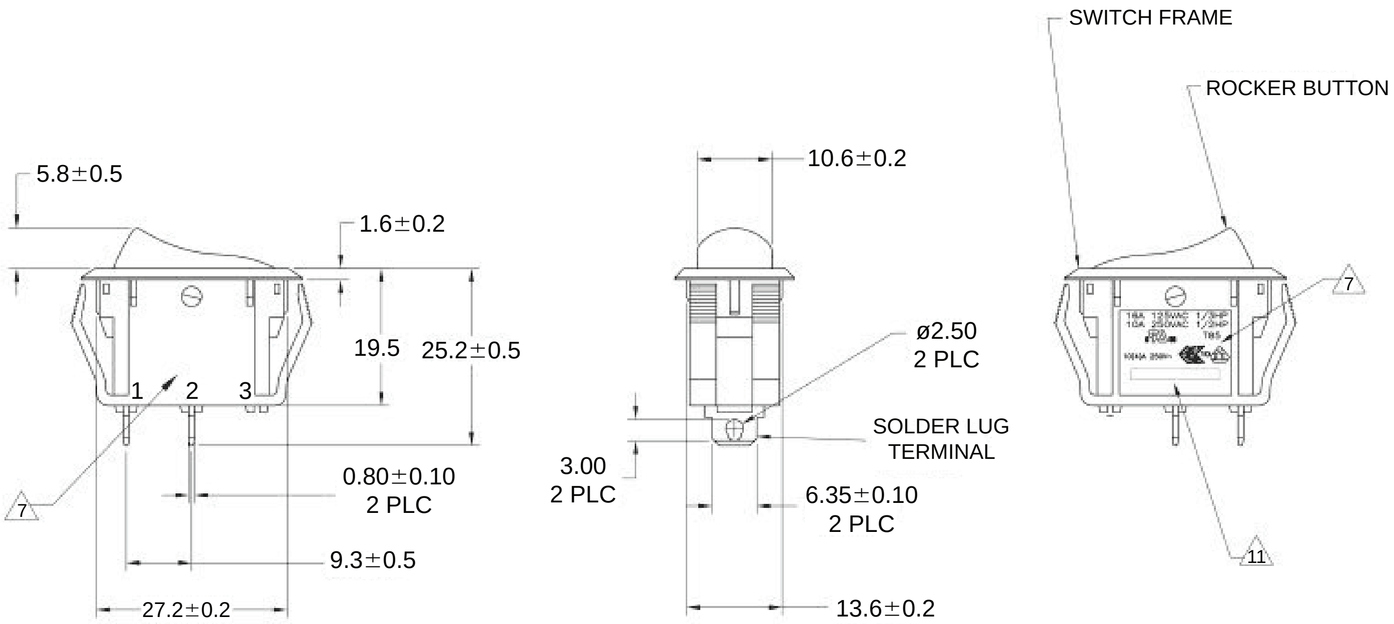


Figure 4.19 Single-pole triple-throw rocker switch

Product Type Features:
- Configuration (Pole-Throw) = Single Pole - Single Throw
- Actuator Style = Rocker
- Actuator Type = Top
- Termination Type = Quick Connect
- Mount = Panel
- UL Listed = File No. E46765

Electrical Characteristics:
- Contact Rating = 16A

Body Related Features:
- Series = PR
- Actuator Colour = Black
- Mount Style = 13.80mm x 27.60mm [.545 in. x 1.086 in.] Panel Cutout
- LED = Without

Contact Related Features:
- Contact Plating = Silver

Some of the terms used to describe switch operation are listed below.

Free position Position of the actuator when no force is applied.

Pretravel Movement of the actuator between the free and operating position.

Operating position Position the actuator takes when contact changeover takes place.

Release position Actuator position when the mechanism resets.

Overtravel Movement of the actuator beyond the operating position.

Total travel Sum of pretravel and overtravel.

Actuating force Force required to move the actuator from the free to the operating position.

Release force Force required to allow the mechanism to reset.

The number of contacts, the number of poles and the type of throw are the further points to be considered in this section. Specific vehicle current consumers require specific switching actions. Figure 4.20 shows the circuit symbols for a selection of switches and switching actions. Relays are also available with contacts and switching action similar to those shown.

So far, all the switches mentioned have been manually operated. Switches are also available, however, that can operate due to temperature, pressure and inertia, to name just three. These three examples are shown in Figure 4.21. The temperature switch shown is typical of those used to operate radiator cooling fans and it operates by a bimetal strip which bends due to temperature and causes

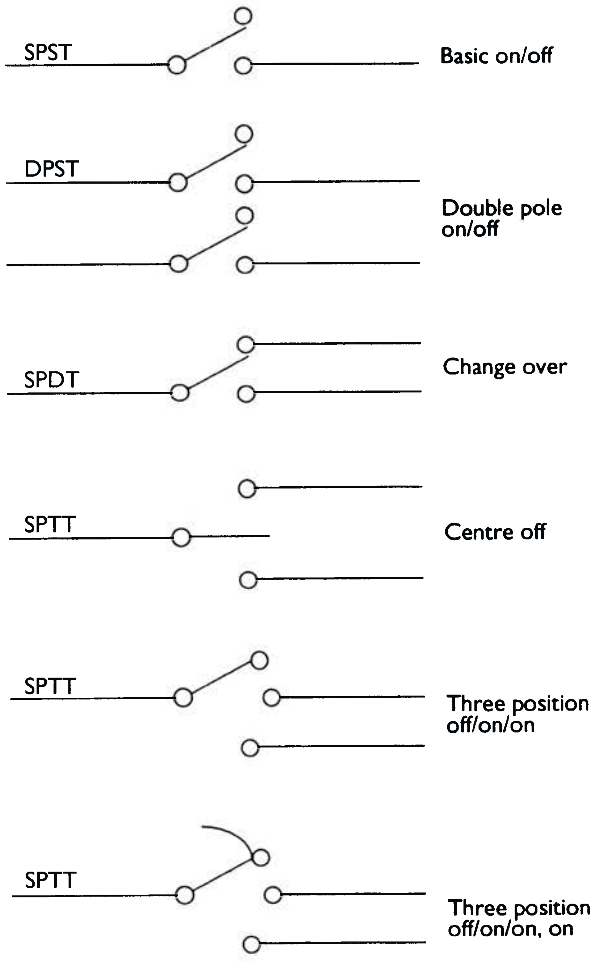


Figure 4.20 Circuit symbols for a selection of switches and switching actions

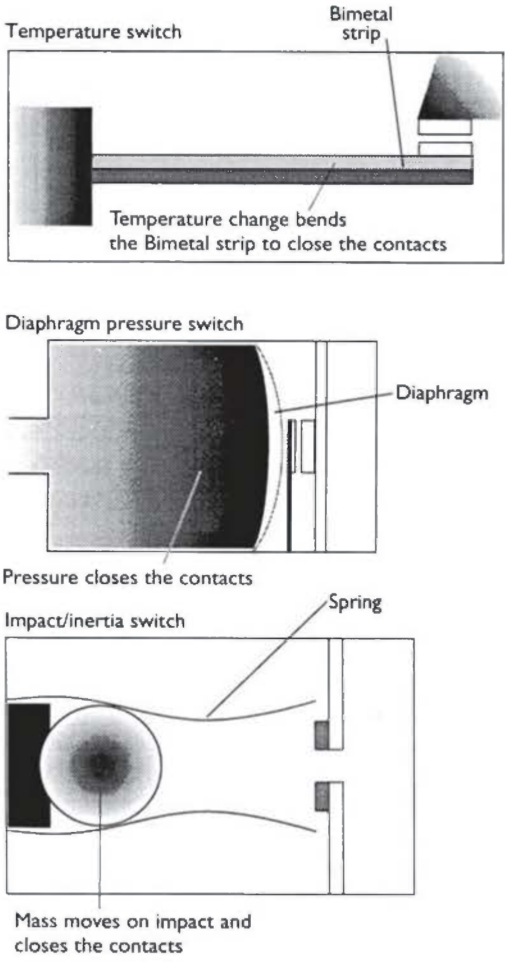


Figure 4.21 Temperature, pressure and inertia switches

a set of contacts to close. The pressure switch shown could be used to monitor over-pressure in an air conditioning system and simply operates by pressure on a diaphragm which, at a pre-determined pressure, will overcome spring tension and close (or open) a set of contacts. Finally, the inertia switch is often used to switch off the supply to a fuel injection pump in the event of an impact to the vehicle.

4.3 Multiplexing

4.3.1 Limits of the conventional wiring system

The complexity of modern wiring systems has been increasing steadily over the last 35 years or so and, in recent years, has increased dramatically. It has now reached a point where the size and weight of the wiring harness is a major problem. The number of separate wires required on a top-of-the-range vehicle can be in the region of 1500! The wiring loom required to control all functions in or from the driver's door can require up to 50 wires, the systems in the dashboard area alone can use over 100 wires and connections. This is clearly becoming a problem as, apart from the obvious issues of size and weight, the number of connections and the number of wires increases the possibility of faults developing. It has been estimated that the complexity of the vehicle wiring system doubles every 10 years.

The number of systems controlled by electronics is continually increasing. A number of these systems are already in common use and the others are becoming more widely adopted. Some examples of these systems are listed below:

- Engine management.
- Anti-lock brakes.
- Traction control.
- Variable valve timing.
- Transmission control.
- Active suspension.
- Communications.
- Multimedia.

Key fact

Many sensors that provide inputs to one electronic control unit can also be used by all or some of the others.

All the systems listed above work in their own right but are also linked to each other. Many of the sensors that provide inputs to one electronic control unit are common to all or some of the others. One solution to this is to use one computer to control all systems. This, however, would be very expensive to produce in small numbers. A second solution is to use a common data bus. This would allow communication between modules and would make the information from the various vehicle sensors available to all sensors.

Taking this idea a stage further, if data could be transmitted along one wire and made available to all parts of the vehicle, then the vehicle wiring could be reduced to just three wires. These wires would be a mains supply, an earth connection and a signal wire. The idea of using just one line for many signals is not new and has been in use in areas such as telecommunications for many years. Various signals can be 'multiplexed' on to one wire in two main ways – frequency division and time division multiplexing. Frequency division is similar to the way radio signals are transmitted. It is oversimplifying a complex subject, but a form of time division multiplexing is generally used for transmission of digital signals.

A ring main or multiplexed wiring system is represented in Figure 4.22. This shows that the data bus and the power supply cables must 'visit' all areas of the vehicle electrical system. To illustrate the operation of this system, consider the events involved in switching the sidelights on and off. First, in response to the driver pressing the light switch, a unique signal is placed on the data bus. This signal is only recognized by special receivers built as part of each light unit assembly, and

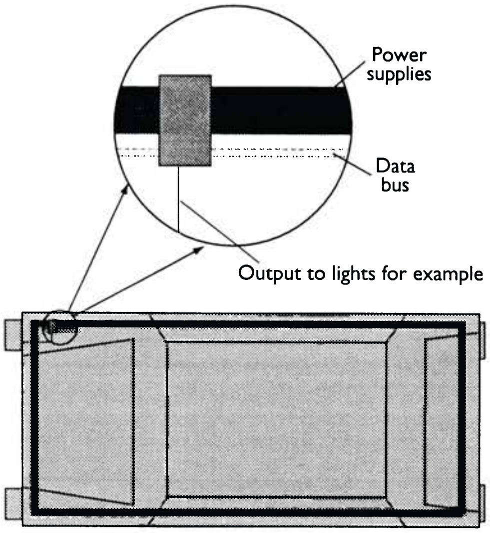


Figure 4.22 Multiplexed 'ring main' wiring system

these in turn will make a connection between the power ring main and the lights. The events are similar to turn off the lights, except that the code placed on the data bus will be different and will be recognized only by the appropriate receivers as an off code.

4.3.2 Multiplex data bus

In order to transmit different data on one line, a number of criteria must be carefully defined and agreed. This is known as the communications protocol. Some of the variables that must be defined are as follows:

- Method of addressing.
- Transmission sequence.
- Control signals.
- Error detection.
- Error treatment.
- Speed or rate of transmission.

The physical layer must also be defined and agreed. This includes the following:

- Transmission medium, e.g. copper wire, fibre optics etc.
- Type of transmission coding, e.g. analogue or digital.
- Type of signals, e.g. voltage, current or frequency etc.

The circuit to meet these criteria is known as the bus interface and will often take the form of a single integrated circuit. This IC will, in some cases, have extra circuitry in the form of memory for example. It may, however, be appropriate for this chip to be as cheap as possible due to the large numbers required on a vehicle.

4.3.3 Overview

The number of vehicle components which are networked, has considerably increased the requirements for the vehicle control systems to communicate with one another. The CAN (Controller Area Network) developed by Bosch is today's communication standard in passenger cars. However, there are a number of other systems.

Multiplexing is a process of combining several messages for transmission over the same signal path. The signal path is called the data bus. The data bus is basically just a couple of wires connecting the control units together. A data bus consists of a communication or signal wire and a ground return, serving all multiplex system nodes. The term node is given to any sub-assembly of a multiplex system (such as a control unit) that communicates on the data bus.

On some vehicles, early multiplex systems used three control units. These were the door control unit, the driver's side control unit and the passenger's side control unit (Figure 4.23). These three units replaced the following:

- Integrated unit.
- Interlock control unit.
- Door lock control unit.
- Illumination light control.
- Power window control unit.
- Security alarm control unit.

When a switch is operated, a coded digital signal is generated and communicated, according to its priority, via the data bus. All control units receive the

Definition

CAN: Controller Area Network.

Definition

Multiplexing is a process of combining several messages for transmission over the same signal path.

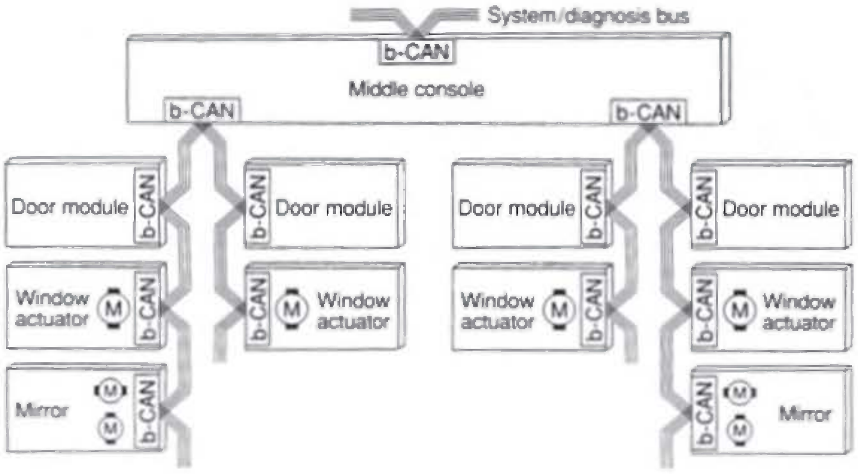


Figure 4.23 Sub-system for doors on an earlier system

signal but only the control unit for which the signal is intended will activate the desired response.

Only one signal can be sent on the BUS at any one time. Therefore each signal has an identifier that is unique throughout the network. The identifier defines not only the content but also the priority of the message. Some systems make changes or adjustments to their operation much faster than other systems. Therefore, when two signals are sent at the same time, it is the system which requires the message most urgently whose signal takes priority.

A multiplex control system has the advantage of self-diagnosis. This allows quick and easy troubleshooting and verification using diagnostic trouble codes (DTCs).

Many vehicles contain over a kilometre of wiring to supply all their electrical components. Luxury models may contain considerably more because of elaborate drivers' aids. The use of multiplexing means that considerably less wiring is used in a vehicle along with fewer multi-plugs and connectors etc (Figure 4.28).

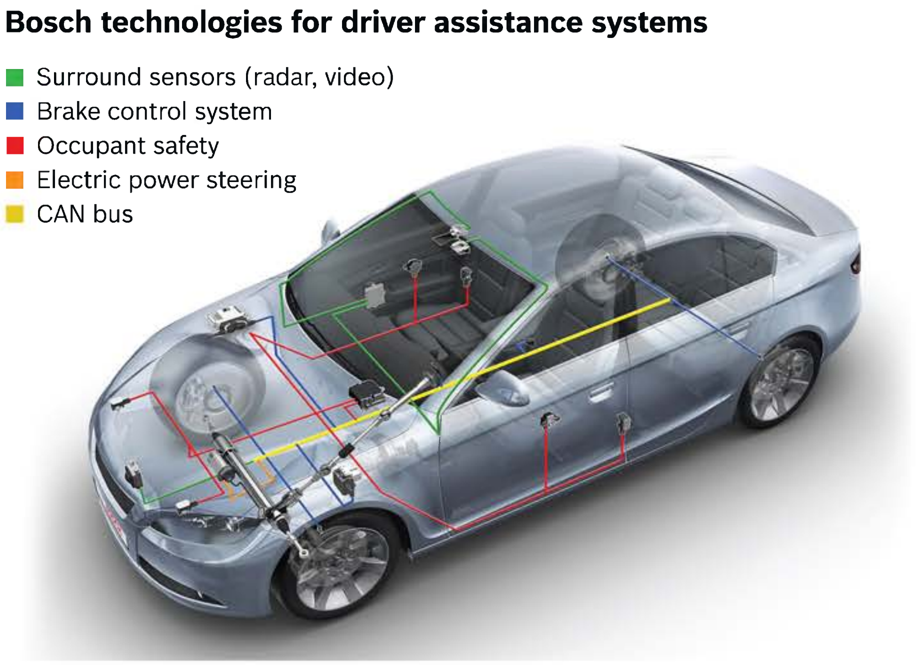


Figure 4.24 A data bus connects all networked components

An additional advantage of multiplexing is that existing systems can be upgraded or added to without modification to the original system.

4.3.4 Controller Area Network (CAN)

CAN is a serial bus system especially suited for networking 'intelligent' devices as well as sensors and actuators within a system or sub-system. It operates in a broadly similar way to a wired computer network. CAN stands for controller area network and means that control units are able to interchange data. CAN is a high-integrity serial data communications bus for real-time applications. It operates at data rates of up to 1Mbit/s. It also has excellent error detection and confinement capabilities. CAN was originally developed by Bosch for use in cars but is now used in many other industrial automation and control applications.

CAN is a serial bus system with multi-master capabilities. This means that all CAN nodes are able to transmit data and several CAN nodes can request use of the bus simultaneously. In CAN networks there is no addressing of subscribers or stations, like on a computer network, but instead, prioritized messages are transmitted. A transmitter sends a message to all CAN nodes (broadcasting). Each node decides on the basis of the identifier received whether it should

Key fact

CAN is a serial bus system with multi-master capabilities.

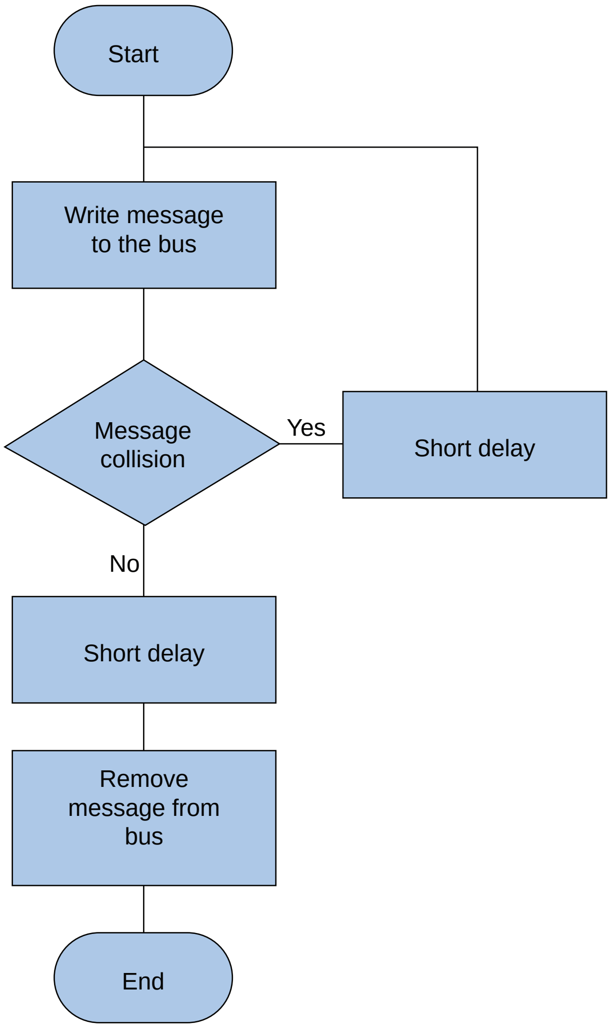


Figure 4.25 Much simplified CAN message protocol flowchart

process the message or not. The identifier also determines the priority that the message enjoys in competition for bus access.

Fast controller area network (F-CAN) and basic (or body) controller area network (B-CAN) share information between multiple electronic control units (ECUs). B-CAN communication is transmitted at a slower speed for convenience related items such as electric windows. F-CAN information moves at a faster speed for real time functions such as fuel and emissions systems. To allow both systems to share information, a control module translates information between B-CAN and F-CAN.

The ECUs on the B-CAN and F-CAN transmit and receive information in the form of structured messages that may be received by several different ECUs on the network at one time. These messages are transmitted and received across a communication circuit that consists of a single wire that is shared by all the ECUs. However, as messages on the F-CAN network are typically of higher importance, a second wire is used for communication circuit integrity monitoring. This CAN-H and CAN-L circuit forms the CAN-bus.

A multiplex control unit is often combined with the under-dash fuse/relay box. It controls many of the vehicle systems related to body electrics and the B-CAN. It also carries out much of the remote switching of various hardwired and CAN controlled systems.

One of the outstanding features of the CAN protocol is its high transmission reliability. The CAN controller registers a station's error and evaluates it statistically in order to take appropriate measures. These may extend to disconnecting the CAN node producing the errors.

Each CAN message can transmit from 0 to 8 bytes of user information. Longer messages can be sent by using segmentation, which means slicing a longer message into smaller parts. The maximum transmission rate is specified as 1 Mbit/s. This value applies to networks up to 40 m which is more than enough for normal cars and trucks.

CAN is a serial bus system designed for networking ECUs as well as sensors and actuators. CAN, originally developed by Bosch, stands for controller area network and means that control units are able to share and exchange data.

Key fact

Fast controller area network (F-CAN) and basic (or body) controller area network (B-CAN) share information between multiple electronic control units (ECUs).

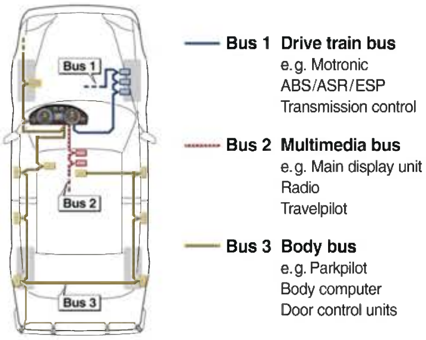


Figure 4.26 Three different speed buses in use (Source: Bosch Media)

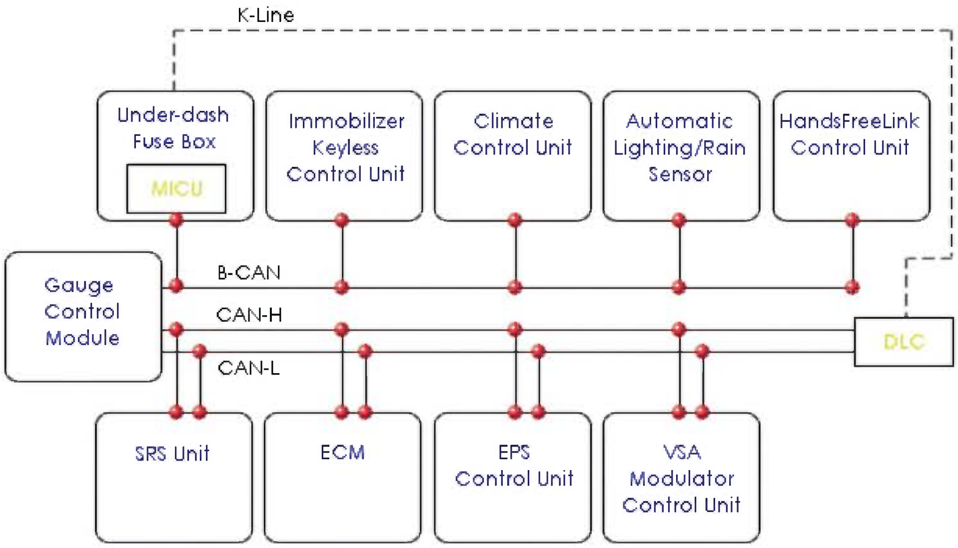


Figure 4.27 F-CAN uses CAN-H (high) and CAN-L (low) wires

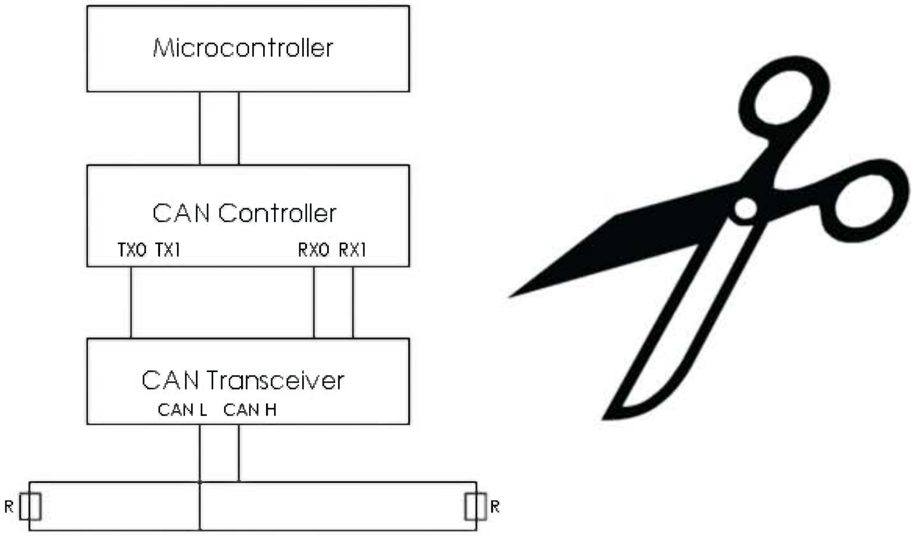


Figure 4.28 CAN nodes can be disconnected by the control program

4.3.5 CAN data signal

The CAN message signal consists of a sequence of binary digits or bits. A high voltage present indicates the value 1, a low or no voltage indicates 0. The actual message can vary between 44 and 108 bits in length. This is made up of a start bit, name, control bits, the data itself, a cyclic redundancy check (CRC) for error detection, a confirmation signal and finally a number of stop bits.

A binary format message can be something like: 100010101000101011111100001 11010111110101010101000111111010111100110011000000111111101010101000011 11111111000000001

The message identifier or name portion of the signal (part of the arbitration field) identifies the message destination and also its priority. As the transmitter puts a message on the data bus it also reads the name back from the bus. If the name is not the same as the one it sent, then another transmitter must be in operation, which has a higher priority. If this is the case it will stop transmission of its own message. This is very important in the case of motor vehicle data transmission.

Key fact

The CAN message signal consists of a sequence of binary digits or bits.

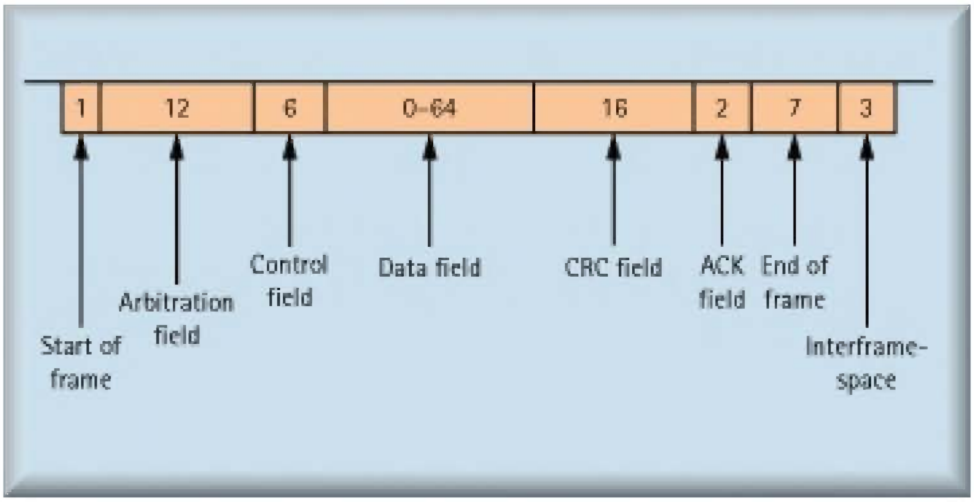


Figure 4.29 Message format (the 3 spaces are not part of the message)

Definition

The cyclic redundancy check (CRC) field is part of the overall message (The basic idea behind CRCs is to treat the message string as a single binary word M, and divide it by a key word k that is known to both the transmitter and the receiver. The remainder r left after dividing M by k constitutes the 'check word' for the given message. The transmitter sends both the message string M and the check word r, and the receiver can then check the data by repeating the calculation, dividing M by the key word k, and verifying that the remainder is r.)

Key fact

A CAN message may vary between 44 and 108 bits in length.

Definition

DLC: Diagnostic/Data link connector.

Errors in a message are recognized by what is known as a cyclic redundancy check (CRC). This is an error detection scheme in which all the bits in a block of data are divided by a predetermined binary number. A check character, known to the transmitter and receiver, is determined by the remainder. If an error is recognised the message on the bus is destroyed. This in turn is recognized by the transmitter, which then sends the message again. This technique, when combined with additional tests, makes it possible to discover all faulty messages.

The cyclic redundancy check (CRC) field is part of the overall message. (The basic idea behind CRCs is to treat the message string as a single binary word M, and divide it by a key word k that is known to both the transmitter and the receiver. The remainder r left after dividing M by k constitutes the 'check word' for the given message. The transmitter sends both the message string M and the check word r, and the receiver can then check the data by repeating the calculation, dividing M by the key word k, and verifying that the remainder is r.)

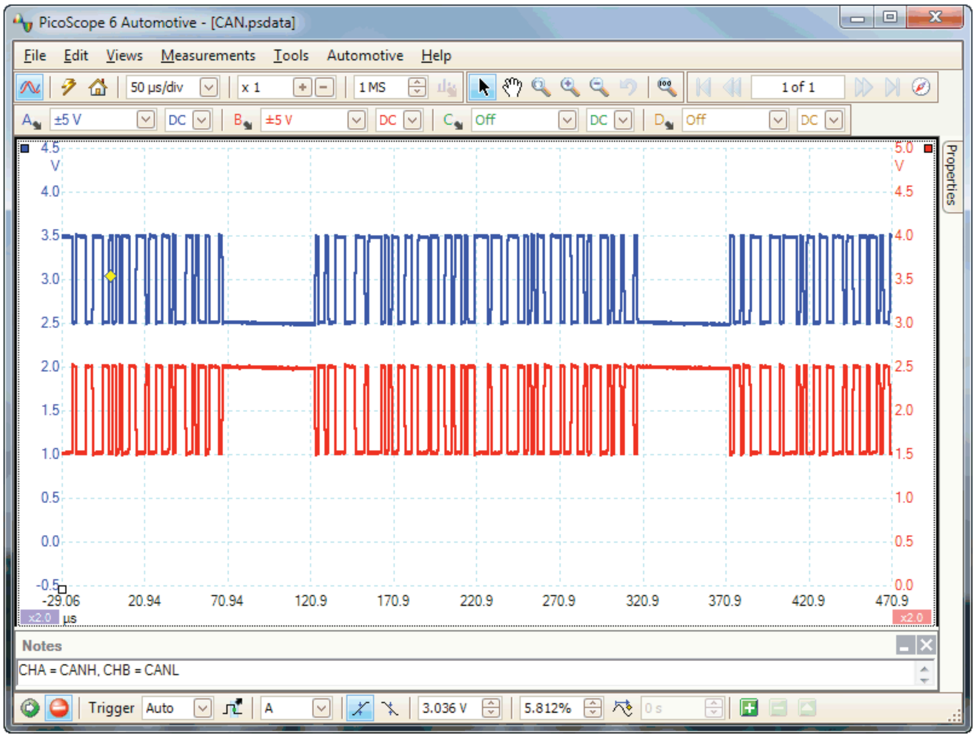
Because each node in effect monitors its own output, interrupts disturbed transmissions, and acknowledges correct transmissions, faulty stations can be recognized and uncoupled (electronically) from the bus. This prevents other transmissions from being disturbed.

A CAN message may vary between 44 and 108 bits in length. This is made up of a start bit, name, control bits, the data itself, CRC error detection, a confirmation signal and finally a number of stop bits.

Controller area network (CAN) signal is made up of voltage pulses that represent ones and zeros, in other words, binary signals. The data is applied to two wires known as CAN-high and CAN-low.

In Figure 4.27, it is possible to verify that data is being continuously exchanged along the CAN bus. It is also possible to check that the peak to peak voltage levels are correct and that a signal is present on both CAN lines. CAN uses a differential signal, and the signal on one line should be a coincident mirror image (the signals should line up) of the data on the other line (Figures 4.31 and 4.32).

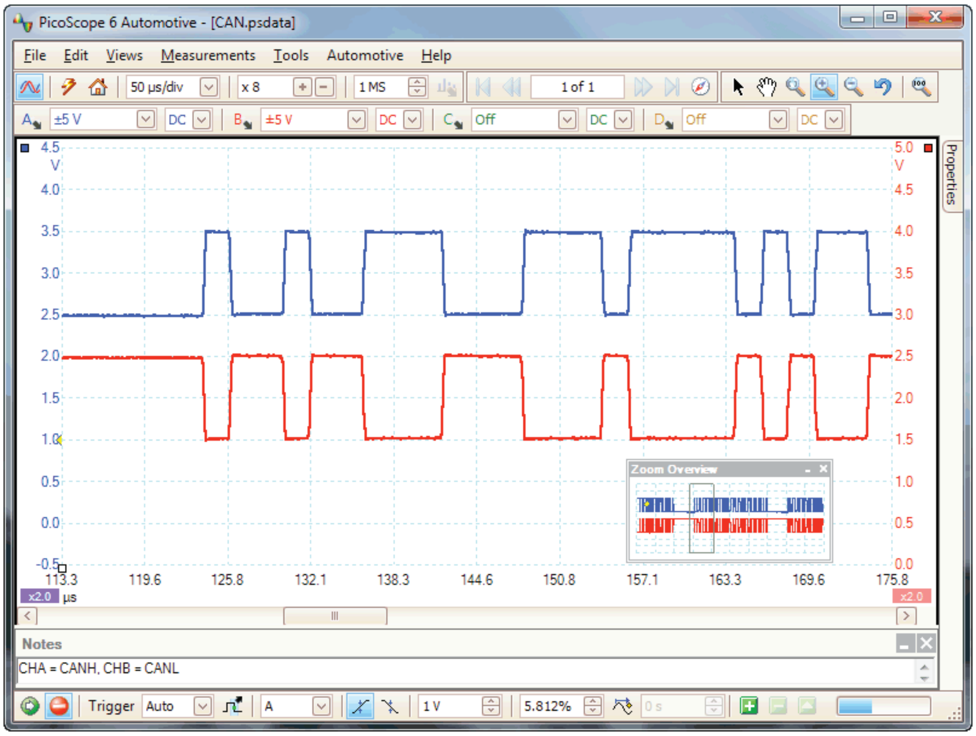
Figure 4.30 DLC socket – pin 6 is CAN-high and pin 14 is CAN-low



CHA = CANH, CHB = CANL

Figure 4.31 CAN high and low signals on a dual trace scope

The usual reason for examining the CAN signals is where a CAN fault has been indicated by OBD, or to check the CAN connection to a suspected faulty CAN node. The vehicle manufacturers' manual should be referred to for precise waveform parameters.



CHA = CANH, CHB = CANL

Figure 4.32 CAN signal zoomed in

When the signal is captured on a fast time base (or zoomed in) it allows the individual state changes to be viewed. This enables the mirror image nature of the signals, and the coincidence of the edges to be verified.

In this display, it is possible to verify that:

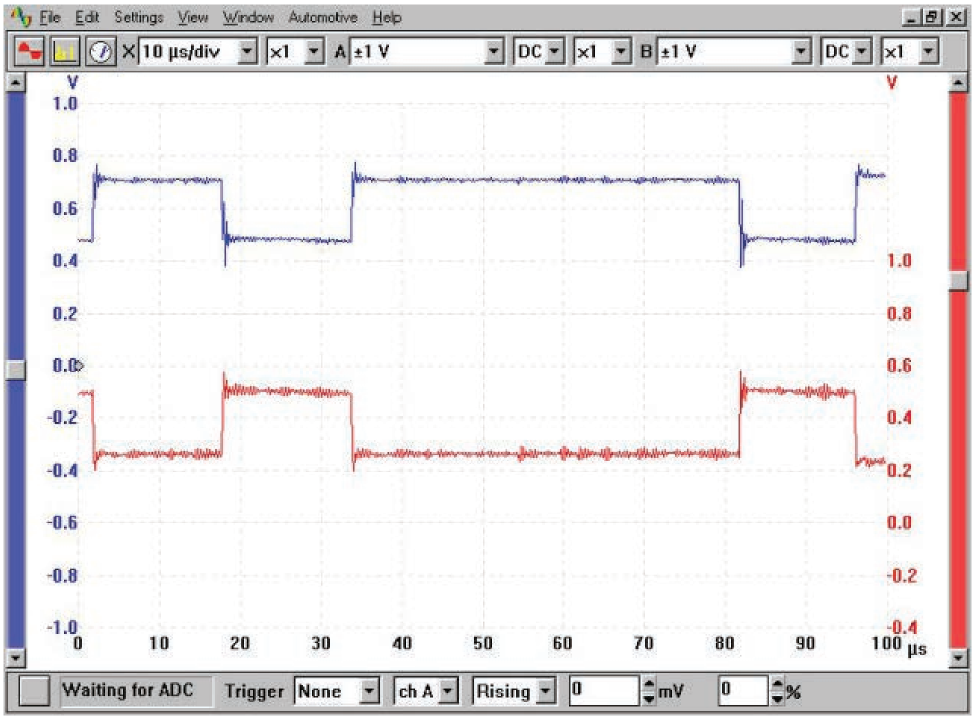
- data is being continuously exchanged along the CAN bus;
- the voltage levels are correct;
- a signal is present on both CAN lines.

Key fact

CAN uses a differential signal so the signal on one line should be a coincident mirror image of the data on the other line.

CAN uses a differential signal so the signal on one line should be a coincident mirror image of the data on the other line. The usual reasons for examining the CAN signals is where a CAN fault has been indicated by OBD, or to check the CAN connection to a suspected faulty CAN node. Manufacturers' data should be referred to for precise waveform parameters.

The following CAN data is captured on a much faster timebase and allows the individual state changes to be examined. This enables the mirror image nature of the signals, and the coincidence of the edges to be verified.



10 µs/div x1 A ±1 V DC x1 B ±1 V DC x1

Waiting for ADC Trigger None ch A Rising 0 mV 0 %

Figure 4.33 CAN signals on a fast timebase

The signals are equal and opposite and they are of the same amplitude (voltage). The edges are clean and coincident with each other. This shows that the vehicle data bus (CAN bus) is enabling communication between the nodes and the CAN controller unit. This test effectively verifies the integrity of the bus at this point in the network. If a particular node is not responding correctly, the fault is likely to be the node itself. The rest of the bus should work correctly.

It is usually recommended to check the condition of the signals present at the connector of each of the ECUs on the network. The data at each node will always be the same on the same bus. Remember that much of the data on the bus is safety critical, so do not use insulation piercing probes!

4.3.6 Local Interconnect Network (LIN)

A local interconnect network (LIN) is a serial bus system especially suited for networking 'intelligent' devices, sensors and actuators within a sub-system. It is a concept for low cost automotive networks, which complements existing automotive multiplex networks such as CAN (Figures 4.34 and 4.35).

LIN enables the implementation of a hierarchical vehicle network. This allows further quality enhancement and cost reduction of vehicles.

The LIN standard includes the specification of the transmission protocol, the transmission medium, the interface between development tools, and the interfaces for software programming. LIN guarantees the interoperability of network nodes from the viewpoint of hardware and software, and predictable electro-magnetic compatibility (EMC) behaviour.

LIN is a time triggered single master, multiple slave network concept. It is based on common interface hardware, which makes it a low cost solution (Figures 4.16 and 4.17). Additional attributes of LIN are:

- Multicast reception with self-synchronization.
- Selectable length of message frames.

Key fact

LIN is a concept for low cost automotive networks, which complements existing automotive multiplex networks such as CAN.

Definition

EMC: Electro-magnetic compatibility. EMC requirements stipulate that a device shall not cause interference within itself or in other devices, or be susceptible to interference from other devices.

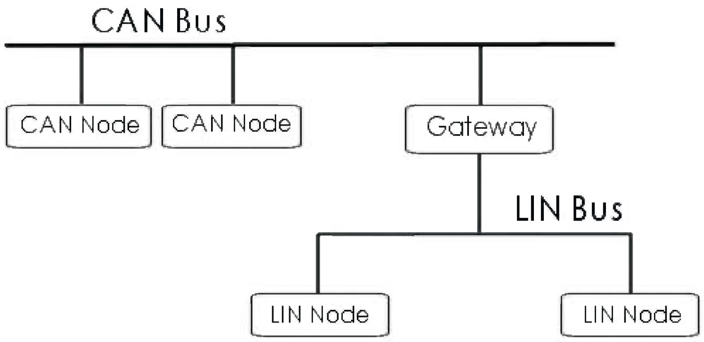


Figure 4.34 Structure using CAN and LIN

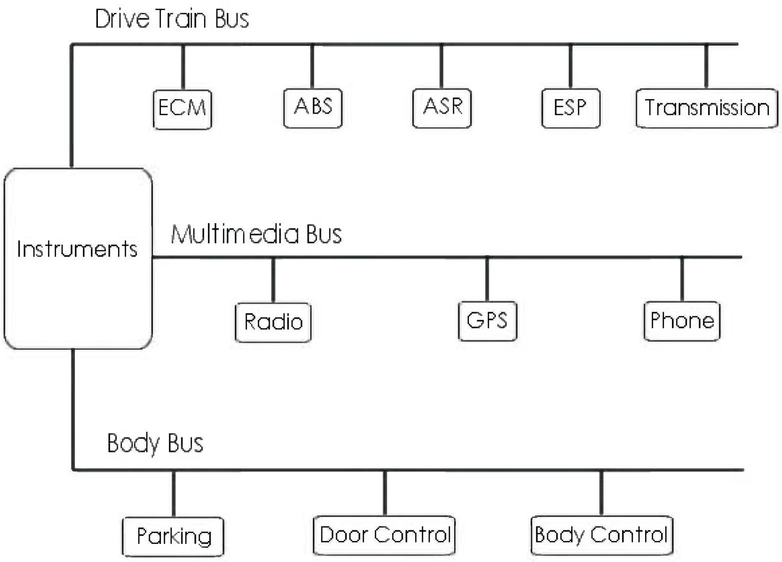


Figure 4.35 Standards allow communication between different systems

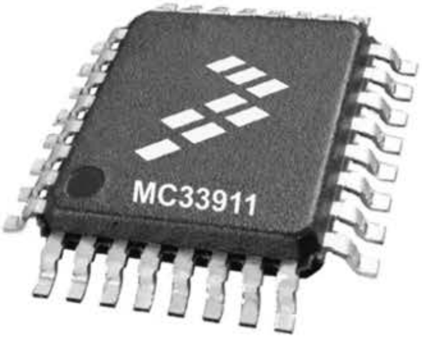


Figure 4.36 LIN package
(Source: FreeScale Semlconductor)

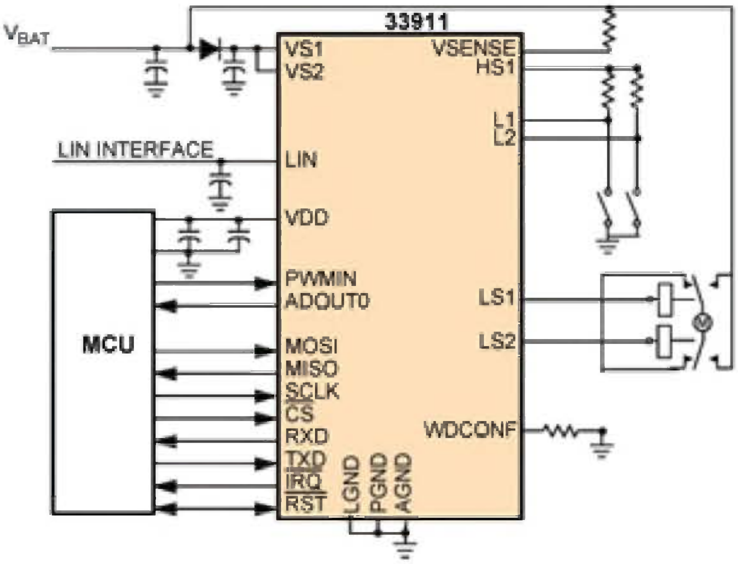


Figure 4.37 System basis chip (SBC) with a LIN transceiver (Source: FreeScaie Semiconductor)

- Data checksum security and error detection.
- Single-wire implementation.
- Speed up to 20 kBit/s.

LIN provides a cost efficient bus communication where the bandwidth and versatility of CAN are not required. It is used for non-critical systems.

The LIN bus is proving popular because of its low cost and also because it reduces the bus load of the supervising CAN network. Local interconnect network (LIN) signals can be measured by connecting between earth/ground

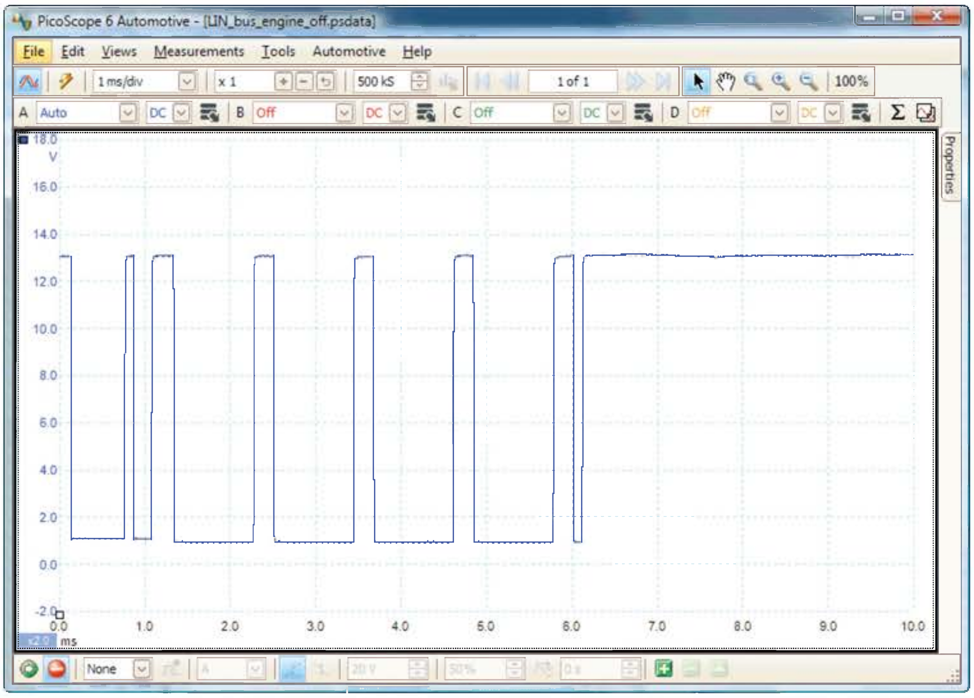

PicoScope 6 Automotive - [LIN_bus_engine_off.psdata]

Figure 4.38 LIN waveform

and the signal wire. It is not possible to decode the signal but a correctly switching square waveform should be as shown in Figure 4.38.

4.3.7 FlexRay

FlexRay is a fast and fault-tolerant bus system for automotive use (Figures 4.40 and 4.41). It was developed, using the experience of well-known OEMs. It is designed to meet the needs of current and future in-car control applications that require a high bandwidth. The bit rate for FlexRay can be programmed to values up to 10 MBit/s.

The data exchange between the control devices, sensors and actuators in automobiles is mainly carried out via CAN systems. However, the introduction of X-by-wire systems has resulted in increased requirements. This is especially so with regard to error tolerance and speed of message transmission. FlexRay meets these requirements by message transmission in fixed time slots, and by fault-tolerant and redundant message transmission on two channels.

The physical layer means the hardware, that is, the actual components and wires. FlexRay works on the principle of: time division multiple access (TDMA). This means that components or messages have fixed time slots in which they have exclusive access to the data bus. These time slots are repeated in a cycle and are just a few milliseconds long.

The fixed allocation of the bus bandwidth to the components or messages by means of fixed time slots has the disadvantage that the bandwidth is not fully used. For example, if a component is simply not in use at its slot-time. To get over this, FlexRay subdivides the cycle into static and dynamic segments. The fixed time slots are situated in the static segment at the beginning of a bus

Figure 4.39 FlexRay logo

Key fact

FlexRay can cope with the requirements of X-by-wire systems.

Key fact

FlexRay communicates via two physically separated lines with a data rate of up to 10 Mbit/son each.

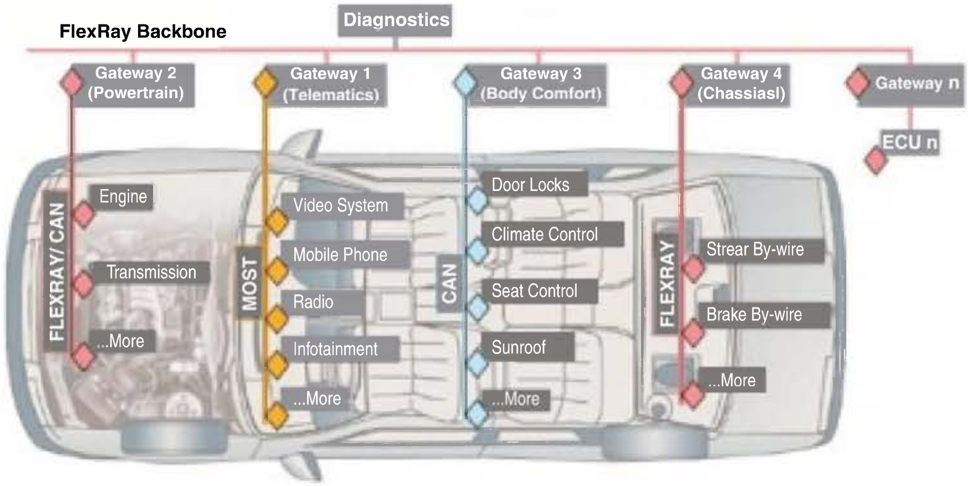


Figure 4.40 FlexRay backbone

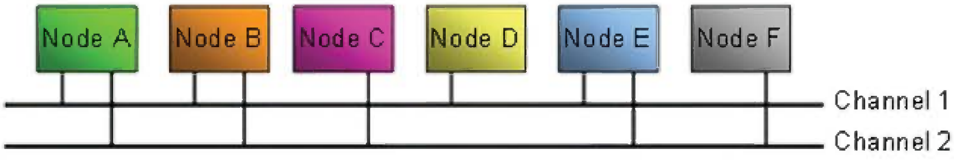


Figure 4.41 FlexRay topology with two channels (Source: Eberspacher)

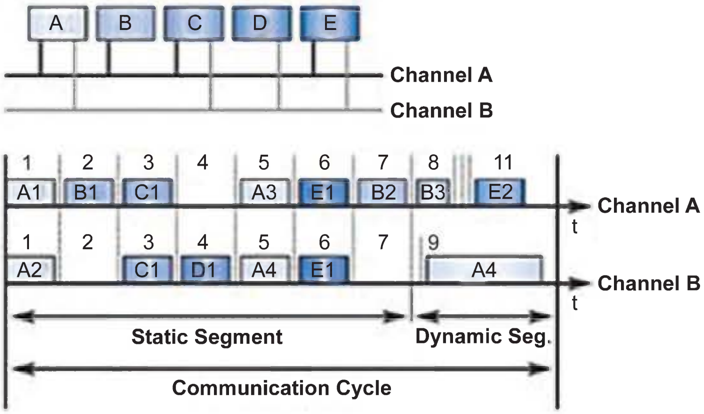


Figure 4.42 Bandwidth timeslots

cycle. In the dynamic segment, the time slots are assigned dynamically, in other words, as they are needed. Exclusive bus access is only enabled for a short time in 'mini-slots'. This mini-slot is then only extended if a bus access occurs. Bandwidth is therefore used up only when it is actually needed (Figure 4.42).

FlexRay communicates via two physically separated lines with a data rate of up to 10 Mbit/s on each. The two lines are mainly used for redundant and therefore fault-tolerant message transmission, but they can also transmit different messages.

In order to implement synchronous functions and use all available bandwidth, the distributed nodes on the network require a common time base. Clock synchronization messages are therefore transmitted in the static segment of each cycle.

A FlexRay ECU consists of a host processor, a FlexRay communication controller (CC), a bus guardian (BG) and a bus driver (BD). The host processor supplies and processes the data, which are transmitted via the controller (Figures 4.43 and 4.44). The process is as follows:

- Bus guardian monitors access to the bus.
- Host processor informs the bus guardian which time slots the communication controller has allocated.
- Bus guardian allows the communication controller to transmit data only in these time slots.
- Bus driver is enabled.

FlexRay is a fast and fault-tolerant bus system that was developed to meet the needs of high bandwidth applications such as X-by-wire systems (Figure 4.45). Error tolerance and speed of message transmission in these systems is essential.

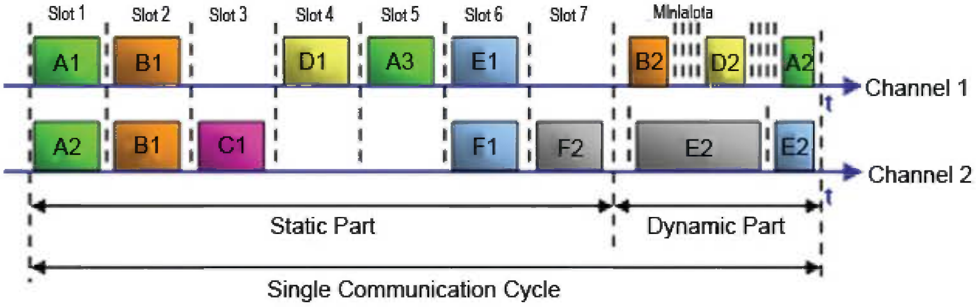


Figure 4.43 FlexRay communication cycle (Source: Eberspacher)

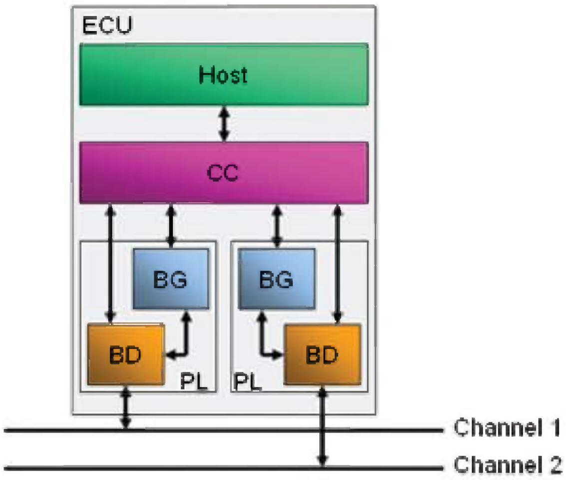


Figure 4.44 FlexRay ECU (Source: Eberspacher)

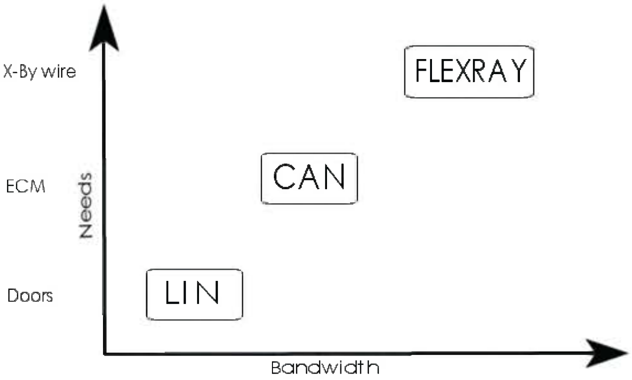


Figure 4.45 Comparing requirements and data rates of the three systems

FlexRay uses very high speed signals so it is necessary to use high speed probes to view the signal. The FlexRay-High and FlexRay-Low pins are usually available at the multi-way connector at each ECU on the network.

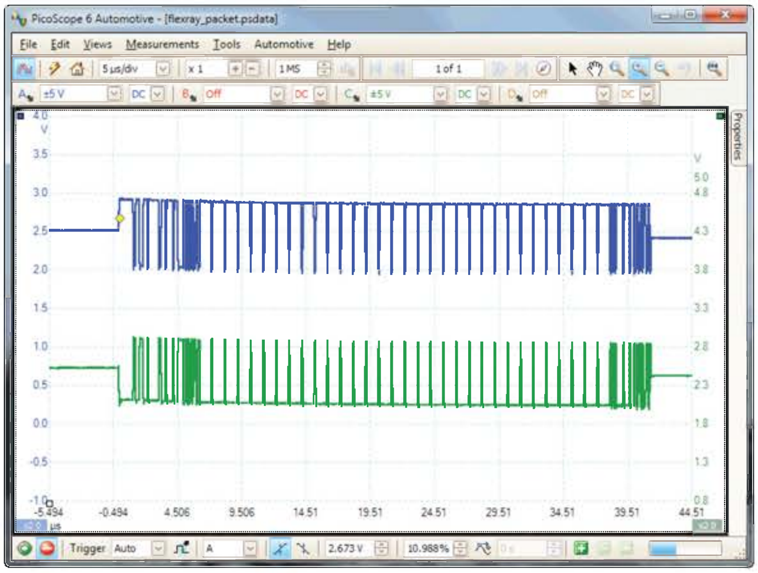


Figure 4.46 FlexRay signal

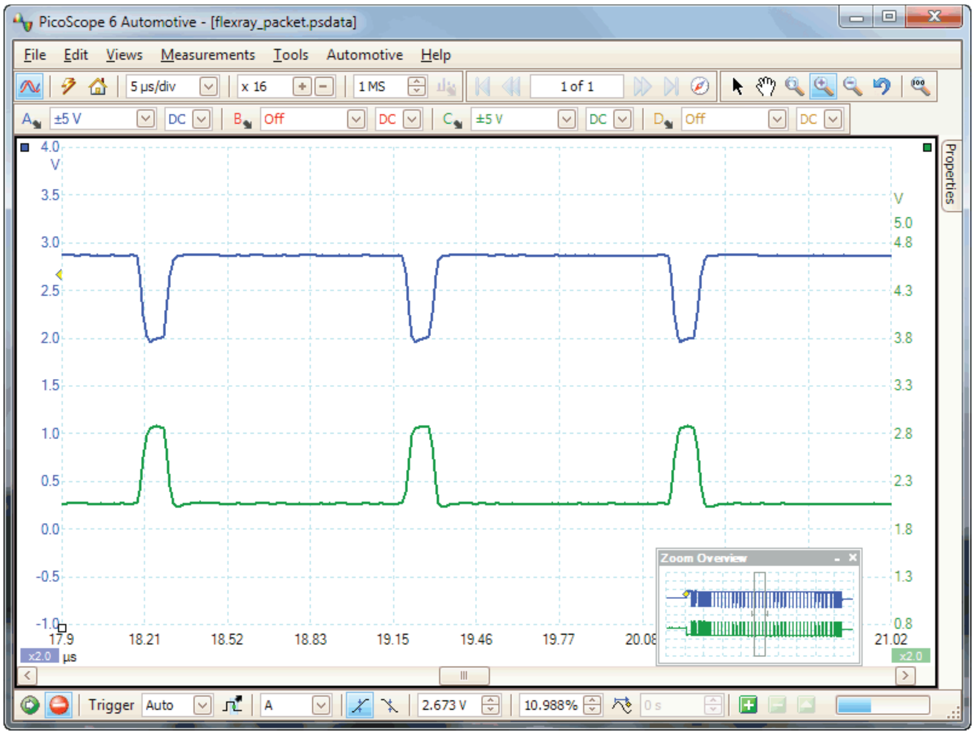


Figure 4.47 A closer view of a FlexRay signal

It is possible to verify that data is being continuously exchanged on the FlexRay network, that the peak-to-peak voltage levels are correct, and that a signal is present on both FlexRay lines. FlexRay uses a differential signal, so the signal on one line should be a mirror image of the data on the other line.

4.4 Media oriented systems transport (MOST)

4.4.1 Introduction

MOST is the de-facto standard for multimedia and infotainment networking in the automotive industry. The technology was designed from the ground up to provide an efficient and cost-effective fabric to transmit audio, video, data and control information between any devices attached even to the harsh environment of an automobile. Its synchronous nature allows for simple devices to provide content and others to render that content with the minimum of hardware. At the same time it provides unique quality of service for transmission of audio and video services. Although its roots are in the automotive industry, MOST can be used for applications in other areas such as other transportation applications, A/V networking, security and industrial applications.

Definition

A/V: Audio visual.

The MOST Cooperation is the organization through which the technology is standardized and refined so that it continues to stay abreast of the latest technology requirements.

4.4.2 MOST network

MOST is a synchronous network. A timing master supplies the clock with a synchronous and continuous data signal and all other devices synchronize their operation to this base signal. This technology eliminates the need for buffering and sample rate conversion so that very simple and inexpensive devices can be connected and the hardware of the network interface itself is lean and cost effective. The technology is similar to what the public switched telephone network uses.

Within the synchronous base data signal, multiple streaming data channels and a control channel are transported. The control channel is used to set up what streaming data channels the sender and receiver are to use. Once the connection is established, data can flow continuously and no further addressing or processing of packet label information is required. The bandwidth of the streaming data channels is always available and reserved for the dedicated stream so there are no interruptions, collisions, or slow-downs in the transport of the data stream. This is the optimum mechanism for delivering streaming data (information that flows continuously) like audio and video.

Key fact

On a MOST network, multiple streaming data channels and a control channel are transported within the synchronous base data signal.

Computer based data, such as Internet traffic or information from a navigation system, is typically sent in short (asynchronous) bursts as packets and is often going to many different places. To accommodate such signals, MOST has defined efficient mechanisms for sending asynchronous, packet based data in addition to the control channel and the streaming data channels (Figure 4.48). These mechanisms run on top of the permanent synchronous data signal. However, they are completely separate from the control channel and the streaming data channels so that none of them interfere with each other.

First conceived in 1997, MOST differs from existing vehicle bus technologies, in that it's intended to be carried largely on an optical fibre bearer, thus providing a bus-based networking system at bit-rates far higher than available on previous vehicle-bus technologies.

The MOST specification defines all seven layers of the ISO/OSI Reference Model for data communication. The MOST network often employs a ring topology, but star configurations and double rings for critical applications are possible and may include up to 64 devices or nodes. A plug and play feature enables easy adding and removing of devices. A Timing Master is one of the nodes which continuously feeds data frame into the ring or acts as the gate for data. The preamble, or packet header, repeatedly synchronizes the rest of the nodes called Timing Slaves.

Definition

ISO/OSI Reference Model: The Open Systems Interconnection (OSI) model is a way of sub-dividing a communications system into smaller parts called layers. Similar communication functions are grouped into logical layers. A layer provided services to its upper layer while receiving services from the layer below.

4.4.3 Protocol

MOST can use both optical and electrical components on the physical layer. A special code know as bi-phase mark) is used to encode data on either physical medium, such as used in S/PDIF. All bits are transmitted in a single clock cycle. A zero bit has a single transition at the end of the clock cycle. A one bit is encoded with two transitions both in the middle of the clock cycle and at the end of the clock cycle. The preamble is encoded using a clock cycle after the third clock cycle.

Definition

S/PDIF: Sony/Philips Digital Interface. A format for sending digital audio information. S/PDIF connections use optical cables or phone connectors.

A most data block has 16 frames and each comprises:

- a 4 bit preamble;
- a 4 bit boundary descriptor;
- 15 quad-byte payload that are split into a synchronous and asynchronous part, the split point is specified by the boundary descriptor;
- a 2 byte control pay-load; the control payload of 16 frames are concatenated to form a single 32 byte control message;

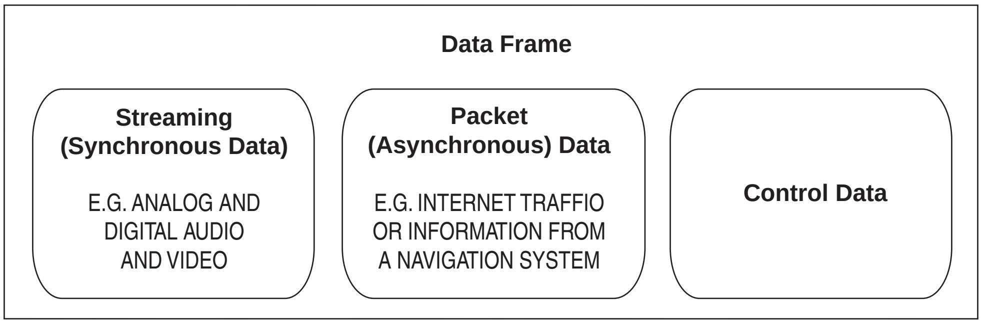


Figure 4.48 MOST data frame

- 7 bits frame control;
- a parity bit.

4.4.4 MOST applications

The simplest MOST applications consist of analog audio gateways. A simple digital-to-analog converter (DAC) can be connected to a MOST transceiver to drive a speaker. A human machine Interface (HMI) sets up the MOST transceiver to receive data from a specific channel and the transceiver drives the DAC clock based on the timing provided by the network timing master. Since both the receiver and the sender of the data are running off the same clock, no special buffering or processing of packet information is required. The MOST transceiver even provides industry-standard to connect to off-the-shelf DACs from many manufacturers.

Key fact

The simplest MOST applications consist of analog audio gateways.

A microphone can be connected to an analog-to-digital converter (ADC) that in turn connects directly to a MOST transceiver. The HMI sets up the communication channel and the ADC just streams its data to the MOST transceiver.

Video sources can be easily connected to a MOST network by using a video ADC and MPEG encoder that digitizes the video source and places it on the network. The video signals previously captured can be decoded by a video display interface. This interface could even take care of any copy protection mechanisms that were used if the data came from a DVD.

Definition

MPEG: A set of standards adopted by the moving pictures experts group for the compression of digital video and audio data.

4.4.5 Consumer device gateway

Rapidly changing technologies can be decoupled from a stable backbone that transport A/V information and control by building gateways to connect the two (Figure 4.49).

4.4.6 Summary

MOST is a multiplex network that has different channels with their own mechanisms to transport all the various signals and data streams that occur in multimedia and infotainment systems. They all run on top of a synchronous base data stream which guarantees high quality of service for the audio and video signals where disturbances are not acceptable. MOST:

- uses a single interconnection to transport audio, video, data and control information;

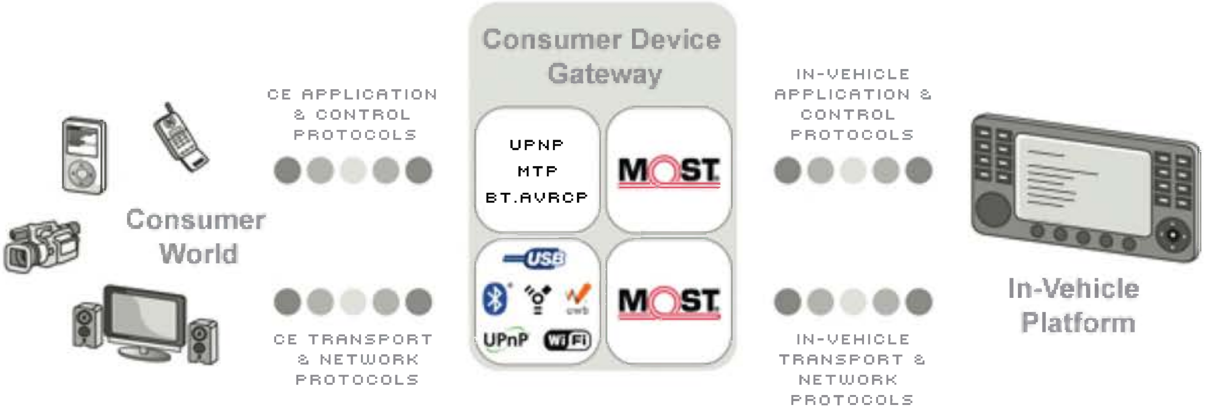


Figure 4.49 Using a gateway on a MOST network (Source: MOST Corporation)

- can use fibre-optic cable, or twisted pair wires;
- supports 25, 50 and 150 Mbps speeds;
- provides the connectivity backbone to network a range of multimedia interfaces.

Please visit: www.mostcooperation.com for more information.

4.5 Automotive Ethernet

4.5.1 Introduction

Automotive systems have tended to use custom standards such as MOST, but one of the leading automotive networking chip suppliers, SMSC, has produced a high-performance single-chip 10/100 Ethernet controller.

The device is designed specifically to meet the high reliability standards required by automotive applications. Using a high-performance Ethernet interface in today’s complex vehicle electrical systems may help diagnose issues faster and lower software maintenance time.

Definition

Ethernet: A set of network cabling and network access protocol standards for bus topology computer networks invented by Xerox but now controlled by the a subcommittee of the IEEE. It is also generally used to describe a computer network which complies with these standards.

4.5.2 Overview

The need for higher speed interfaces is driven by the increasing size of embedded program and data memories. For example, a recent BMW 7 series has more than 1 GB of memory while the previous model had just short of 100 MB. Repair shops diagnose and fix problems, but also update the software and data embedded in the various control devices inside the car via the data link connector (DLC).

This standardized connector only provides a slow communication interface so updating the software of a modern car via this interface can take hours. As a result, many car companies are working on an upgrade of the OBD connector to provide the car with a high-performance data interface for diagnostics and software downloads. This initiative is expected to lead to a new ISO/SAE standard that mandates Ethernet as part of the OBD interface for all cars.

Definition

GB: Giga bytes.

Gb: Giga bits.

Byte: 8 bits.

The device from SMSC provides a simple, parallel host bus interface to the typical automotive embedded microcontrollers used inside a car. It can function as a network branch to the outside world connecting the car to a personal computer, diagnostic tool or a complex Ethernet network in the repair shop with power management, wake-on-LAN support allows network to wake-up electronics devices from sleep state, multiple low-power modes and built-in flow control support.

For more information, visit www.smsc.com

4.6 Circuit diagrams and symbols

4.6.1 Symbols

The selection of symbols given in Figure 2.14 (in Chapter 2), is intended as a guide to some of those in use. Some manufacturers use their own variation but a standard is developing. The idea of a symbol is to represent a component in a very simple but easily recognizable form. The symbol for a motor or for a small electronic unit deliberately leaves out internal circuitry in order to concentrate on the interconnections between the various devices.

Examples of how these symbols are used are given in the next three sections, which show three distinct types of wiring diagram. Due to the complexity of modern wiring systems it is now common practice to show just part of the whole system on one sheet. For example, lights on one page, auxiliary circuits on the next, and so on. In many cases, even one system is split into two parts (see the circuits in the central electrical control section for examples).

4.6.2 Conventional circuit diagrams

The conventional type of diagram shows the electrical connections of a circuit but makes no attempt to show the various parts in any particular order or position. Figure 4.50 shows an example of this type of diagram.

4.6.3 Layout or wiring diagrams

A layout circuit diagram makes an attempt to show the main electrical components in a position similar to those on the actual vehicle. Owing to the complex circuits and the number of individual wires, some manufacturers now use two diagrams – one to show electrical connections and the other to show the actual layout of the wiring harness and components. Citroën, amongst others, have used this system. An example of this is reproduced in Figure 4.51.

4.6.4 Terminal diagrams

A terminal diagram shows only the connections of the devices and not any of the wiring. The terminal of each device, which can be represented pictorially, is marked with a code. This code indicates the device terminal designation, the destination device code and its terminal designation and, in some cases, the wire colour code. Figure 4.52 shows an example of this technique.

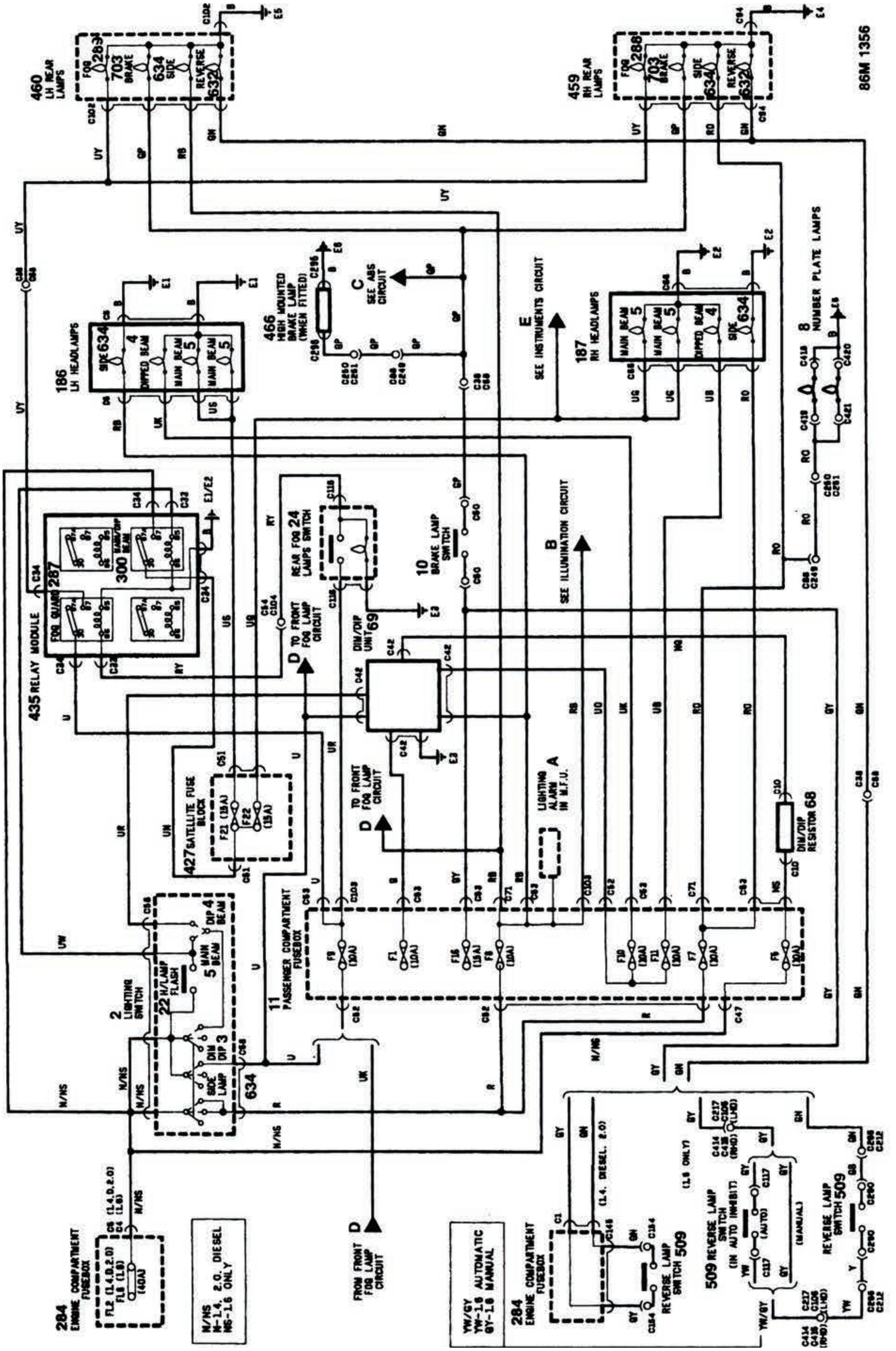


Figure 4.50 Conventional circuit diagram

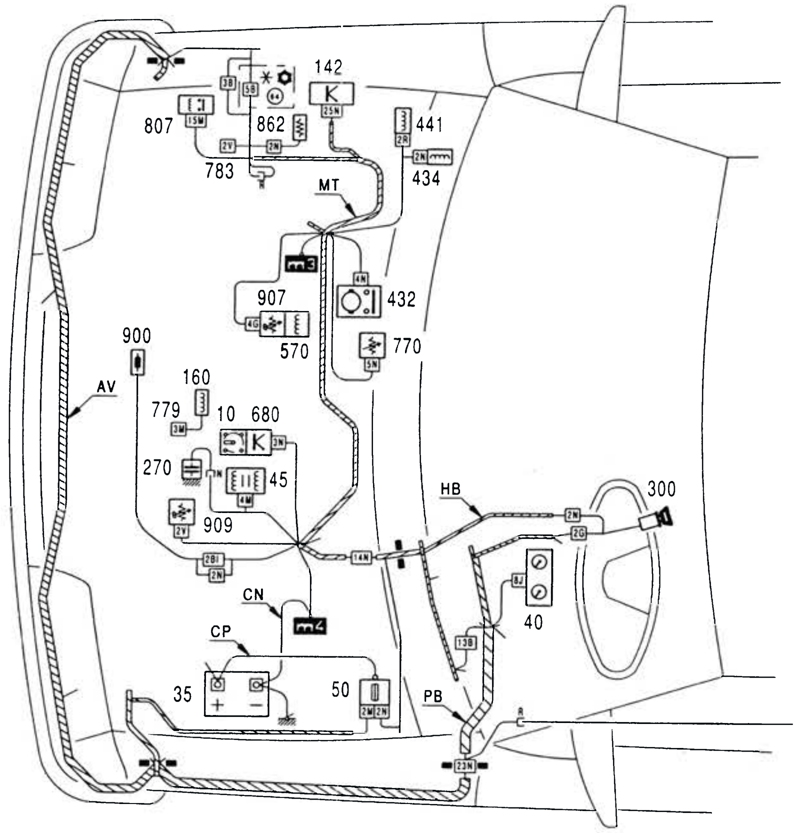


Figure 4.51 Layout diagram

Key fact

Current flow diagrams have two supply lines at the top of the page marked 30 (main battery positive supply) and 15 (ignition controlled supply) and earth/ground marked 31 at the bottom.

4.6.5 Current flow diagrams

Current flow diagrams are now very popular. The idea is that the page is laid out such as to show current flow from the top to the bottom. These diagrams often have two supply lines at the top of the page marked 30 (main battery positive supply) and 15 (ignition controlled supply). At the bottom of the page is a line marked 31 (earth or chassis connection). Figure 4.53 is an example of this technique.

4.7 Electromagnetic compatibility

Definition

EMC: Electromagnetic compatibility.

4.7.1 Introduction

Electromagnetic compatibility (EMC), describes the ability of a device or system to function without error in its intended electromagnetic environment.

Electromagnetic interference (EMI) refers to electromagnetic emissions from a device or system that interferes with the normal operation of another device or system.

Definition

EMI: Electromagnetic interference.

4.7.2 EMC problems

The following list gives some examples where EMC can become a problem:

- A computer interferes with FM radio reception.

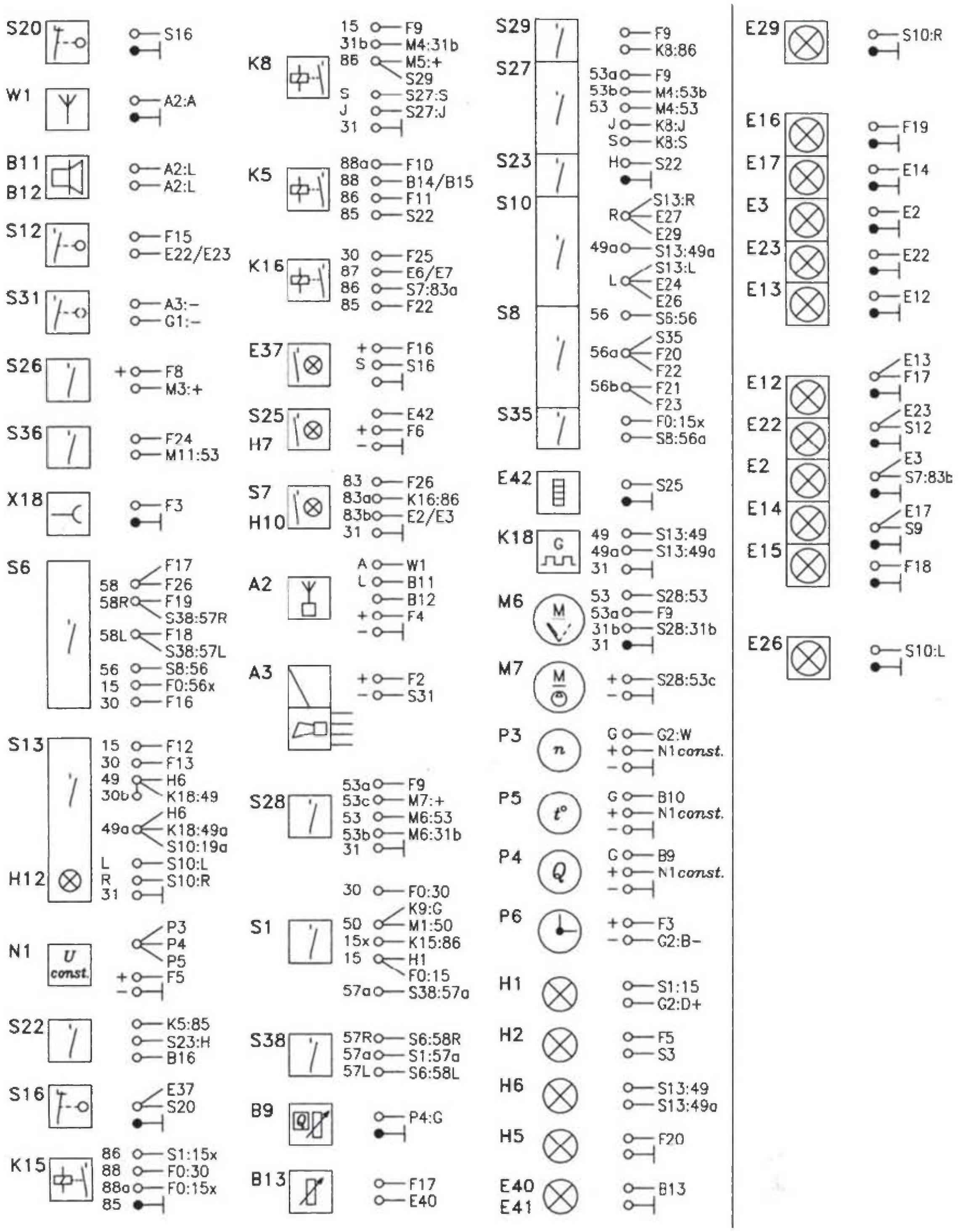


Figure 4.52 Terminal designation diagram

- A car radio buzzes when you drive under a power line.
- A car misfires when you drive under a power line.
- A helicopter goes out of control when it flies too close to a radio tower.
- CB radio conversations are picked up on the stereo.
- The screen on a video display jitters when fluorescent lights are on.
- The clock resets every time the air conditioner kicks in.
- A laptop computer interferes with an aircraft's rudder control!

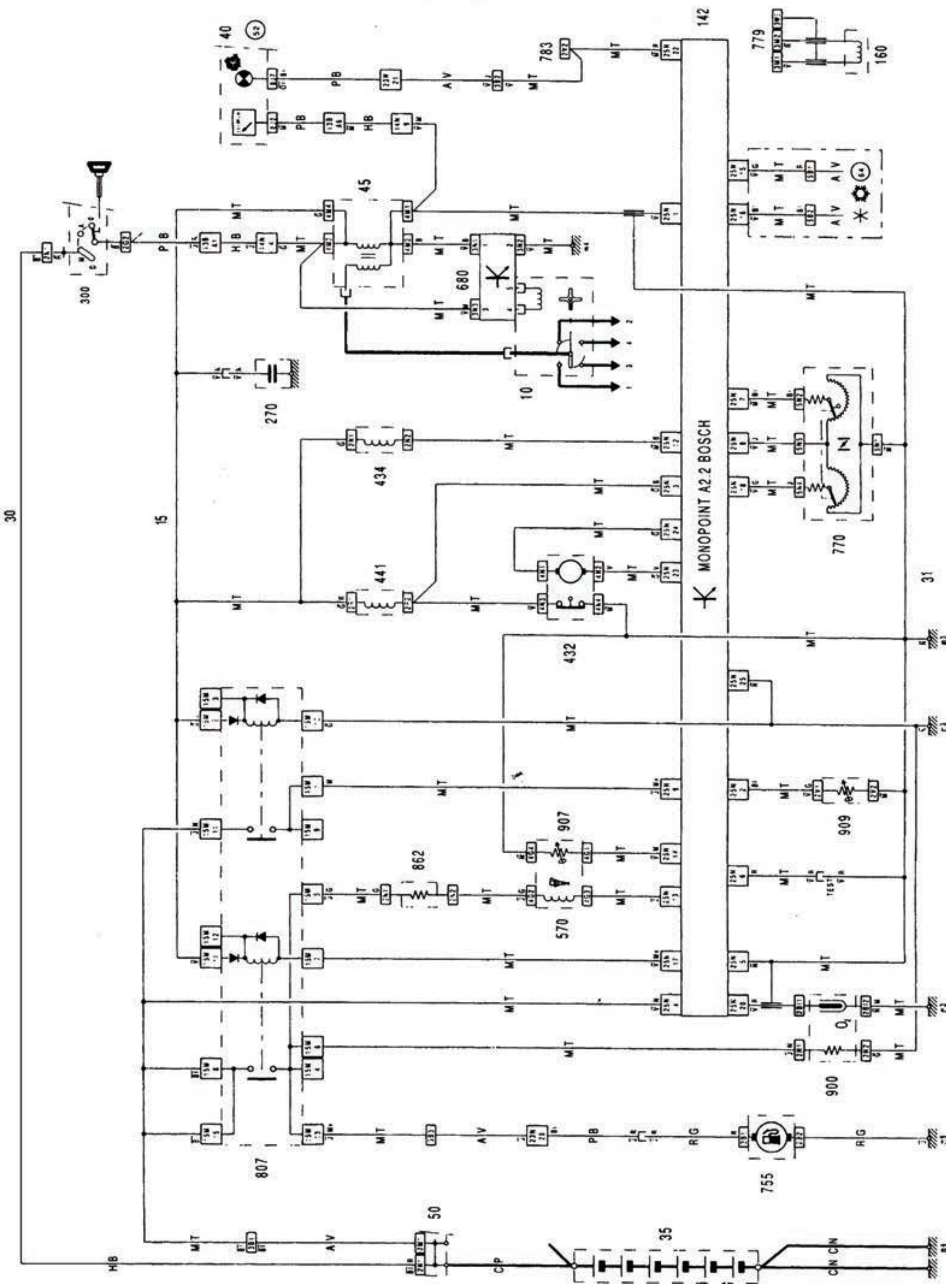


Figure 4.53 Current flow diagram

- The airport radar interferes with a laptop computer display.
- A heart pacemaker picks up cellular telephone calls!

There are three essential elements to any EMC problem:

1. Source of an electromagnetic phenomenon.
2. Receptor (or victim) that cannot function properly due to the electromagnetic phenomenon.
3. Path between them that allows the source to interfere with the receptor.

Each of these three elements must be present, although they may not be readily identified in every situation. Identifying at least two of these elements and eliminating (or attenuating) one of them generally solves electromagnetic compatibility (EMC) problems.

For example, suppose it was determined that radiated emissions from a mobile telephone were inducing currents on a cable that was connected to an ECU controlling anti-lock brakes. If this adversely affected the operation of the circuit a possible coupling path could be identified.

Shielding, filtering, or re-routing of the cable may be the answer. If necessary, filtering or redesigning the circuit would be further possible methods of attenuating the coupling path to the point where the problem is non-existent.

Potential sources of electromagnetic compatibility problems include radio transmitters, power lines, electronic circuits, lightning, lamp dimmers, electric motors, arc welders, solar flares and just about anything that utilizes or creates electromagnetic energy. On a vehicle, the alternator and ignition system are the worst offenders. Potential receptors include radio receivers, electronic circuits, appliances, people, and just about anything that utilizes or can detect electromagnetic energy.

Methods of coupling electromagnetic energy from a source to a receptor fall into one of the following categories:

1. Conducted (electric current).
2. Inductively coupled (magnetic field).
3. Capacitively coupled (electric field).
4. Radiated (electromagnetic field).

Coupling paths often utilize a complex combination of these methods making the path difficult to identify even when the source and receptor are known. There may be multiple coupling paths and steps taken to attenuate one path may enhance another. EMC therefore is a serious issue for the vehicle designer.

4.8 Central electrical control

4.8.1 Overview

4.8.1.1 Introduction

For many years the trend with automotive electrical systems has been towards some sort of networked central control. This makes sense because many systems can share one source of information and, with the proper equipment, diagnostics can be made easier. Also, centralization allows facilities to be linked and improved. For example, networking and centralization of control units makes it easier to have a system where the engine will not start if a door is open, or selection of reverse gear can operate rear wiper when the fronts are switched on (Figure 4.62).

The basic central control system can be simplified in a way that is represented by Figure 4.54.

The most common usage of central control is for body systems such as lighting, wipers, doors, seats and windows. In some cases these systems are controlled by slave units via a communication network, in other cases, one unit controls everything. In almost all cases, this central unit is networked to other ECUs

Key fact

Centralization systems allow facilities to be linked and improved.

Key fact

The most common usage of central control is for body systems such as lighting, wipers, doors, seats and windows.

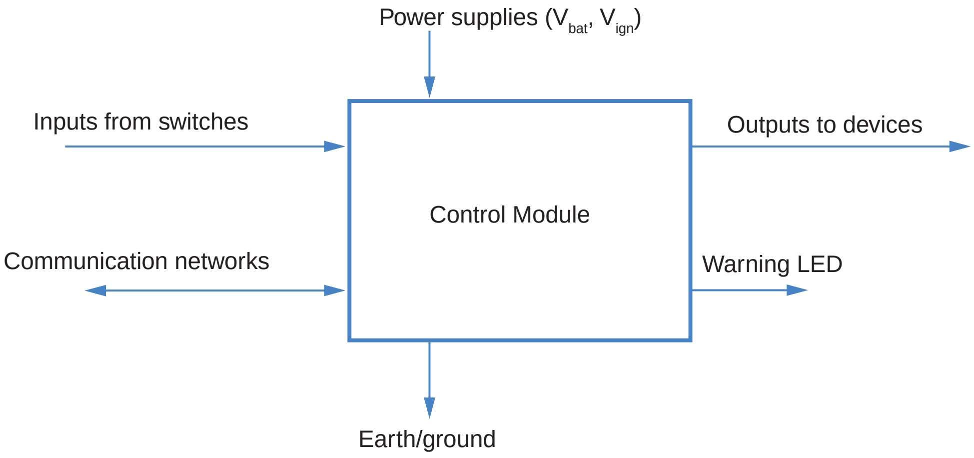


Figure 4.54 Central control system (simplified)

Some central control modules connect via normal wires to switches that supply normal voltage on/off signals, others use switches that communicate on the CAN or LIN networks. The outputs from the module are sent via relays or solid state switches on standard wires.

Manufacturers have different names for these systems and the control units but most have a similar function. Four example names follow:

- Body control module (BCM).
- General electronic module (GEM).
- Central control unit (CCU).
- Central control module (CCM).

4.8.1.2 Detailed system and circuit diagrams

Figures 4.55 and 4.56 show a full circuit diagram (in two halves) that has been adapted from materials supplied by Ford Motor Company. The circuit shows a general electronic module (GEM) and how it is used to control the wipers (in this case). Note that the multifunction wiper switch contains a series of switch contacts that all connect directly back to the GEM. Also note the CAN connection to the module.

Figure 4.57 below is from a Ford vehicle with adaptive front lighting. The light switch in this case has a supply (30) and an earth/ground (31) connection, but that all commands to operate the lights are sent via the LIN bus. The GEM supplies outputs to operate the lights and a separate module is used in this case for the adaptive features of the lights.

4.8.1.3 Control units

A central body control module (BCM) is the primary hub that maintains body functions, such as:

- internal and external lighting;
- security and access control;
- comfort features for doors and seats;
- other convenience controls.

A leading OEM company called FreeScale produces high quality 16-bit MCU families that target many BCM applications. A single board computer (SBC) combines voltage regulation with a CAN or LIN physical interface in a single package. H-bridge drivers and a series of high-side switches drive high-current loads and replace relays.

The gateway serves as the information bridge between various in-car communication networks, including Ethernet, FlexRay, CAN, LIN and MOST protocols. It also serves as the car's central diagnostic interface.

Exterior lighting plays an important role in the safety of car passengers and other road users. Different types of lamps (e.g. halogen, xenon or LED) are used in a variety of lighting functions, such as brake lights, turn indicators, low and high beam headlights, daytime running lights and others. More advanced functions include light bending, levelling and shaping to adapt to changing driving conditions.

FreeScale produces components as 8-, 16- and 32-bit devices to address the processing requirements of these different applications. Their 'eXtreme' switch product family of intelligent high-side switches use performance profiles tailored for different lamp types. They feature extensive diagnostic functionalities to detect faults and malfunctions and provide 'wave-shaping' to improve system-level EMC performance.

Figure 4.59 shows the internal configuration of a lighting control module that uses solid-state switching.

An example of the 'eXtreme' switch (high side switch) family of devices is shown as Figure 4.60. It is designed for low-voltage automotive lighting applications. Its four low RDS(ON) MOSFETs (dual 10 mΩ/dual 12 mΩ) can control four separate 55W/28W bulbs, and/or Xenon modules, and/or LEDs.

Programming, control and diagnostics are accomplished using a 16-bit SPI interface. Its output with selectable slew rate improves electromagnetic compatibility (EMC) behaviour. Additionally, each output has its own parallel input or SPI control for pulse-width modulation (PWM) control if desired. The 10XS3412 allows the user to program via the SPI the fault current trip levels and duration of acceptable lamp inrush. The device (Figure 4.58) has fail-safe mode to provide fail-safe functionality of the outputs in case of MCU damaged. The internal circuit of this device is shown as Figure 4.61.

Key fact

H-bridge drivers and high-side switches drive high-current loads and replace relays.

4.8.1.4 Summary

The previous sections outlined and showed examples of how central control systems are configured. At first view they can appear complex but actually compared to separate switched and wires and relays for every electrical component on the car, centralization actually simplifies the system as well as making it easy to add new features.

4.8.2 Ford generic electronic module (GEM)

4.8.2.1 Overview

On many Ford cars, the GEM is installed under the instrument panel, behind the glove compartment. It controls a multitude of functions in the generic electronics. The GEM is a separate module and does not contain any current distribution section (no fuses or relays).

Depending on equipment level, different GEMs are installed in the factory. One of the highest equipment versions supports the following:

- central locking;
- opening/closing function via radio remote control (radio receiver built into the GEM);

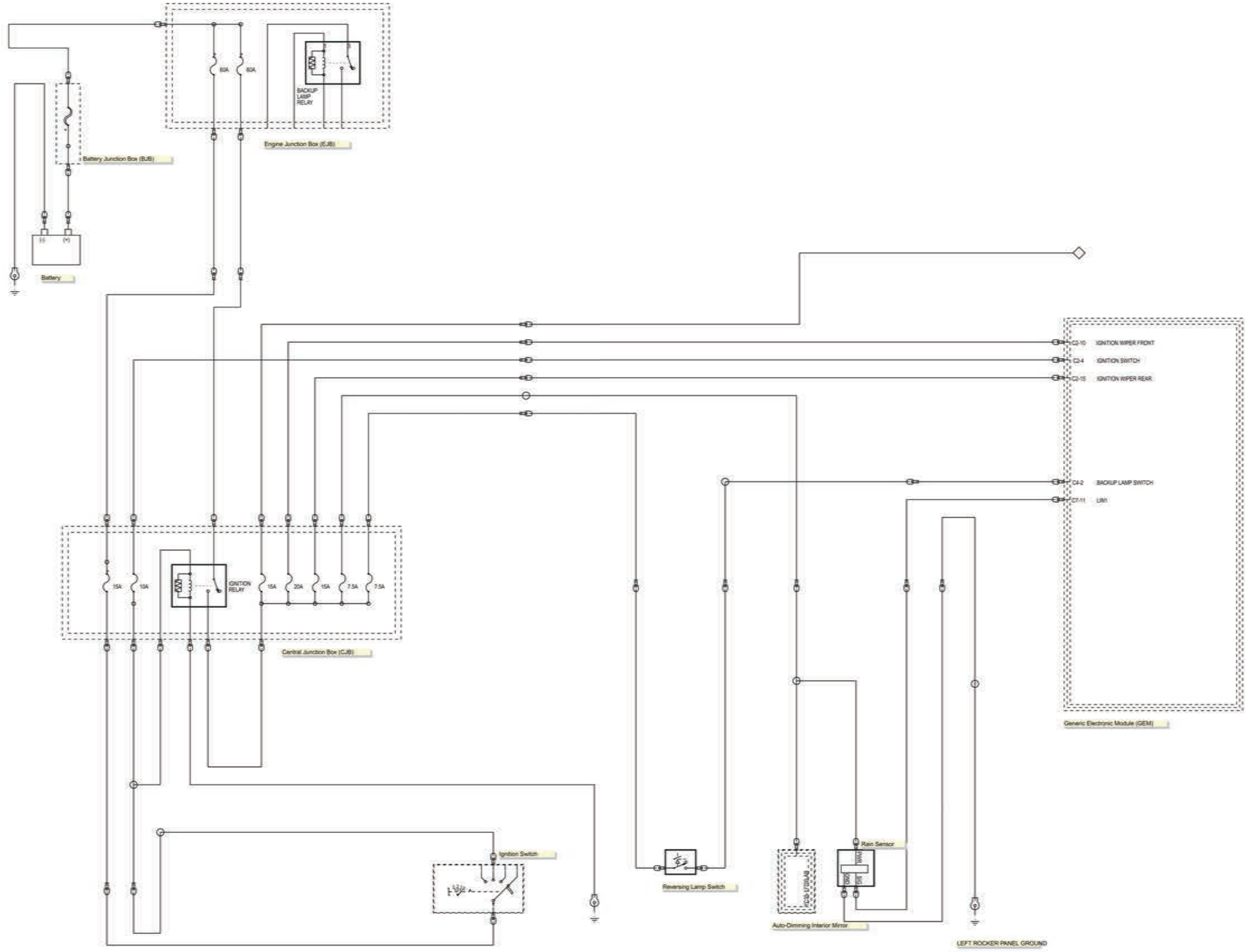


Figure 4.55 Part 1 of a wiper circuit using a central control module (Source: Ford Motor Company)

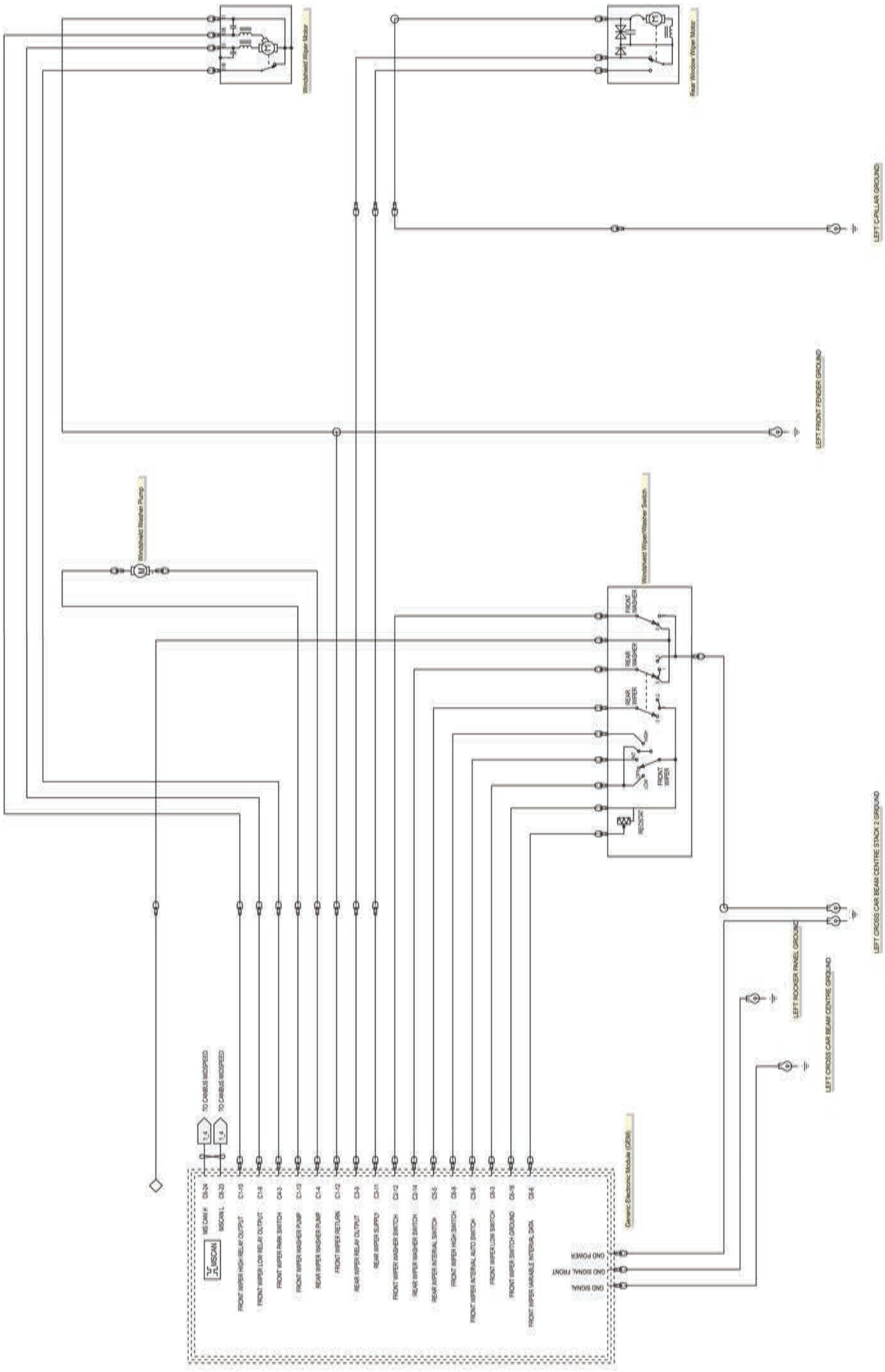


Figure 4.56 Part 2 of a wiper circuit using a central control module (Source: Ford Motor Company)

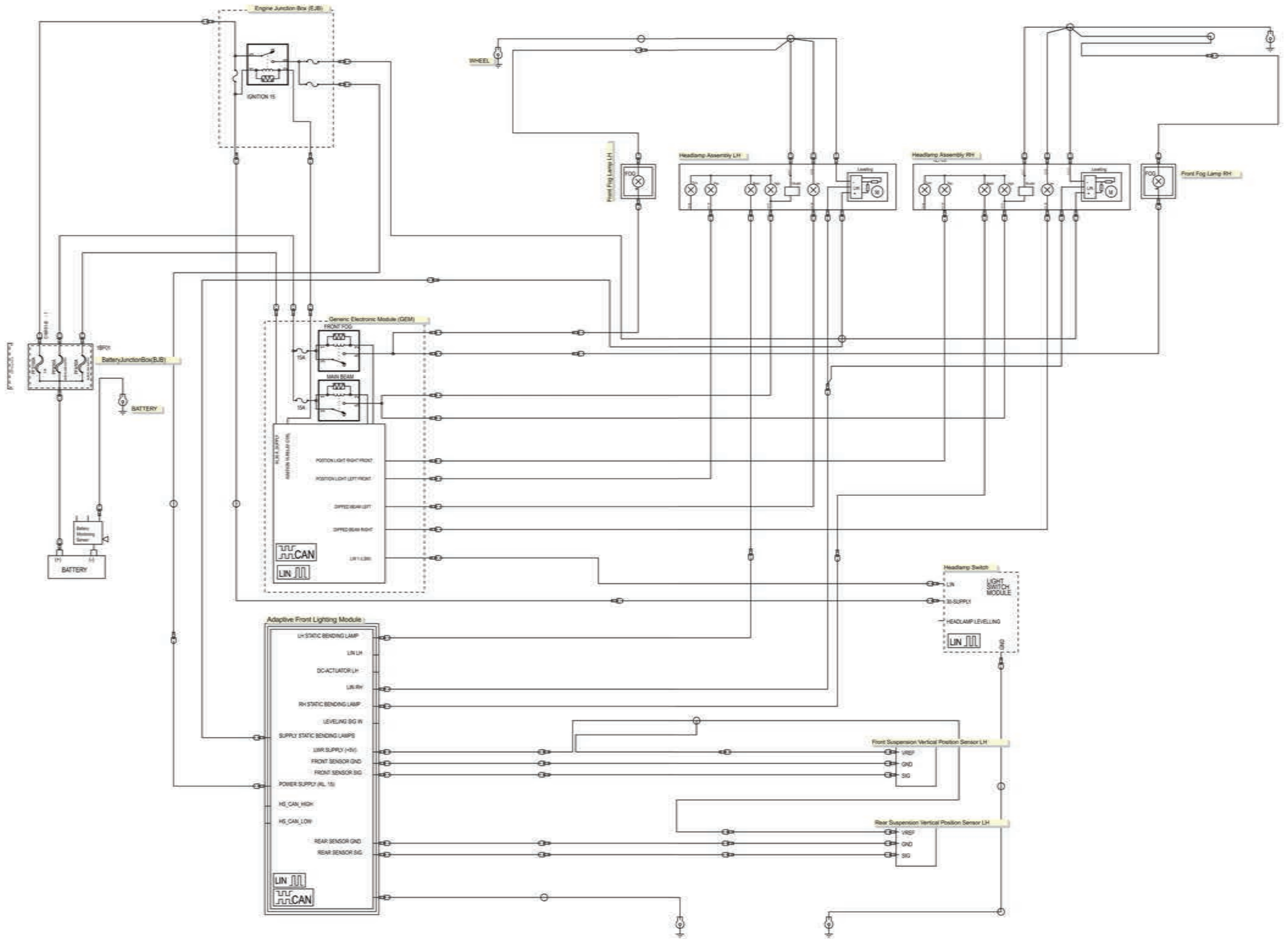


Figure 4.57 Lighting circuit using a general electronic module (Source: Ford Motor Company)

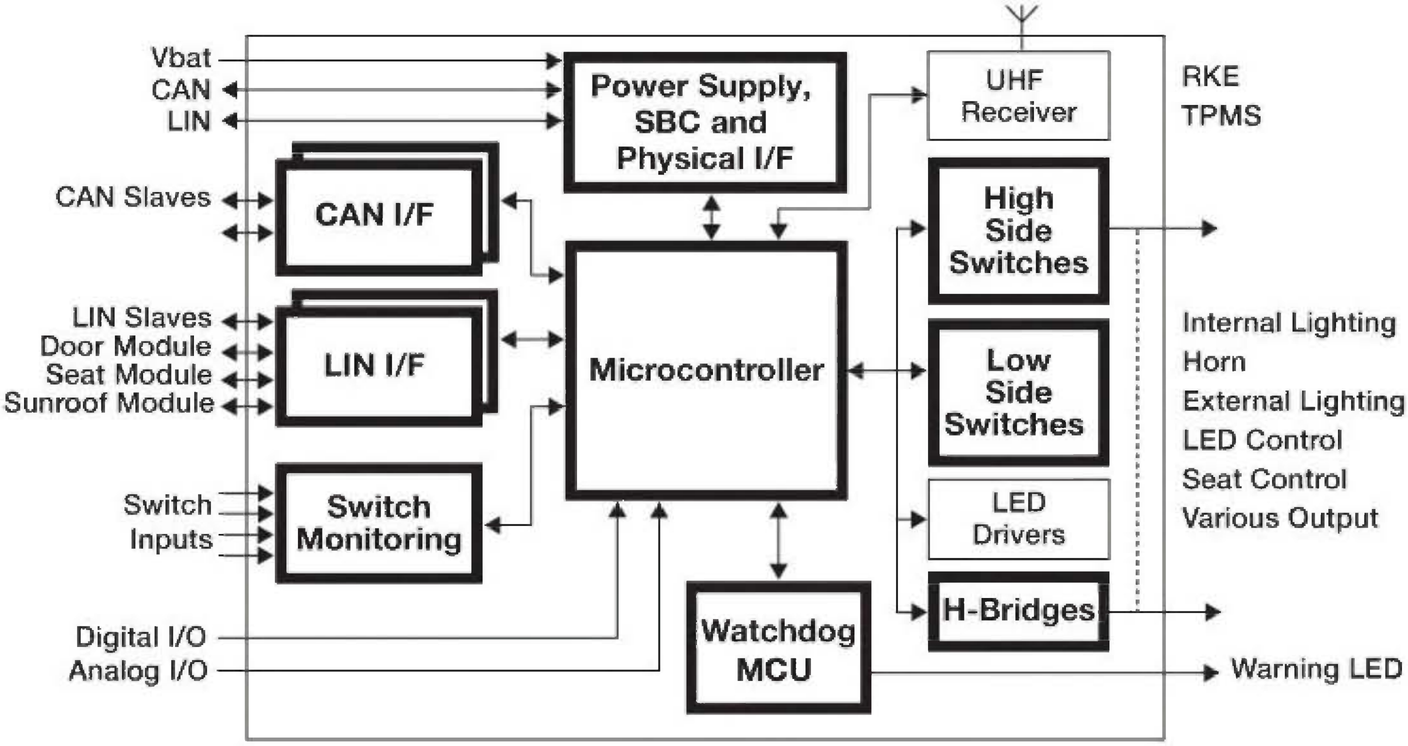


Figure 4.58 Details of the body control module and central gateway (Source: FreeScale Electronics)

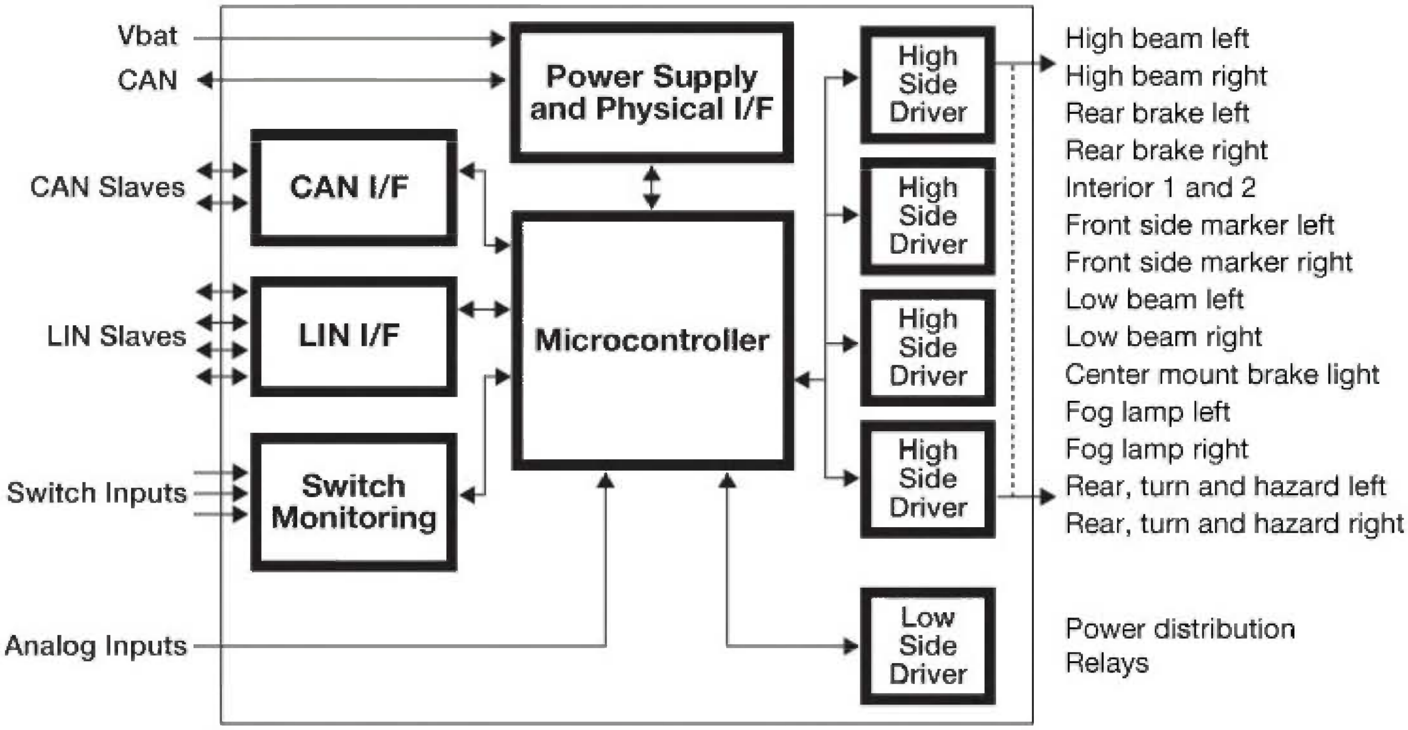


Figure 4.59 Lighting control unit (Source: FreeScale Electronics)

Figure 4.60 eXtreme or high side switch devices (Source: FreeScale)

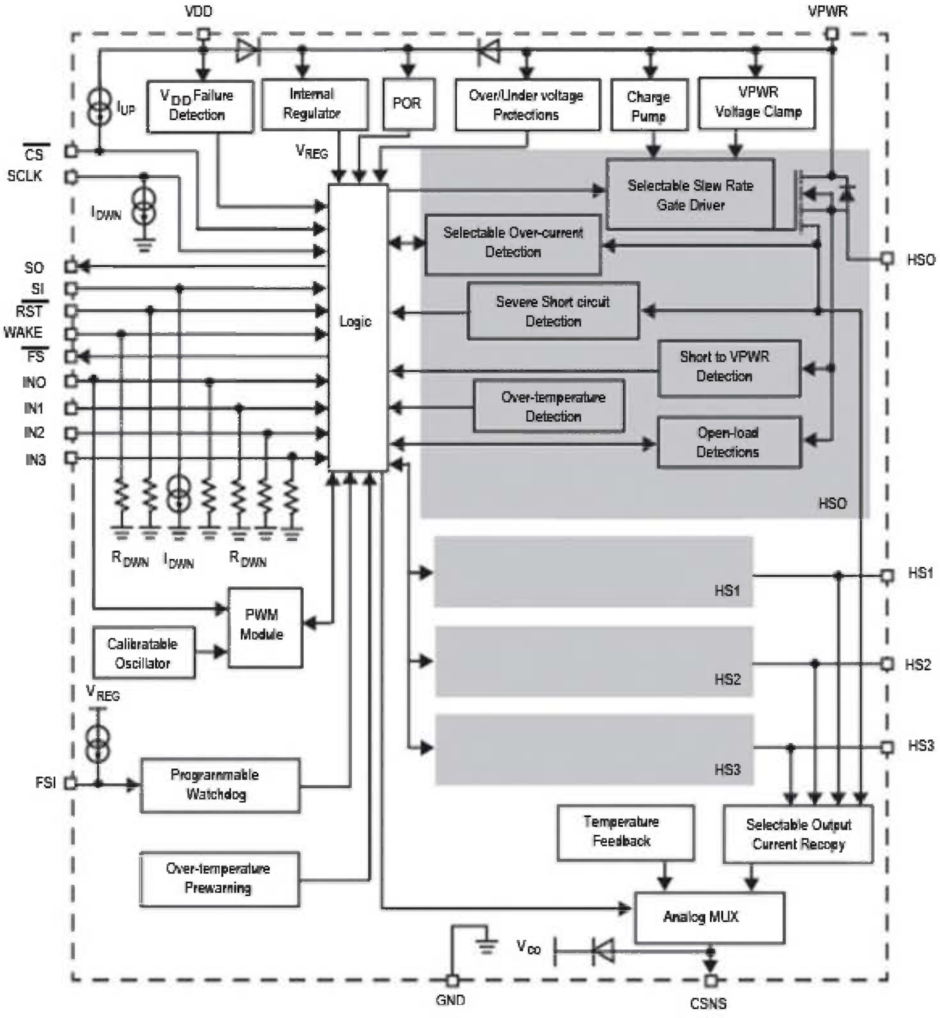


Figure 4.61 Internal configuration of an eXtreme or high side switch (Source: FreeScale)

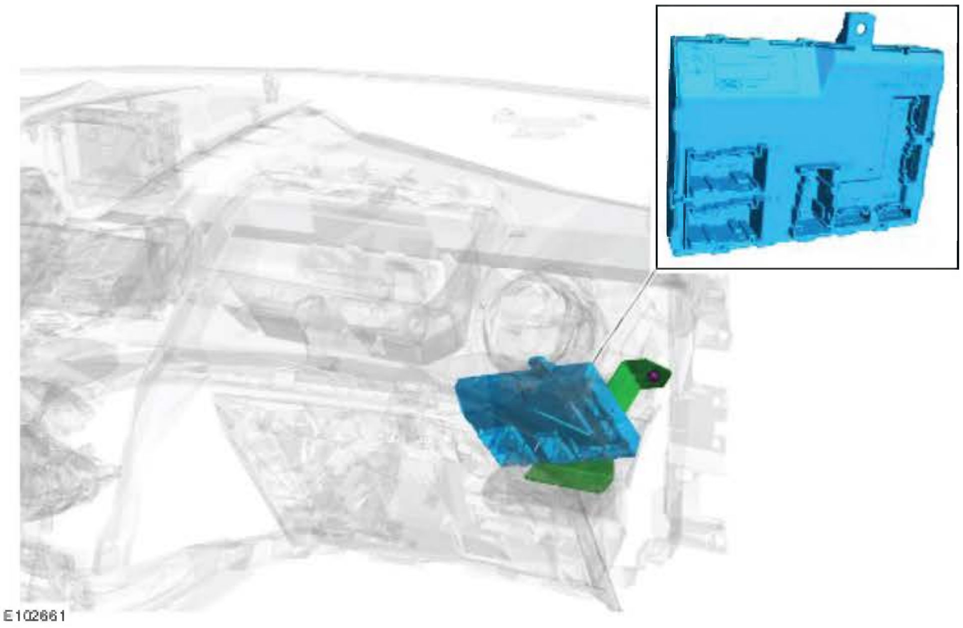


Figure 4.62 Generic electronic module location (Source: Ford Motor company)

- fold-in/fold-out external mirrors;
- ambient lighting;
- automatic light and wiper control;
- anti-theft warning system (perimeter monitoring);
- double locking.

An emergency running mode is also available. If a serious fault occurs in the GEM (a defective microprocessor or failure of the voltage supply for example) the following functions are still maintained:

- dipped beam (will then be switched on every time the ignition is switched on);
- windscreen wipers (only slow speed).

One many vehicles, if the GEM has been changed, the new one will configure itself automatically when the ignition is switched on, after about eight seconds.

4.8.2.2 Service mode

Various input and output signals can be checked using the service mode. Service mode is activated as follows:

1. turn the ignition off;
2. press the heated rear window switch and keep it pressed;
3. switch the ignition on and then release the heated rear window switch.

The GEM now requests the user to perform a set sequence of different functions (for instance, to operate the warning flashers, the light switch, and the door locking). If the test is completed successfully, a signal tone will be output.

The GEM can contain configuration data about the vehicle systems as well as the VIN. This can be backed up before replacement using manufacturer's equipment, the integrated diagnostic system (IDS) to transfer it to the new GEM. In many cases, the system will operate without the back-up in the GEM. However, if the instrument cluster fails, there will no longer be any module configuration data available.

Key fact

The GEM can contain configuration data about the vehicle systems as well as the VIN.

4.8.2.3 Anti-theft

The anti-theft protection of the vehicle is a feature of the GEM. The following perimeter-monitoring components are used:

- door ajar switch;
- engine bonnet switch;
- tailgate switch.

The alarm state of the vehicle is signalled by the turn signal lamps and a horn with its own battery. With the ignition switched off, the anti-theft system is activated about eleven seconds after the vehicle is locked. If the bonnet, tailgate or one of the doors is not fully closed, it can be opened without triggering the alarm. In this case the system is not armed.

4.8.3 Communication between modules

4.8.3.1 Introduction

In a communications network (using a data bus), various modules of different systems are connected to one another via one or several lines. The purpose of the data bus system is the transmission of data between the connected modules themselves, as well as between the modules and a diagnostic tool.

In a data bus system, complete data blocks are transmitted instead of single on/off pulses. In addition to the actual information, these blocks also contain data regarding the address of the module to be controlled or accessed, the size of the data block and information for monitoring the content of each individual data block. Please see the multiplexing section for more information.

Data bus systems offer various advantages:

- simplified data transmission between the modules due to a standardised protocol;
- fewer sensors and connectors;
- improved diagnostics;
- lower costs.

Diagnostic equipment is connected to the various bus systems and to the power supply via the standard 16-pin Data Link Connector (DLC). Signals for module programming or backup are also transmitted via the DLC. In order to be able to establish communication with one another, the modules of the individual systems must use the same language. This language is called a protocol.

Many manufacturers such as Ford use three different data bus systems. Depending upon model and equipment level, all three data bus systems may be used. Each of these data bus systems has its own protocol.

4.8.3.2 Data bus systems

Standard Corporate Protocol (SCP) bus

This system consists of two twisted wires. It is used for communication between the powertrain control module (PCM) and the diagnostic equipment connected via the DLC. Depending upon engine version and year of manufacture, a third wire (ACP bus) is used for programming the PCM. This bus is only used in conjunction with the SCP bus.

International Organisation for Standardisation ISO 9141 bus

This consists of a single wire and is used exclusively for communication between the modules and a diagnostic tool. The fault memories of the various modules are read out via the ISO 9141 bus.

Controller Area Network (CAN) bus

This consists of two twisted wires and operates serially (data is transmitted sequentially). It is used for communication between the modules themselves and between the modules and diagnostic equipment. The modules are connected to the data bus in parallel. New modules can be incorporated easily, without modifying the other wiring or modules. The transmitted data is received by every module connected to the CAN bus. As each data packet has an identifier, in which the priority of the message is determined as well as the content identification, each module can detect whether or not the data is relevant for its own information processing. This enables several modules to be addressed with a particular data packet and supplied with data simultaneously. For this purpose, it is ensured that important data (for example from the Anti-lock Brake System (ABS)) is transmitted first. The other modules are only able to submit their data to the data bus after the high-priority messages have been received.

Key fact

CAN connections consist of two twisted wires and operates serially (data is transmitted sequentially).

The advantages of the CAN bus are:

- minimization of wiring requirements;
- high degree of error protection (fault/fail-proof);
- robustness;
- good extendibility;
- prioritization of messages;

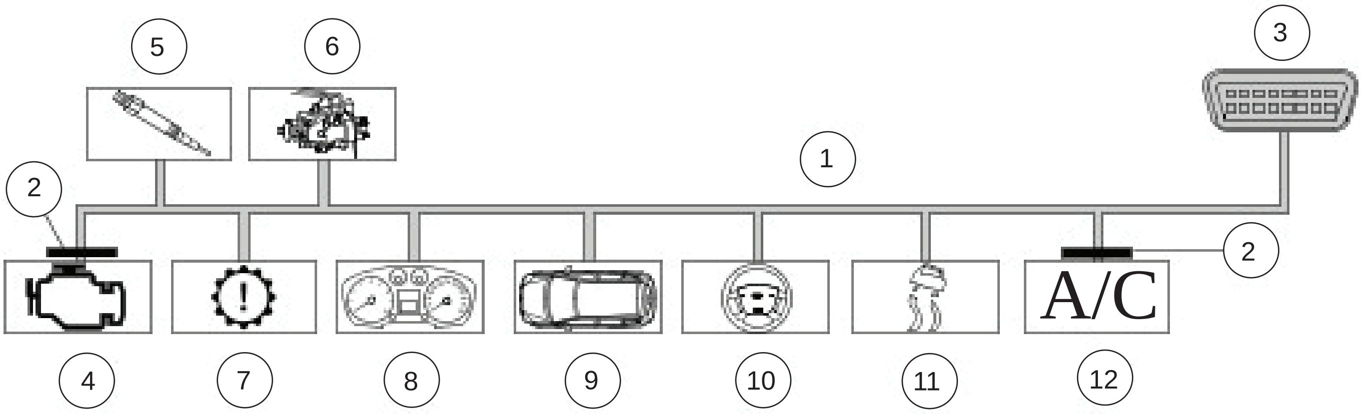


Figure 4.63 1, High-speed CAN bus; 2, Terminating resistors; 3, DLC; 4, PCM; 5, Injector Driver Module (IDM); 6, Fuel injection pump; 7, TCM; 8, Instrument cluster; 9, Yaw rate sensor; 10, Steering wheel rotation sensor; 11, Electronic stability program module; 12, EATC module or ATC module (Source: Ford Motor Company)

- inexpensive;
- automatic repetition of faulty messages;
- independent system monitoring and option for automatic disconnection of faulty modules from the data bus.

Over the years the complexity of the networks has increase and a number of methods are used to combine different protocols. The device that allows on protocol to communicate with another is known as a gateway.

4.8.3.3 Single network system

The modules connected to an ISO 9141 bus depend upon the equipment level of the vehicle. The ISO 9141 bus is used solely for reading out the fault memories of the modules connected to the data bus. The ISO 9141 bus connects the various modules to diagnostic equipment via the DLC.

Generally, vehicles built prior to 2003 have only one CAN bus. This is designated as high-speed CAN bus (HS CAN). The modules exchange data with one another via this bus. The modules connected to the HS CAN bus depend upon the equipment level of the vehicle.

One 120 ohm terminating resistor is installed in the PCM and in the instrument cluster respectively. These terminating resistors are used for suppression of the data bus system. In order to be able to ensure correct functioning of the data bus system, the modules must always be connected with an integral terminating resistor.

4.8.3.4 Twin network system

Due to the increased number of modules and the resulting ever-increasing data transmission, a second CAN bus (mid-speed CAN bus (MS CAN)) is used in vehicles built from around 2003 (but this does vary) (Figure 4.64). It operates at a lower speed and is mainly used for communication relating to convenience electronics.

A gateway (portal) is used in order to enable data exchange between the HS CAN bus and the MS CAN bus. The gateway serves as interface between the two CAN data bus systems and is installed in the air-conditioning module (in this example).

4.8.3.5 Multiple network system

An ISO bus (C in Figure 4.65) is used to connect to the keyless entry radio receiver (Figure 4.65).

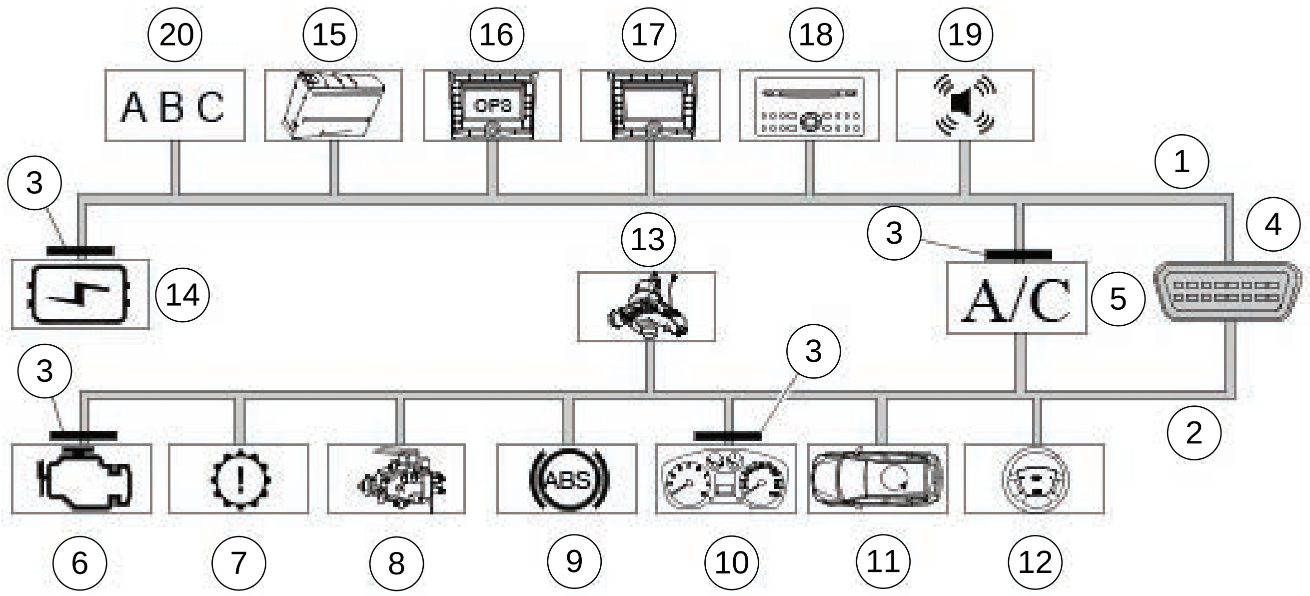


Figure 4.64 CAN bus communication: 1, Mid-speed CAN bus; 2, High-speed CAN bus; 3, Terminating resistors; 4, DLC5, EATC module or ATC module; 6, PCM; 7, TCM; 8, Fuel injection pump; 9, ABS module or ESP module; 10, Instrument cluster; 11, Yaw rate sensor; 12, Steering wheel rotation sensor; 13, Electrical actuator, turbocharger guide vane adjustment; 14, GEM; 15, Compact Disc (CD) changer; 16, Navigation module; 17, Touchscreen; 18, Audio unit; 19, Audio system module; 20, Bluetooth voice control module (Source: Ford Motor Company)

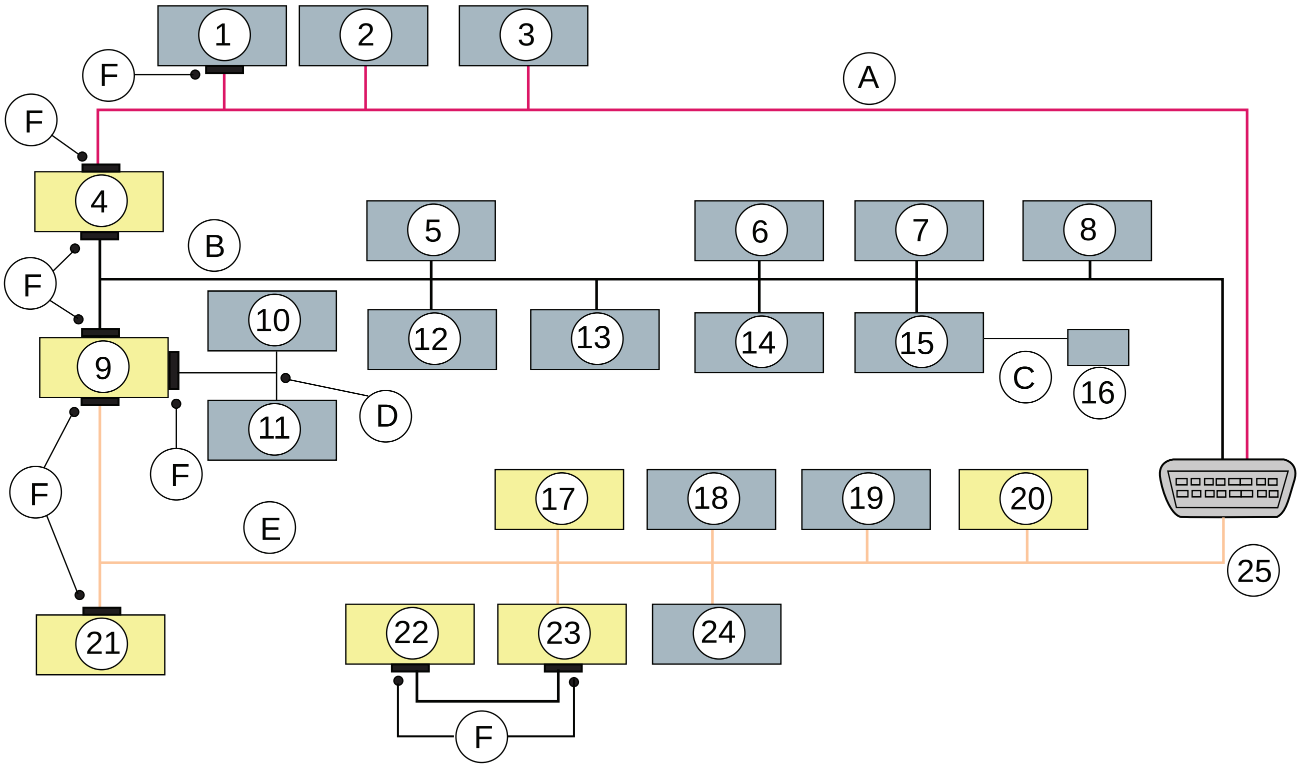


Figure 4.65 Networks and communication diagram: A; HS-CAN bus (multimedia system); B, MS-CAN bus (generic electronics); C, ISO bus; D, private MS-CAN bus; E, HS-CAN bus; F, terminating resistor; 1, audio unit; 2, multi-function display; 3, hands-free phone kit/Bluetooth®/ voice control module; 4, instrument Cluster; 5, PDM (passenger door module); 6, EATC (electronic automatic temperature control) module; 7, fuel fired booster heater; 8, electric liftgate module; 9, GEM (generic electronic module); 10, right-hand module, blind spot monitoring; 11, left-hand module, blind spot monitoring; 12, DDM (driver door module); 13, trailer Module; 14, reversing camera module; 15, keyless vehicle module; 16, radio receiver, keyless vehicle system; 17, EPS (electric power steering) module; 18, gas discharge headlamp module; 19, parking aid module; 20, steering wheel rotation sensor; 21, PCM (powertrain control module); 22, ABS (anti-lock brake system) module; 23, RCM (restraints control module); 24, TCM (transmission control module); 25, DLC (data link connector) (Source: Ford Motor Company)

Two separate HS-CAN networks are used on this system:

HS-CAN bus (E in Figure 4.65):

- EPS module
- Gas discharge headlamp module
- Parking Aid Module
- Steering wheel rotation sensor
- PCM
- ABS module
- RCM
- TCM
- GEM.

HS-CAN multimedia system (A in Figure 4.65):

- Instrument Cluster
- Audio unit
- Multi-function display
- Hands-free phone kit/Bluetooth®/voice control module.

The two terminating resistors, 120 ohms each, of the HS-CAN data bus are integrated into the PCM and the GEM. Since both 120 ohm resistors are connected in parallel, the total resistance is 60 ohms (measured at the DLC between pin 6 and 14). The HS-CAN bus and the HS-CAN multimedia system have a speed of 500 kBit/s.

Two separate MS-CAN networks are used:

MS-CAN generic electronics (B in Figure 4.65):

- PDM
- EATC module
- Fuel fired booster heater
- Liftgate release module
- Instrument Cluster
- GEM
- DDM
- Trailer Module
- Reversing camera module
- keyless vehicle module.

Private MS-CAN (D in Figure 4.65):

- right-hand module, blind spot monitoring;
- left-hand module, blind spot monitoring.

On some cars a LIN data bus system is also used. LIN is a standard especially for the cost-effective communication between intelligent sensors and actuators in motor vehicles. LIN is used everywhere that the bandwidth and versatility of CAN is not needed. The transmission speed within the LIN databus system is 9.6 kBit/s. A LIN databus system consists of a LIN master and one or more LIN slaves. The LIN is a single conductor databus system, i.e. the data is transmitted in the cable on one wire. No termination resistors are used in the LIN network.

The following systems are often connected to a LIN bus:

- Rear PDM
- Rear DDM

- Front driver's side switch unit
- Steering wheel module
- Light switch unit
- Interior scanning sensors
- Anti-theft alarm system horn with own battery
- Battery monitoring sensor
- Rain sensor
- Electronic steering lock unit
- Audio system control unit
- Select-shift switch module
- generator
- Gas discharge headlamps.

4.8.4 Summary

Central electrical and electronic control has evolved to become a system that may be better described as distributed control. However, because units such as the body control module have wide ranging functions it is often described as central control. A key point however is that almost all modules are now connected using one or more networking protocol.

4.9 Connected cars

4.9.1 Introduction

There are a number of features and systems now in use and under development where, one could argue, the car is becoming more intelligent – a scary thought! However, there are also a number of developments whereby different technologies are merging or, to be more accurate, converging (Figure 4.66).

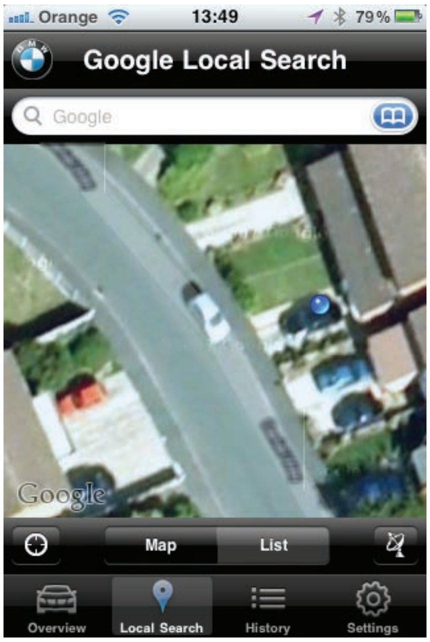


Figure 4.66 This iPhone BMW app allows the user to remotely lock or unlock the car and other features such as route planning and then sending the destination to the vehicle.

Right at the start of this book I speculated on the 'Future Car'. Actually what is most interesting is that most of that speculation has come true, in a remarkably short time. We may not yet be able to sit in our cars and let them do all the driving, but the technology exists to make it happen.

In the following sections I have outlined some interesting developments ranging from linking a smartphone with a car, to vision enhancement and Wi-Fi.

4.9.2 Smart cars and traffic systems

There are a number of examples of technologies that allow cars to drive themselves to a greater or lesser extent. The first example outlined here is a concept vehicle that will come when called and can park itself. The second example is a technology that allows cars to follow each other closely at high speed; a system known as platooning.

Definition

Platooning: a road train of wirelessly linked cars following a lead vehicle.

4.9.2.1 General Motors EN-V

General Motors have developed, as a concept car, an electric networked vehicle called EN-V (get it?). The EN-V represents one vision of the future of urban personal mobility. The car includes features that allow the vehicle to park itself and automatically return to the user when summoned from a smartphone application.

Figure 4.67 General Motors electric networked vehicle (EV-V) concept car (Source: GM Media)

The vehicle can rotate 360 degrees and be driven in manual mode with a driver – or without. It runs on battery power for about 25 miles on a charge, with top speeds of 25 miles-per-hour. A significant downside to this pint-sized vehicle is its inability to withstand a collision with a full-sized vehicle. However, cities could perhaps create EN-V-only lanes, or alternatively, create enclosed areas solely for use by this type of vehicle.

Several EN-Vs can fit into a standard parking space and the interior has room for two passengers. Using sophisticated sensing technology, EN-V can detect obstacles in its path, including pedestrians or other vehicles, and automatically come to a stop if necessary.

This concept is really interesting because while it would not be suitable for our current road systems, it is a new way of thinking about transport. A small step in this direction has already been made. In London (and many other cities) it is now possible to hire a bicycle from roadside areas using a credit or debit card. The bike can be returned to any other ‘station’. In some busy areas there are also dedicated lanes on the roads where they can be used.

4.9.2.2 Road trains

An EU project called SARTRE has been launched to develop and test technology for vehicles that can drive themselves in long road trains on motorways. This technology has the potential to improve traffic flow and journey times, offer greater comfort to drivers, reduce accidents, and improve fuel consumption and hence lower CO_2 emissions.

The automotive industry has long been focused on the development of active safety systems that operate preventively, traction control and braking assistance programs for example. But automakers have now gone much further in proposing a technology that allows vehicles to be operated without any input whatsoever from the driver. Known as autonomous driving, this technology means that the vehicles is able to take control over acceleration, braking and steering, and can be used as part of a road train of similarly controlled vehicles.

Definition

The SARTRE project stands for Safe Road Trains for the Environment. Part-funded by the European Commission, SARTRE is led by Ricardo UK Ltd and with collaboration between: Idiada and Robotiker-Tecnalia of Spain, Institut für Kraftfahrwesen Aachen (IKA) of Germany, and SP Technical Research Institute, Volvo Car Corporation and Volvo Technology of Sweden.

Figure 4.68 One car following the lead vehicle (Source: Volvo Media)

The vehicles will be equipped with a navigation system and a transmitter/receiver unit that communicates with a lead vehicle. Since the system is built into the cars, there is no need for any infrastructure change to the road network.

The idea is that each road train or platoon will have a lead vehicle that drives exactly as normal, with full control of all the various functions. This lead vehicle is driven by an experienced driver who is thoroughly familiar with the route. For instance, the lead may be taken by a taxi, a bus or a truck. Each such platoon or road train will consist of six to eight vehicles.

As a driver approaches their destination, they take over control of the vehicle and leave the convoy by exiting off to the side. The other vehicles in the road train close the gap and continue on their way until the convoy splits up.

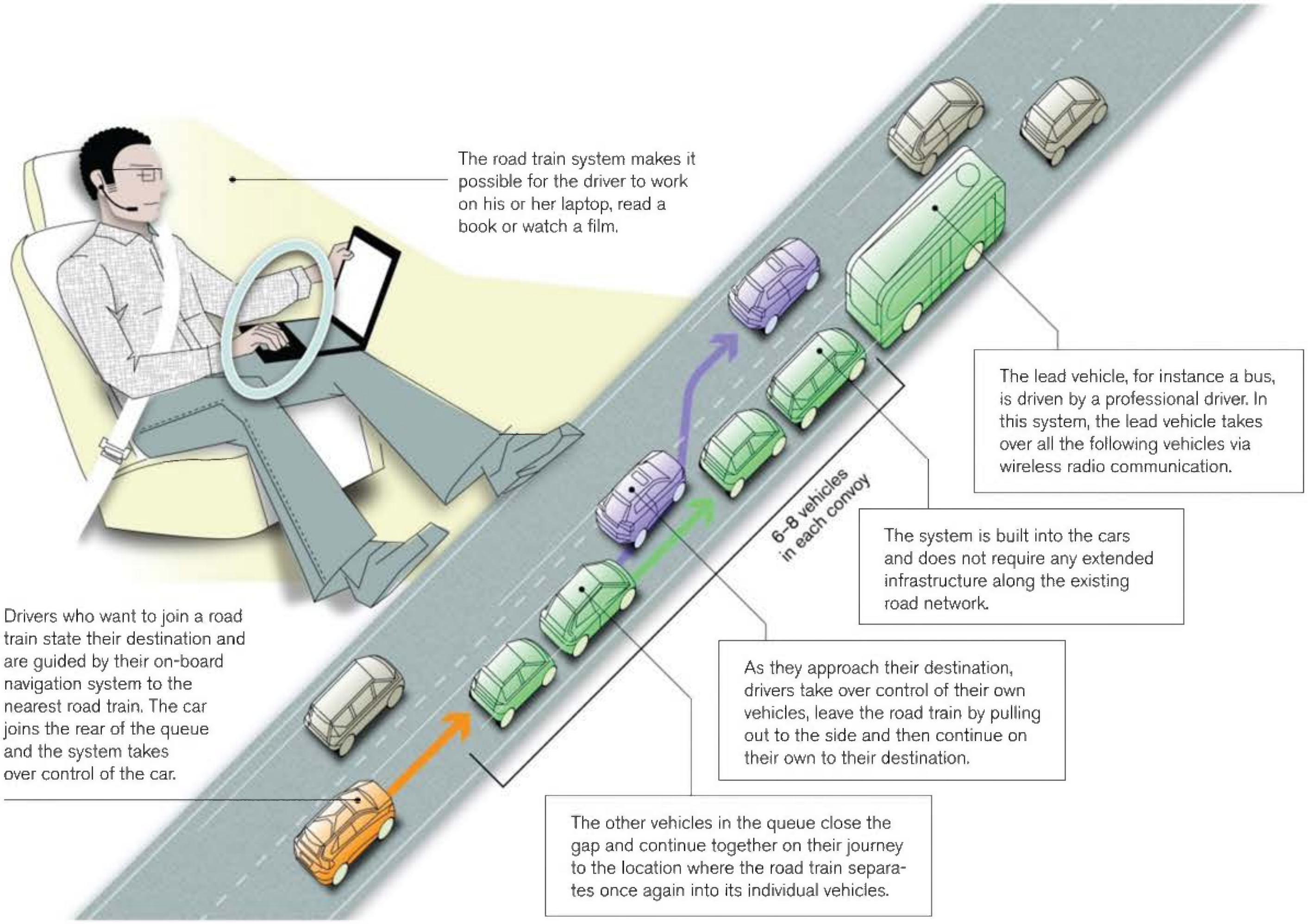
The advantage of these road trains is that all the other drivers in the convoy have time to get on with other business while on the road, for instance when driving to or from work. The road trains increase safety and reduce environmental impact thanks to lower fuel consumption compared with cars being driven individually. The reason for this is that the cars in the train are close to each other, exploiting the resultant lower air drag. The energy saving is expected to be in the region of 20 per cent. Road capacity will also be able to be utilised more efficiently.

As the participants meet, each vehicle's navigation system is used to join the convoy, where the autonomous driving program then takes over. As the road train approaches its final destination, the various participants can each disconnect from the convoy and continue to drive as usual to their individual destinations. Figure 4.69 outlines the process of joining and leaving a platoon.

The tests carried out at the time of writing (2011) included a lead vehicle and single following car. The steering wheel of the following car moves by itself as the vehicle smoothly follows the lead truck around a test track. The driver is able to drink coffee or read a paper, as no input is required to operate the vehicle.

The platooning technique is designed to achieve a number of things:

- Road safety, as it minimises the human factor that is the cause of at least 80% of accidents.
- Fuel consumption and CO2 emissions improve by up to 20%.
- Convenience for the driver as it frees up time for other activities.
- Traffic congestion will be reduced because the vehicles travel at highway speed with only a few meters gap.



The road train system makes it possible for the driver to work on his or her laptop, read a book or watch a film.

Drivers who want to join a road train state their destination and are guided by their on-board navigation system to the nearest road train. The car joins the rear of the queue and the system takes over control of the car.

The lead vehicle, for instance a bus, is driven by a professional driver. In this system, the lead vehicle takes over all the following vehicles via wireless radio communication.

The system is built into the cars and does not require any extended infrastructure along the existing road network.

As they approach their destination, drivers take over control of their own vehicles, leave the road train by pulling out to the side and then continue on their own to their destination.

The other vehicles in the queue close the gap and continue together on their journey to the location where the road train separates once again into its individual vehicles.

Figure 4.69 Road train methodology (Source: Volvo Media)

The technology development is well underway and may go into production in a few years. What may take longer is getting public acceptance and the legislation where 25 EU governments must pass similar laws.

4.9.3 Wi-Fi cars

Wi-Fi is becoming a feature that may soon be standard in many cars. Connectivity is achieved by plugging a 3G dongle or a SIM card into the car. This then creates a wireless hotspot and any suitable device can be connected. From this point, all the features of 'Internet control' become possible. For example Skype, connection to remote systems at home or connection to cameras in the car from home, as well as the obvious emails and web downloads. The roll out of 4G mobile broadband promises to make facilities like video calls to and from the car a possibility.

The main uses are likely to be connectivity for business 'road warriors' and entertainment for laptop of console-using passengers.

There is now and IEEE standard for in-car and associated Wi-Fi; this is 802.11p, which is dedicated to traffic applications. As well as the normal Wi-Fi features such as Internet connection, this system can allow cars, sensors, traffic signs and more, to become a giant mesh network. Clearly there are security

Definition

4G: In telecommunications, 4G is the fourth generation of cellular wireless standards and successor to the 3G and 2G families of standards. The specified requirements for 4G standards are peak speed of 100 Mbit/s for high mobility communication (such as from trains and cars) and 1 Gbit/s for low mobility communication (such as pedestrians and stationary users).

Figure 4.70 A road train test (Source: Volvo Media)

implications here but the hope is that the system will allow a car to broadcast its speed and position – and if it is heading for or involved in an accident. If this is broadcast from car to car it could prevent following cars from making the accident worse. Further, in communication with say, traffic lights, the car could be told that the light it is approaching will turn red in 5 seconds, so it can be ready to stop (or maybe the 'human element' would tell it to speed up – we will see!).

This Wi-Fi standard may also be used for the platooning system discussed earlier.

Definition

IEEE: The Institute of Electrical and Electronics Engineers or IEEE (read eye-triple-ee) is the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. IEEE and its members inspire a global community through IEEE's highly cited publications, conferences, technology standards, and professional and educational activities (Source: www.ieee.org).

4.9.4 Bluetooth

Bluetooth is a proprietary open wireless technology standard for exchanging data via radio transmissions over short distances from fixed and mobile devices. These connections are known as personal area networks (PANs) or piconets, and have a high level of security. Bluetooth can connect several devices at the same time, overcoming problems of synchronization.

Key fact

Bluetooth can connect several devices at the same time, overcoming problems of synchronization.

The radio technology used is called frequency-hopping spread spectrum, which chops up the data being sent and transmits chunks of it on up to 79 bands (1 MHz each; centred from 2402 to 2480 MHz) in the 2400–2483.5 MHz range (allowing for guard bands). This range is in the globally unlicensed Industrial, Scientific and Medical (ISM) 2.4 GHz short-range radio frequency band, which is the same frequency band used for Wi-Fi.

Definition

The word Bluetooth is an anglicised version of the Scandinavian Blåtand/Blåtann, the epithet of the tenth-century king Harald I of Denmark and parts of Norway. Harald united dissonant Danish tribes into a single kingdom. The idea is that Bluetooth does the same with communications protocols.

Bluetooth is a packet-based protocol with a master–slave structure. One master may communicate with up to 7 slaves in a piconet. Packet exchange is based on the basic clock, defined by the master, which ticks at 312.5 μs intervals. All devices use this clock. Two clock ticks make up a slot of 625 μs; two slots make up a slot pair of 1250 μs. In the simple case of single-slot packets the master transmits in even slots and receives in odd slots. The slave does the opposite. Packets may be 1, 3 or 5 slots long but in all cases the master transmission starts in even slots and the slave transmission in odd slots.

Key fact

Bluetooth is a packet-based protocol with a master–slave structure.

Bluetooth provides a secure way to connect and exchange information between devices such as mobile phones, hands-free headsets, telephones, laptops, personal computers, printers, Global Positioning System (GPS) receivers, digital cameras, and video game consoles.

Many cars now have Bluetooth connectivity or the option to have it included. Interestingly, in 2011, to add Bluetooth to a computer costs as much as a small round of drinks, the option from some carmakers is as much as a good laptop!

The main item that will connect to the car is a standard mobile phone, or more often no a smartphone so that applications on the device can be controlled from the car.

4.9.5 Applications (apps)

It has been possible to connect phones and MP3 players, for example, to cars for some time now. Music players can be connected using a simple 3.5 mm stereo jack-to-jack cable and the in car entertainment system then used to amplify the sound. Bluetooth connection of phones is also now becoming common. These basic connections do not allow sharing of data or applications. Smartphones are now in very common use and are the result of convergence of cameras, MP3 players, games consoles, GPS and, oh yes, mobile telephones.

Definition

Convergence: The tendency for different technological systems to evolve towards performing similar tasks or being contained in one device. The convergence of a camera and mobile phone for example, now allows sharing of resources and they interact with each other, creating new possibilities.

For many people, the smartphone is at the centre of their life and the apps it contains are what they have chosen. To connect with a car and use different applications is therefore confusing for the users. However, there are two ways in which manufacturers are working to deal with this:

1. Integrating the smartphone and the car.
2. Replace the car systems with the smartphone's.

The first is that many carmakers are now developing (or already have) systems where the smartphone integrates with the car in a way that the car becomes an extension for some of the apps on the phone. Ford, for example, have a system known as 'Sync AppLink' which allows the driver to control specific applications on an Android, iPhone or BlackBerry smartphone using the car's inbuilt voice recognition, controls on the steering wheel or controls on the dashboard. This would make it possible to say 'mobile apps, iPod, play' or something similar. The idea is that the smartphone is integrated with, and operated to some extent by the car.

The second method some manufacturers are investigating is to replace the in-car system with the smartphone. For example, a company known as Oxygen Audio have developed the 'O Dock', which is an iPhone docking station for the car. When docked into the O Dock, the user can access all of the iPhone's applications in both horizontal and vertical mode. Listen to internet radio, make a hands-free or Bluetooth call and play their entire iTunes music catalogue. In addition, they can access GPS, maps or any other useful applications. The iPhone also charges while docked.

I suspect the real answer will be a combination of the two. However, another good reason of why a user would prefer to use smartphone apps in conjunction with the car is cost. For example, a few years ago, it cost about as ten times as much as a standalone satnav (such as a TomTom) to have a factory fitted version in a car. Now, it is possible to buy a smartphone application for about a third of the price of the standalone device. And further, as long as you don't download data when roaming, many smartphones have a pretty good satnav app included that costs nothing.

An interesting idea is that the same app could direct the car to a car park and the same app when the smartphone is removed from the car, could direct you right to the door.

Figure 4.71 The O Dock for an iPhone (Source: Oxygen Audio)

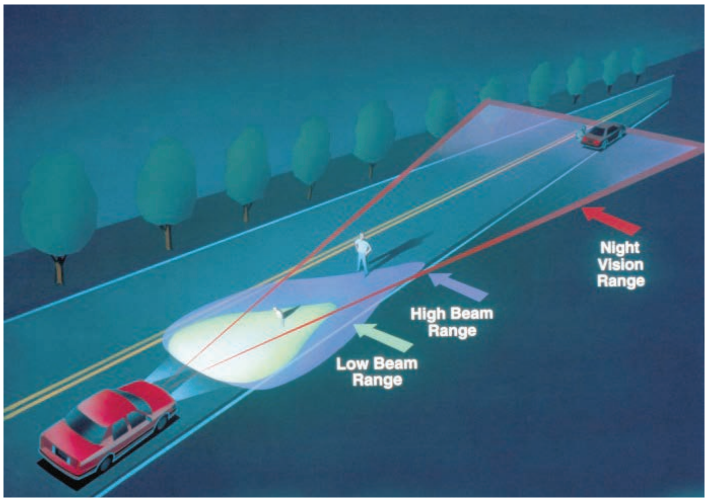


Figure 4.72 Vision enhancement using infra-red (Source: GM Media)

4.9.6 Vision enhancement

Cadillac developed a windscreen head-up display (HUD) technology, combined with night vision technology around the year 2000. This allows objects, such as animals (or humans), to be highlighted for the driver, preventing potential accidents.

In particular the range in which objects can be perceived is increased considerably. A view from the driver's perspective is shown in Figure 4.73.

Figure 4.74 shows a magnified view of the pyroelectric detector structure used for infra-red detection. The detector consists of a barium strontium titanate reticulate structure bonded to a readout integrated circuit. Each reticulated section corresponds to a single detector pixel. The pixels are on 48.5 μm centres and are less than 20 μm thick.

Researchers at General Motors R&D Labs in Warren, Michigan continue to investigate new ways to assist the vision challenges of the aging driver population. In Figure 4.75, an enhanced vision system (EVS), comprised of a transparent display technology and advanced sensors, is used to display information that enhances the external world. In this example, the system

Figure 4.73 Vision enhancement from the driver's perspective (Source: GM Media)

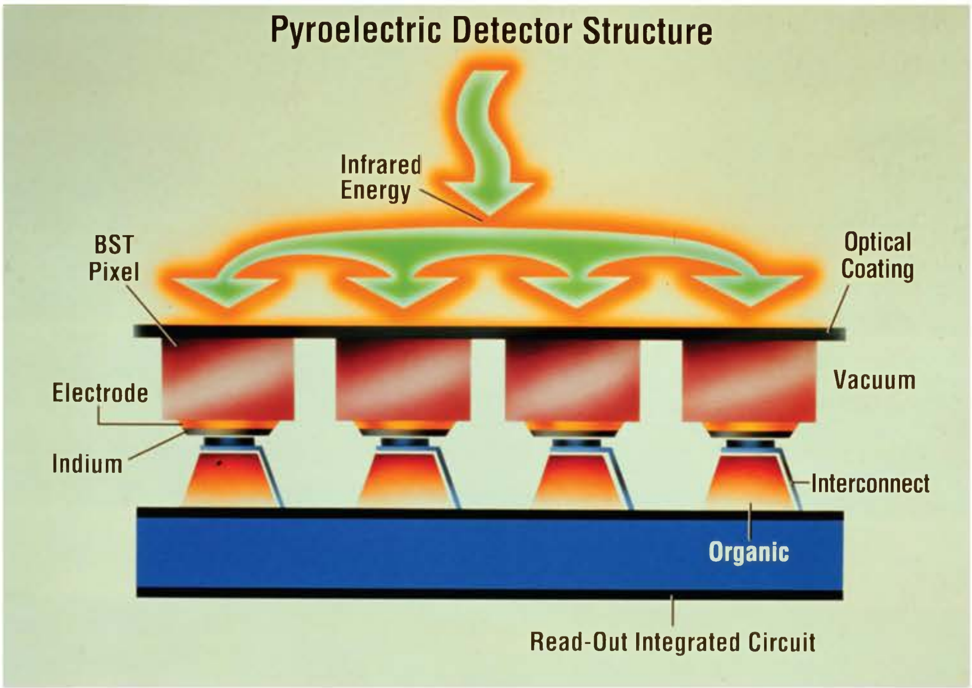


Figure 4.74 Pyroelectric detector structure (Source: GM)

highlights the road edge in a low visibility condition. Infra-red sensors and video cameras are the two main sensors.

Figure 4.76 shows the edge of the road enhanced in very poor visibility conditions. Still under development, this system has great potential to reduce accidents.

Key fact

Infra-red sensors and video cameras are two of the main sensors used for vision enhancement.

4.9.7 Self-help

On board diagnostic systems are now quite advanced but require a scanner or code reader to be connected to access the information. Perhaps the next

Figure 4.75 Enhancement of the road edges and a possible animal obstruction (Source: GM Media)

Figure 4.76 Enhancement of the road edge in foggy conditions (Source: GM Media)

development will be to integrate this function into the car so that any faults can be displayed to the driver. Or maybe the faults would be relayed to the nearest suitable garage, and the satnav programmed to take you there – before the car breaks down.

4.9.8 Big brother

As a somewhat scary thought, one system under development uses the GPS to set the maximum speed for the car so, if you are in a 40 mph area, the car will not exceed that limit. Further, researchers are looking at how to adapt the way the car can be driven in order to save fuel. For example, by only allowing a level of acceleration that is appropriate to the prevailing conditions, for where the car actually is at that time, economy and safety could be improved. I think I will take the train!

4.9.9 When computers go wrong

Key fact

Volvo is a well-respected company when it comes to vehicle safety.

Volvo is a well-respected company when it comes to vehicle safety and indeed is often at the forefront of new technologies geared to reducing accidents. However, in 2010 Volvo gathered the world's media to show off new safety features, and they went spectacularly wrong, twice.

In fairness, these are systems under development but when they were demonstrating the crash-avoidance system on an S60, the results were not what they had hoped for. A car was fired out of a testing tunnel towards the back of a stationary lorry. The car was supposed to recognise the impending collision, but a problem between the control system and the battery meant the car ploughed into the back of the lorry.

The next test that went wrong was a display of a pedestrian avoidance system also on an S60. The system uses radar and a camera to spot pedestrians and initiate an emergency stop. This worked fine for 9 out of 12 dummies – but the others all died. Again in fairness to Volvo, it would be unusual for a driverless car to encounter 12 careless dummies at the same time!

Some security experts suggest that as cars become ‘smarter’ and have more onboard computing power, they could be open to attack by anyone bearing a grudge or the usual sad characters who cause trouble just for their own bizarre reasons.

University of Washington researchers have hacked into several car systems using a variety of attack methods and said they could: *‘adversarially control a wide range of automotive functions and completely ignore driver input, including disabling the brakes or selectively braking individual wheels on demand’*. Oh dear!

4.9.10 Summary

As well as some speculation on the future, this section outlined some of the technologies that make cars more intelligent. Well, if not more intelligent, then more interactive. One of the key developments seems to be the understanding that many of us carry smartphones around, so it is better to integrate them with the car than to add another layer of technology. However, there is also the worry of being tracked and hacked! Your thoughts welcome via the website: www.automotive-technology.co.uk.

CHAPTER 5

Batteries

5.1 Vehicle batteries

5.1.1 Requirements of the vehicle battery

The vehicle battery is used as a source of energy in the vehicle when the engine, and hence the alternator, is not running. The battery has a number of requirements, which are listed below broadly in order of importance.

- To provide power storage and be able to supply it quickly enough to operate the vehicle starter motor.
- To allow the use of parking lights for a reasonable time.
- To allow operation of accessories when the engine is not running.
- To act as a swamp to damp out fluctuations of system voltage.
- To allow dynamic memory and alarm systems to remain active when the vehicle is left for a period of time.

The first two of the above list are arguably the most important and form a major part of the criteria used to determine the most suitable battery for a given application. The lead-acid battery, in various similar forms, has to date proved to be the most suitable choice for vehicle use. This is particularly so when the cost of the battery is taken into account.

Key fact

The lead-acid battery, has to date proved to be the most suitable choice for vehicle use.

Figure 5.1 Batteries

Key fact

Batteries should be able to work from –30 to +70 °C.

The final requirement of the vehicle battery is that it must be able to carry out all the above listed functions over a wide temperature range. This can be in the region of –30 to 70 °C. This is intended to cover very cold starting conditions as well as potentially high under-bonnet temperatures.

5.1.2 Choosing the correct battery

The correct battery depends, in the main, on just two conditions.

1. The ability to power the starter to enable minimum starting speed under very cold conditions.
2. The expected use of the battery for running accessories when the engine is not running.

The first of these two criteria is usually the deciding factor. Figure 5.2 shows a graph comparing the power required by the starter and the power available from the battery, plotted against temperature. The point at which the lines cross is the cold start limit of the system (see also the chapter on starting systems). European standards generally use the figure of 18 °C as the cold start limit and a battery to meet this requirement is selected.

Research has shown that under ‘normal’ cold operating conditions in the UK, most vehicle batteries are on average only 80% charged. Many manufacturers choose a battery for a vehicle that will supply the required cold cranking current when in the 80% charged condition at 7 °C.

5.1.3 Positioning the vehicle battery

Several basic points should be considered when choosing the location for the vehicle battery:

- Weight distribution of vehicle components.
- Proximity to the starter to reduce cable length.
- Accessibility.
- Protection against contamination.
- Ambient temperature.
- Vibration protection.

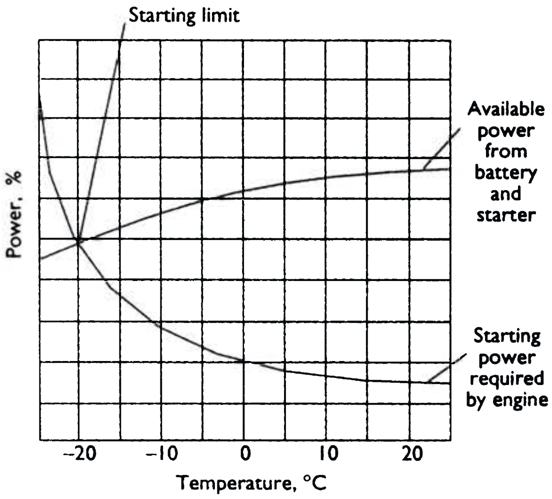


Figure 5.2 Comparison of the power required by the starter and the power available from the battery plotted against temperature

As usual, these issues will vary with the type of vehicle, intended use, average operating temperature and so on. Extreme temperature conditions may require either a battery heater or a cooling fan. The potential build-up of gases from the battery may also be a consideration.

5.2 Lead-acid batteries

5.2.1 Construction

Even after well over 100 years of development and much promising research into other techniques of energy storage, the lead-acid battery is still the best choice for motor vehicle use. This is particularly so when cost and energy density are taken into account.

Incremental changes over the years have made the sealed and maintenance-free battery now in common use very reliable and long lasting. This may not always appear to be the case to some end-users, but note that quality is often related to the price the customer pays. Many bottom-of-the-range cheap batteries, with a 12 month guarantee, will last for 13 months!

The basic construction of a nominal 12 V lead-acid battery consists of six cells connected in series. Each cell, producing about 2 V, is housed in an individual compartment within a polypropylene, or similar, case. Figure 5.4 shows a cut-away battery showing the main component parts. The active material is held in grids or baskets to form the positive and negative plates. Separators made from a micro-porous plastic insulate these plates from each other.

The grids, connecting strips and the battery posts are made from a lead alloy. For many years this was lead antimony (PbSb) but this has now been largely

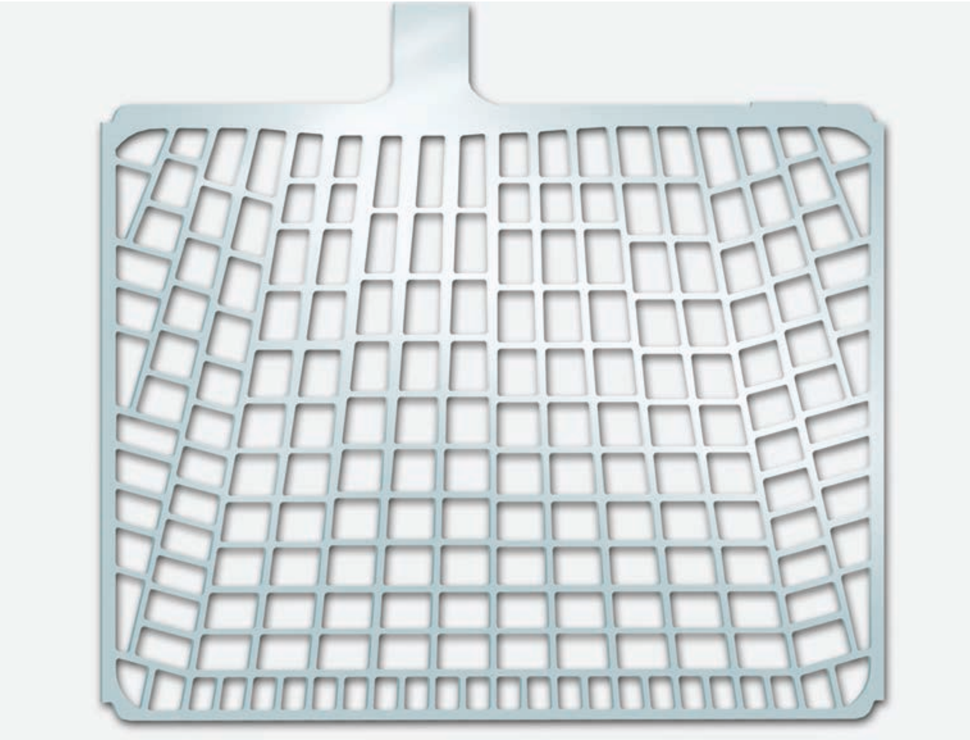


Figure 5.3 Battery grid before the active materials are added

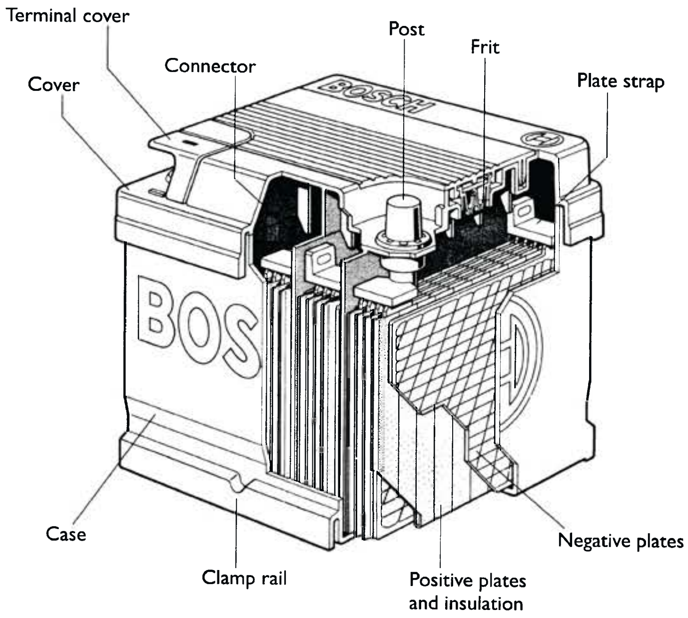


Figure 5.4 Lead-acid battery

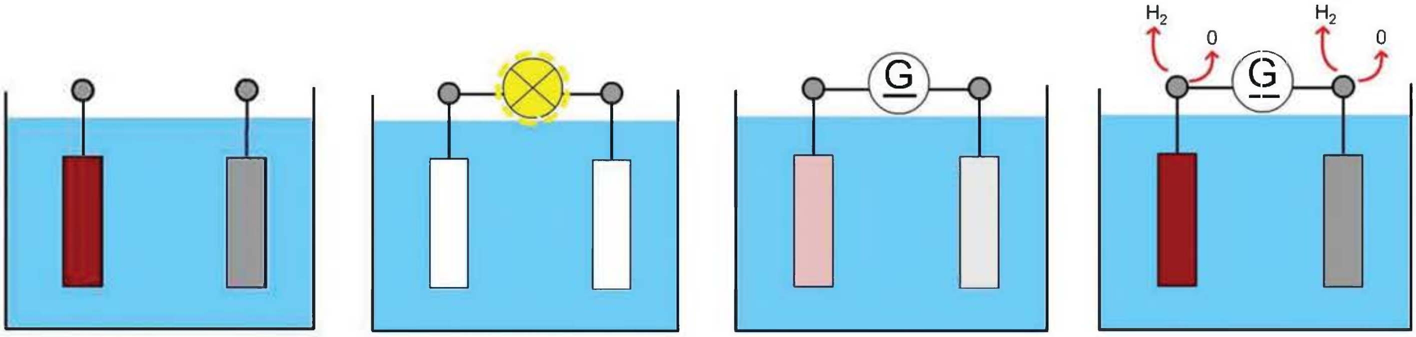


Figure 5.5 Battery discharge and charging process (left to right): Fully charged; discharging; charging; charging and gassing

replaced by lead calcium (PbCa). The newer materials cause less gassing of the electrolyte when the battery is fully charged. This has been one of the main reasons why sealed batteries became feasible, as water loss is considerably reduced (Figure 5.5).

However, even modern batteries described as sealed do still have a small vent to stop the pressure build-up due to the very small amount of gassing. A further requirement of sealed batteries is accurate control of charging voltage.

5.2.2 Battery rating

In simple terms, the characteristics or rating of a particular battery are determined by how much current it can produce and how long it can sustain this current.

The rate at which a battery can produce current is determined by the speed of the chemical reaction. This in turn is determined by a number of factors:

- Surface area of the plates.
- Temperature.
- Electrolyte strength.
- Current demanded.

Figure 5.6 Battery showing the information and status panel (Source: Bosch Media)

The actual current supplied therefore determines the overall capacity of a battery. The rating of a battery has to specify the current output and the time.

This method describes how much current the battery is able to supply for either 10 or 20 hours. The 20-hour figure is the most common. For example, a battery quoted as being 44 Ah (ampere-hour) will be able, if fully charged, to supply 2.2 A for 20 hours before being completely discharged (cell voltage above 1.75 V).

Key fact

The rate at which a battery can produce current is determined by the speed of the chemical reaction.

Reserve capacity

A system used now on all new batteries is reserve capacity. This is quoted as a time in minutes for which the battery will supply 25 A at 25 °C to a final voltage of 1.75 V per cell. This is used to give an indication of how long the battery could run the car if the charging system was not working. Typically, a 44 Ah battery will have a reserve capacity of about 60 minutes.

Cold cranking amps

Batteries are given a rating to indicate performance at high current output and at low temperature. A typical value of 170 A means that the battery will supply this current for one minute at a temperature of 18 °C, at which point the cell voltage will fall to 1.4 V (BS – British Standards).

Note that the overall output of a battery is much greater when spread over a longer time. As mentioned above, this is because the chemical reaction can only work at a certain speed. Figure 5.7 shows the above three discharge characteristics and how they can be compared.

The cold cranking amps (CCA) capacity rating methods do vary to some extent; British standards, DIN standards and SAE standards are the three main examples.

Table 5.1 CCA standards

Standard	Time (seconds)
BS	60
DIN	30
SAE	30

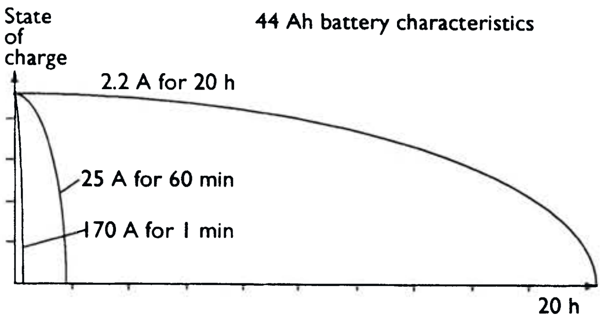


Figure 5.7 Battery discharge characteristics compared

In summary, the capacity of a battery is the amount of electrical energy that can be obtained from it. It is usually given in ampere-hours (Ah), reserve capacity (RC) and cold cranking amps (CCA).

- A 44 Ah battery means it should give 2.2 A for 20 hours.
- The reserve capacity indicates the time in minutes for which the battery will supply 25 A at 25 °C.
- Cold cranking current indicates the maximum battery current at 18°C (0°F) for a set time (standards vary).

A battery for normal light vehicle use may be rated as follows: 44 Ah, 60 RC and 170 A CCA (BS). Figure 5.7 shows the discharge characteristics of this battery. A 'heavy duty' battery will have the same Ah rating as its 'standard duty' counterpart, but it will have a higher CCA and RC.

5.3 Maintenance, charging and testing batteries

5.3.1 Maintenance

By far the majority of batteries now available are classed as 'maintenance free'. This implies that little attention is required during the life of the battery. Earlier batteries and some heavier types do, however, still require the electrolyte level to be checked and topped up periodically. Battery posts are still a little prone to corrosion and hence the usual service of cleaning with hot water if appropriate and the application of petroleum jelly or proprietary terminal grease is still recommended. Ensuring that the battery case and, in particular, the top remains clean, will help to reduce the rate of self-discharge.

Key fact

By far the majority of batteries are now maintenance free.

The state of charge of a battery is still very important and, in general, it is not advisable to allow the state of charge to fall below 70% for long periods as the sulphate on the plates can harden, making recharging difficult. If a battery is to be stored for a long period (more than a few weeks), then it must be recharged every so often to prevent it from becoming sulphated. Recommendations vary but a recharge every six weeks is a reasonable suggestion.

5.3.2 Charging the lead-acid battery

The recharging recommendations of battery manufacturers vary slightly. The following methods, however, are reasonably compatible and should not cause any problems. The recharging process must 'put back' the same ampere-hour capacity as was used on discharge plus a bit more to allow for losses. It is therefore clear that the main question about charging is not how much, but at what rate.

Figure 5.8 Battery charger (Source: Bosch Media)

The traditional recommendation was that the battery should be charged at a tenth of its ampere-hour capacity for about 10 hours or less. This is assuming that the ampere-hour capacity is quoted at the 20 hour rate, as a tenth of this figure will make allowance for the charge factor. This figure is still valid, but as ampere-hour capacity is not always used nowadays, a different method of deciding the rate is necessary.

One way is to set a rate at 1/16 of the reserve capacity, again for up to 10 hours. The final suggestion is to set a charge rate at 1/40 of the cold start performance figure, also for up to 10 hours. Clearly, if a battery is already half charged, half the time is required to recharge to full capacity.

Key fact

The ideal charge rate is determined as:

1/10 of the Ah capacity;

1/16 of the RC;

1/40 of the CCA.

The above suggested charge rates are to be recommended as the best way to prolong battery life. They do all, however, imply a constant current charging source. A constant voltage charging system is often the best way to charge a battery. This implies that the charger, an alternator on a car for example, is held at a constant level and the state of charge in the battery will determine how much current will flow. This is often the fastest way to recharge a flat battery. The two ways of charging are represented in Figure 5.9. This shows the relationship between charging voltage and the charging current. If a constant voltage of less than 14.4 V is used then it is not possible to cause excessive gassing and this method is particularly appropriate for sealed batteries. A combination of the two charging methods is the best and fastest method.

Boost charging (Figure 5.10) is a popular technique often applied in many workshops. It is not recommended as the best method but, if correctly administered and not repeated too often, is suitable for most batteries. The key to fast or boost charging is that the battery temperature should not exceed 43 °C. With sealed batteries it is particularly important not to let the battery create excessive gas in order to prevent the build-up of pressure. A rate of about five times the ‘normal’ charge setting will bring the battery to 78–80% of its full capacity within approximately one hour.

There are now a number of ‘Smart’ or ‘Intelligent’ battery chargers that are able to determine the ideal rate from the battery voltage and the current it will accept. Some also have features such as a ‘recond’ mode, which allows you to correct the acid stratification that often occurs in deeply discharged

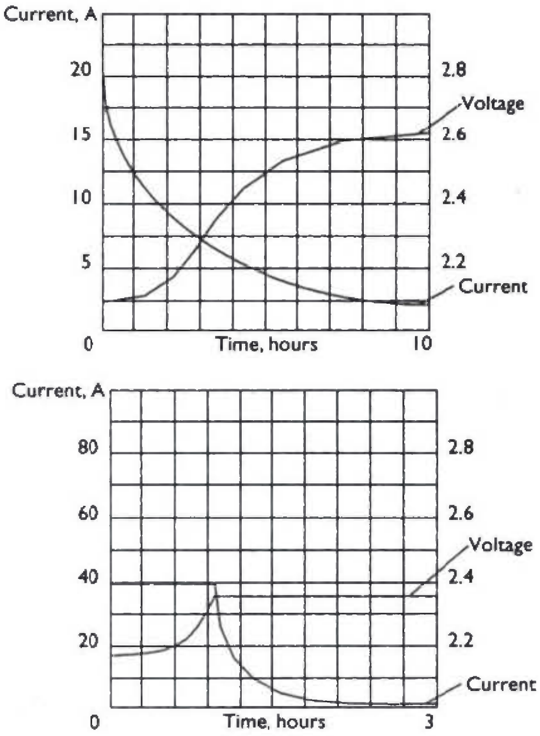


Figure 5.9 Two ways of charging a battery showing the relationship between charging voltage and charging current

Figure 5.10 Battery charger and engine starter (Source: Bosch Media)

batteries – particularly leisure batteries. Some key features of a charger produced by a company called Ctek are:

- Safe: No sparks and cannot harm vehicle electrics, so no need to disconnect the battery.
- Suitable for all types of 12 V lead-acid batteries up to 150 Ah.
- Connect and forget – can be left connected for months – ideal for vehicles used occasionally.
- Analysis mode to check if battery can hold charge.
- 10 day float maintenance for maximum charge level.
- ‘Recond’ mode – special programme to revive deeply discharged batteries.
- Supply mode – can be used as a 12 V power source to protect electrical settings.

Table 5.2 Charging methods summary

Charging method	Notes
Constant voltage	Will recharge any battery in seven hours or less without any risk of overcharging (14.4 V maximum).
Constant current	Ideal charge rate can be estimated as: 1/10 of Ah capacity, 1/16 of reserve capacity or 1/40 of cold start current (charge time of ten to twelve hours or pro rata original state).
Boost charging	At no more than five times the ideal rate, a battery can be brought up to about 70% of charge in about one hour.
Smart charging	Let the charger do all the calculations and all the work!

Figure 5.11 Smart battery charger (Source: www.ctek.com)

5.3.3 Servicing batteries

In use, a battery requires very little attention other than the following when necessary:

- Clean corrosion from terminals using hot water.
- Terminals should be smeared with petroleum jelly or Vaseline, not ordinary grease.
- Battery tops should be clean and dry.
- If not sealed, cells should be topped up with distilled water 3 mm above the plates.
- The battery should be securely clamped in position.

5.3.4 Battery faults

Any electrical device can suffer from two main faults; these are either open circuit or short circuit. A battery is no exception but can also suffer from other problems, such as low charge or low capacity. Often a problem – apparently with the vehicle battery – can be traced to another part of the vehicle such as the charging system. Table 5.3 lists all of the common problems encountered with lead-acid batteries, together with typical causes.

Repairing modern batteries is not possible. Most of the problems listed will require the battery to be replaced. In the case of sulphation it is sometimes possible to bring the battery back to life with a very long low current charge. A fortieth of the ampere-hour capacity or about a 1/200 of the cold start performance, for about 50 hours, is an appropriate rate.

5.3.5 Testing batteries

For testing the state of charge of a non-sealed type of battery, a hydrometer can be used, as shown in Figure 5.12. The hydrometer comprises a syringe that draws electrolyte from a cell, and a float that will float at a particular depth in the electrolyte according to its density. The density or specific gravity is then read from the graduated scale on the float. A fully charged cell should show 1.280, 1.200 when half charged and 1.130 if discharged.

Table 5.3 Common problems with lead-acid batteries and their likely causes

Symptom or fault	Likely causes
Low state of charge	Charging system fault Unwanted drain on battery Electrolyte diluted Incorrect battery for application
Low capacity	Low state of charge Corroded terminals Impurities in the electrolyte Sulphated Old age – active material fallen from the plates
Excessive gassing and temperature	Overcharging Positioned too near exhaust component
Short circuit cell	Damaged plates and insulators Build-up of active material in sediment trap
Open circuit cell	Broken connecting strap Excessive sulphation Very low electrolyte
Service life shorter than expected	Excessive temperature Battery has too low a capacity Vibration excessive Contaminated electrolyte Long periods of not being used Overcharging

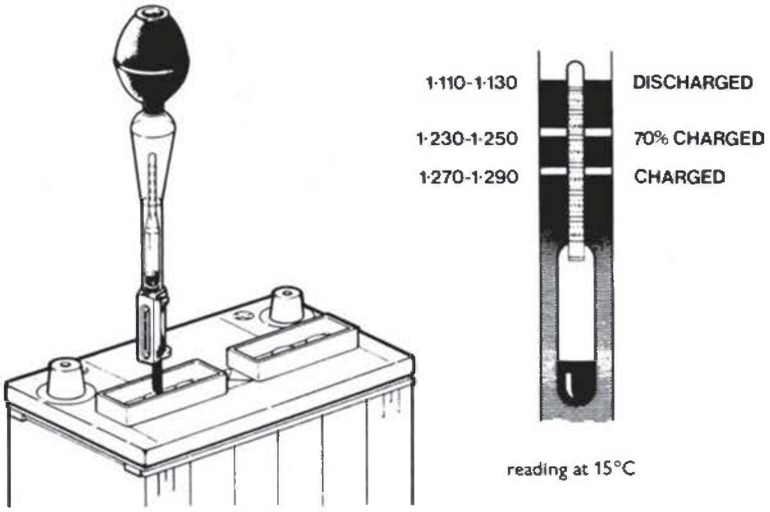


Figure 5.12 Hydrometer test of a battery

Almost all vehicles are now fitted with maintenance-free batteries and a hydrometer cannot be used to find the state of charge. This can only be determined from the voltage of the battery, as given in Table 5.4. An accurate voltmeter is required for this test.

A heavy-duty (HD) discharge tester as shown in Figure 5.14 is an instrument consisting of a low-value resistor and a voltmeter connected to a pair of heavy test prods. The test prods are firmly pressed on to the battery terminals. The voltmeter reads the voltage of the battery on heavy discharge of 200–300 A.

Figure 5.13 Checking battery voltage. In this case the engine had just been switched off so the reading shows a 'surface charge'

Table 5.4 Battery voltage readings and state of charge

Battery volts at 20 °C	State of charge
12.0	Discharged (20% or less)
12.3	Half charged (50%)
12.7	Charged (100%)

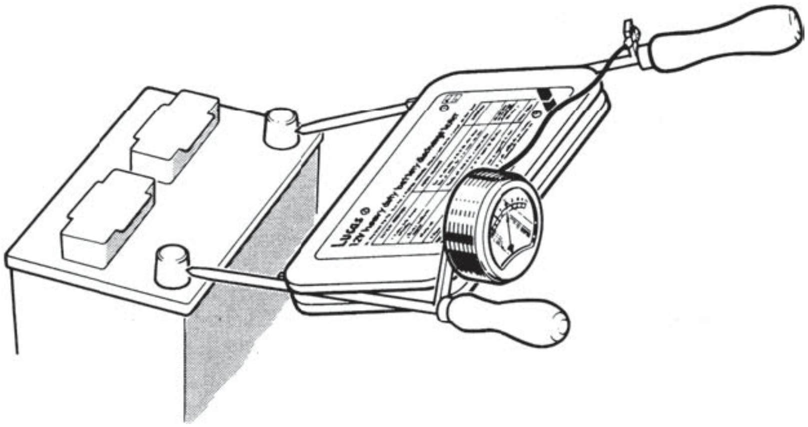


Figure 5.14 An early type of heavy duty discharge tester

Assuming a battery to be in a fully charged condition, a serviceable battery should read about 10 V for a period of about 10 s. A sharply falling battery voltage to below 3 V indicates an unserviceable cell. Note also if any cells are gassing, as this indicates a short circuit. A zero or extremely low reading can indicate an open circuit cell. When using the HD tester, the following precautions must be observed:

- Blow gently across the top of the battery to remove flammable gases.
- The test prods must be positively and firmly pressed into the lead terminals of the battery to minimize sparking.
- It should not be used while a battery is on charge.

Figure 5.15 MicroVAT (Source: Snap-on)

To test a battery safely and more thoroughly, it is now preferred to use a volt, amp tester (VAT). There are many variations on the market; however, this section will outline just one type. Snap-on produce a compact and very useful tester called the MicroVAT. This equipment will carry out a range of diagnostic tests.

The device, as with many similar types, will do not only battery condition tests, but also tests on the charging and starting system.

This VAT takes advantage of new impedance/ current test technology to detect the full range of battery failure modes including bad cells, sulphation, internal short circuits, and other chemical and physical failures. Testing takes less than 5 seconds and will even work on batteries discharged down to as low as one volt. Some of the key features of this tester are:

- Automated system test of battery, alternator and starter in under a minute.
- Detailed test data: alternator ripple, internal resistance, starter draw, state of charge, charging amps, and volts.
- Tests discharged batteries down to one volt.
- Impedance/current (IC) test technology.
- Wireless printer option.
- Integrated high and low amp probe options.

The MicroVAT uses a fan cooled 50 A load and integrated amp probe to test the quantity and quality of alternator output with an alternator ripple test. Many late model computer-controlled charging systems virtually shut down under no load conditions. Diagnostic tests that can be carried out with this tester, when an amps probe is also used, are as follows.

Starting test data

- Average cranking current
- Maximum cranking current
- Pre-set voltage
- Pre-set load voltage
- Average cranking voltage
- Minimum cranking voltage.

Battery test data

- Diagnosis
- Actual CCA
- Percentage capacity
- Open circuit voltage
- Impedance (often described as internal resistance).

Alternator test data

- Diagnosis
- Failure mode
- Charging at idle
- Charging volts under load
- Average current at idle

Figure 5.16 Heavy duty batteries

- Peak current
- Peak to peak ripple at idle
- Peak to peak ripple under load.

5.3.6 Safety

The following points must be observed when working with batteries:

- Good ventilation.
- Protective clothing.
- Supply of water available (running water preferable).
- First aid equipment available, including eye-wash.
- No smoking or naked lights permitted.

5.4 Advanced battery technology

5.4.1 Electrochemistry

Electrochemistry is a very complex and wide-ranging science. This section is intended only to scratch the surface by introducing important terms and concepts. This will be helpful with the understanding of vehicle battery operation.

The branch of electrochemistry of interest here is the study of galvanic cells and electrolysis. When an electric current is passed through an electrolyte it causes certain chemical reactions and a migration of material. Some chemical reactions when carried out under certain conditions will produce electrical energy at the expense of the free energy in the system.

The reactions of most interest are those, which are reversible, in other words, can convert electrical energy into chemical energy and vice versa. Some of the terms associated with electrochemistry can be confusing. The following is a selection of terms and names with a brief explanation of each.

Table 5.5 Electrical and chemical terminology

Anion	The negative charged ion that travels to the positive terminal during electrolysis.
Anode	Positive electrode of a cell.
Catalyst	A substance that significantly increases the speed of a chemical reaction without appearing to take part in it.
Cation	The positive charged ion which travels to the negative terminal during electrolysis.
Cathode	Negative electrode of a cell.
Diffusion	The self-induced mixing of liquids or gasses.
Dissociation	The molecules or atoms in a solution decomposing into positive and negative ions. For example sulphuric acid (H_2SO_4), dissociates into H^{++}, H^{++} (two positive ions or cations, which are attracted to the cathode), and SO_4^{--} (negative ions or anions, which are attracted to the anode).
Electrode	Plates of a battery or an electrolysis bath suspended in the electrolyte.
Electrolysis	Conduction of electricity between two electrodes immersed in a solution containing ions (electrolyte), which causes chemical changes at the electrodes.
Electrolyte	An ion conducting liquid covering both electrodes.
Ion	A positively or negatively charged atomic or molecular particle.
Secondary galvanic cell	A cell containing electrodes and electrolyte, which will convert electrical energy into chemical energy when being charged, and the reverse during discharge.

5.4.2 Electrolytic conduction

Electricity flows through conductors in one of two ways. The first is by electron movement as is the case with most metals. The other type of flow is by ionic movement, which may be charged atoms or molecules. For electricity to flow through an electrolyte, ion flow is required.

To explain electrolytic conduction, which is a current flow through a liquid, sulphuric acid (H_2SO_4) is the best electrolyte example to choose. When in an aqueous solution (mixed with water), sulphuric acid dissociates into H^{++}, H^{++} and SO_4^{--}, which are positive and negative ions. The positive charges are attracted to the negative electrode and the negative charges are attracted to the positive electrode. This movement is known as ion flow or ion drift.

5.4.3 Ohm's Law and electrolytic resistance

The resistance of any substance depends on the following variables:

- nature of the material
- temperature
- length
- cross sectional area.

This is true for electrolyte as well as solid conductors. Length and cross sectional area have straightforward effects on the resistance of a sample be it a solid or a liquid. Unlike most metals however, which have a positive temperature coefficient, electrolytes are generally the opposite and have a negative temperature coefficient.

The nature of the material or its conductance (the reciprocal of resistance) is again different between solids and liquids. Different substances have different values of resistivity but with electrolytes the concentration is also important.

5.4.4 Electrochemical action of the lead-acid battery

A fully charged lead-acid battery consists of lead peroxide (PbO_2) as the positive plates, spongy lead (Pb) as the negative plates and diluted sulphuric acid (H_2SO_4) + (H_2O). The dilution of the electrolyte is at a relative density of 1.28. The lead is known as the active material and in its two forms has different valencies. This means a different number of electrons exists in the outer shell of the pure lead than when present as a compound with oxygen. The lead peroxide has in fact a valency of +iv (four electrons missing!).

As discussed earlier, when sulphuric acid is in an aqueous solution (mixed with water), it dissociates into charged ions H^{++}, H^{++} and SO_4^{--}. From the 'outside' the polarity of the electrolyte appears to be neutral as these charges cancel out. The splitting of the electrolyte into these parts is the reason that a charging or discharging current can flow through the liquid.

The voltage of a cell in is created due to the ions (charged particles) being forced into the solution from the electrodes by the solution pressure. Lead will give up two positively charged atoms, which have given up two electrons, into the liquid. As a result of giving up two positively charged particles the electrode will now have an excess of electrons and hence take on a negative polarity with respect to the electrolyte. If a further electrode is immersed into the electrolyte, different potentials will develop at the two electrodes and therefore a potential difference will exist between the two. A lead-acid battery has a nominal potential difference of 2V. The electrical pressure now present between the plates results in equilibrium within the electrolyte. This is because the negative charges on one plate exert an attraction on the positive ions, which have entered the solution. This attraction has the same magnitude as the solution pressure and hence equilibrium is maintained.

When an external circuit is connected to the cell the solution pressure and attraction force is disrupted. This allows additional charged particles to be passed into and through the electrolyte. This will only happen however if the external voltage pressure is greater than the electrical tension within the cell. In simple terms this is known as the charging voltage.

When a lead-acid cell is undergoing charging or discharging certain chemical changes take place. This can be considered as two reactions, one at the positive plate and one at the negative plate. The electrode reaction at the positive plate is a combination of equations (a) and (b).

(a) $PbO_2 + 4H^+ + 2e^- \rightarrow Pb^{++} + 2H_2O$

The lead peroxide combines with the dissociated hydrogen and tends to become lead and water.

(b) $Pb^{++} + SO_4^{--} \rightarrow PbSO_4$

The lead now tends to combine with the sulphate from the electrolyte to become lead sulphate. This gives the overall reaction at the positive pole (a + b) as:

(c) $PbO_2 + 4H^+ + SO_4^{--} + 2e^- \rightarrow PbSO_4 + 2H_2O$

There is a production of water (a) and a deposition of lead sulphate (b) together with a consumption of sulphuric acid.

The electrode reaction at the negative plate is:

(d) $Pb \rightarrow Pb^{++} + 2e^-$

The neutral lead loses two negative electrons to the solution becoming positively charged.

(e) $Pb^{++}\ SO_4^{--} \rightarrow PbSO_4$

This then tends to attract the negative charged sulphate from the solution and the pole becomes lead sulphate. The overall reaction at the negative pole is therefore (d + e):

(f) $Pb + SO_4^{--} \rightarrow PbSO_4 + 2e^-$

This reaction leads to a consumption of sulphuric acid and the production of water as the battery is discharged.

The reverse of the above process is when the battery is being charged. The process is the reverse of that described above. The reactions involved in the charging process are listed below.

The charging reaction at the negative electrode:

(g) $PbSO_4 + 2e^- + 2H^+ \rightarrow Pb + H_2SO_4$

The electrons from the external circuit ($2e^-$) combine with the hydrogen ions in the solution ($2H^+$) and then the sulphate to form sulphuric acid as the plate tends to become lead. The reaction at the positive pole is:

(h) $PbSO_4 - 2e^- + 2H_2O \rightarrow PbO_2 + H_2SO_4 + 2H^+$

The electrons given off to the external circuit ($2e^-$), releases hydrogen ions into the solution ($2H^+$). This allows the positive plate to tend towards lead peroxide and the concentration of sulphuric acid in the electrolyte to increase.

The net two-way chemical reaction is the sum of the above electrode processes (c + f or g + h):

(i) $PbO_2 + 2H_2SO_4 + Pb \Leftrightarrow 2PbSO_4 + 2H_2O$

This two-way or reversible chemical reaction (charged on the left and discharged on the right), describes the full process of the charge and discharge cycle of the lead-acid cell.

The other reaction of interest in a battery is that of gassing after it has reached full charged condition. This occurs because one the plates of the battery have become ‘pure’ lead and lead peroxide, the external electrical supply will cause the water in the electrolyte to decompose. This gassing voltage for a lead-acid battery is about 2.4V. This gassing causes hydrogen and oxygen to be given off resulting in loss of water (H_2O), and an equally undesirable increase in electrolyte acid density.

The reaction as before can be considered for each pole of the battery in turn.

At the positive plate:

(j) $2H_2O - 4e^- \rightarrow O_2 + 4H^+$

At the negative plate:

(k) $4H^+ + 4e^- \rightarrow 2H_2$

The sum of these two equations gives the overall result of the reaction;

(l) (j + k) $2H_2O \rightarrow O_2 + 2H_2$

It is acceptable for gassing to occur for a short time to ensure all the lead sulphate has been converted to either lead or lead peroxide. It is the material of the grids inside a battery, which contribute to the gassing. With sealed batteries this is a greater problem but has been overcome to a large extent by using lead-calcium for the grid material in place of the more traditional lead-antimony.

The voltage of a cell and hence the whole battery is largely determined by the concentration of the acid in the electrolyte. The temperature also has a marked effect. This figure can be calculated from the mean electrical tension of the plates and the concentration of ions in solution. The following table lists the results of these calculations at 27 °C. As a rule of thumb guide the cell voltage is about 0.84 plus the value of the relative density.

Table 5.6 Factors affecting the voltage of a battery

Acid density	Cell voltage	Battery Voltage	% Charge
1.28	2.12 V	12.7 V	100
1.24	2.08 V	12.5 V	70
1.20	2.04 V	12.3 V	50
1.15	1.99 V	12.0 V	20
1.12	1.96 V	11.8 V	0

It is accepted that the terminal voltage of a lead-acid cell must not be allowed to fall below 1.8 V as apart from the electrolyte tending to become very close to pure water, the lead sulphate crystals grow markedly making it very difficult to recharge the battery.

Key fact

The terminal voltage of a standard lead-acid lead-acid battery must not be allowed to fall below 1.8 V. However, some batteries are designed to allow deep discharge.

5.4.5 Characteristics

The following headings are characteristics of a battery that determine its operation and condition.

Internal resistance

Any source of electrical energy can be represented by the diagram shown as Figure 5.17. This shows a perfect voltage source in series with a resistor. This is used to represent why the terminal voltage of a battery drops when a load is placed across it. As an open circuit no current flows through the internal resistance and hence no voltage is dropped. When a current is drawn from the source a voltage drop across the internal resistance will occur.

The actual value can be calculated as follows:

Connect a voltmeter across the battery and note the open circuit voltage, for example 12.7 V. Connect an external load to the battery, measure the current, say 50 A. Note again the on load terminal voltage of the battery, for example 12.2 V.

A calculation will determine the internal resistance:

$$R_i = (U - V)/I$$

where: U = open circuit voltage, V = on load voltage, I = current, R_i = internal resistance.

For this example the result of the calculation is 0.01 Ω.

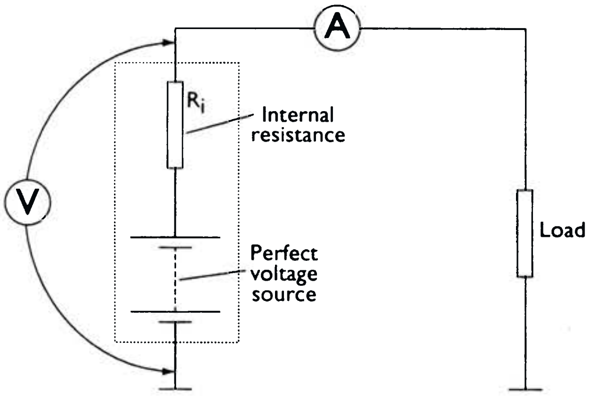


Figure 5.17 Internal resistance representation (the resistor R_i represents internal component resistance, an actual resistor is not part of a battery!)

Temperature and state of charge affect the internal resistance of a battery. It can also be used as an indicator of battery condition, the lower the figure then the better condition.

Efficiency

The efficiency of a battery can be calculated in two ways, either as the ampere hour efficiency or the power efficiency.

$$\text{Ah efficiency} = (\text{Ah discharging}/\text{Ah charging}) \times 100\%$$

At the 20 hour rate this can be as much as 90%. This is often quoted as the reciprocal of the efficiency figure, in this example about 1.1 which is known as the charge factor.

$$\text{Energy efficiency} = (P_d \times t_d)/(P_c \times t_c) \times 100\%$$

where: P_d *= discharge power,* t_d *= discharge time,* P_c *= charging power,* t_c *= charging time.*

A typical result of this calculation is about 75%. This figure is lower than the Ah efficiency as it takes into account the higher voltage required to force the charge into the battery.

Self-discharge

All batteries suffer from self-discharge which means that even without an external circuit the state of charge is reduced. The rate of discharge is in the order of 0.2 to 1% of the Ah capacity per day. This increases with temperature and the age of the battery. It is caused by two factors; firstly the chemical process inside the battery due to the material of the grids forming short circuit voltaic couples between the antimony and the active material. Using calcium as the mechanical improver for the lead grids reduces this. Impurities in the electrolyte in particular trace metals such as iron can also add to self-discharge.

Key fact

All batteries suffer from self-discharge which means that even without an external circuit the state of charge is reduced.

A leakage current across the top of the battery particularly if it is in a poor state of cleanliness, also contributes to the self-discharge. The fumes from the acid together with particles of dirt can form a conducting film. This problem is much reduced with sealed batteries.

5.4.6 Peukert's Law

Peukert's Law expresses the capacity of a battery in terms of the rate at which it is discharged. As the rate increases, the battery's capacity decreases.

$$T = \frac{C}{\left(\frac{I}{C/R}\right)^n} \times \frac{R}{C}$$

where: I = Discharge current (A); T = Time (h); C = Capacity of the battery (Ah); n = Peukerts exponent for that particular battery type, Ideal battery = 1, generally between 1.1 and 1.3 depending on type; R = the battery hour rating, i.e. 20 hour rating, 10 hour rating etc.

5.5 Developments in electrical storage

5.5.1 Lead-acid

Lead-acid batteries have not changed much from the very early designs (invented by Gaston Plante in 1859). Incremental changes and, in particular, the development of accurate charging system control has allowed the use of sealed and maintenance-free batteries. Figure 5.18 shows a typical modern battery.

Figure 5.18 Modern vehicle battery

The other main developments have been to design batteries for particular purposes. This is particularly appropriate for uses such as supplementary batteries in a caravan or as power supplies for lawn mowers and other traction uses. These batteries are designed to allow deep discharge and, in the case of caravan/trailer batteries, may also have vent tubes fitted to allow gases to be vented outside. Some batteries are designed to withstand severe vibration for use on plant-type vehicles.

The processes in lead-acid batteries are very similar, even with variations in design. However, batteries using a gel in place of liquid electrolyte are worth a mention. These batteries have many advantages in that they do not leak and are more resistant to poor handling.

The one main problem with using a gel electrolyte is that the speed of the chemical reaction is reduced. Whilst this is not a problem for some types of supply, the current required by a vehicle starter is very high for a short duration. The cold cranking amps (CCA) capacity of this type of battery is therefore often lower than the equivalent-sized conventional battery.

The solid-gel type electrolyte used in some types of these batteries is thixotropic. This means that, due to a high viscosity, the gel will remain immobile even if the battery is inverted. A further advantage of a solid gel electrolyte is that a network of porous paths is formed through the electrolyte. If the battery is overcharged, the oxygen emitted at the positive plate will travel to the negative plate, where it combines with the lead and sulphuric acid to form lead sulphate and water:

$$O_2 + 2Pb = 2PbO$$
$$PbO + H_2SO_4 = PbSO_4 + H_2O$$

This reforming of the water means the battery is truly maintenance free. The recharging procedure is very similar to the more conventional batteries.

To date, gel-type batteries have not proved successful for normal motor vehicle use, but are an appropriate choice for specialist performance vehicles that are started from an external power source. Ordinary vehicle batteries using a gel electrolyte appeared on the market some years ago accompanied by great claims of reliability and long life. However, these batteries did not become very popular. This could have been because the cranking current output was not high enough due to the speed of the chemical reaction.

A development in ‘normal’ lead-acid batteries is the use of lead-antimony (PbSb) for the positive plate grids and lead-calcium (PbCa) for the negative plate grids. This results in a significant reduction in water loss and an increase in service life. The plates are sealed in micro-porous pocket-type separators, on each side of which are glass-fibre reinforcing mats. The pocket separators collect all the sludge and hence help to keep the electrolyte in good condition.

5.5.2 Alkaline

Lead-acid batteries traditionally required a considerable amount of servicing to keep them in good condition, although this is not now the case with the advent of sealed and maintenance-free batteries. However, when a battery is required to withstand a high rate of charge and discharge on a regular basis, or is left in a state of disuse for long periods, the lead-acid cell is not ideal. Alkaline cells on the other hand require minimum maintenance and are far better able to withstand electrical abuse such as heavy discharge and over-charging.

The disadvantages of alkaline batteries are that they are more bulky, have lower energy efficiency and are more expensive than a lead-acid equivalent. When the lifetime of the battery and servicing requirements are considered, the extra initial cost is worth it for some applications. Bus and coach companies and some large goods-vehicle operators have used alkaline batteries.

Alkaline batteries used for vehicle applications are generally the nickel-cadmium type, as the other main variety (nickel-iron) is less suited to vehicle use. The main components of the nickel-cadmium (NiCad) cell for vehicle use are as follows:

- positive plate – nickel hydrate (NiOOH);
- negative plate – cadmium (Cd);
- electrolyte – potassium hydroxide (KOH) and water (H_2O).

The process of charging involves the oxygen moving from the negative plate to the positive plate, and the reverse when discharging. When fully charged, the negative plate becomes pure cadmium and the positive plate becomes nickel hydrate. A chemical equation to represent this reaction is given next, but note that this is simplifying a more complex reaction.

$$2NiOOH + Cd + 2H_2O + KOH == 2Ni(OH)_2 + CdO_2 + KOH$$

Key fact

NiCad batteries do not suffer from over-charging because once the cadmium oxide has changed to cadmium, no further reaction can take place.

The $2H_2O$ is actually given off as hydrogen (H) and oxygen (O_2) as gassing takes place all the time during charge. It is this use of water by the cells that indicates they are operating, as will have been noted from the equation. The electrolyte does not change during the reaction. This means that a relative density reading will not indicate the state of charge. These batteries do not suffer from over-charging because once the cadmium oxide has changed to cadmium, no further reaction can take place.

The cell voltage of a fully charged cell is 1.4 V but this falls rapidly to 1.3 V as soon as discharge starts. The cell is discharged at a cell voltage of 1.1 V. Figure 5.19 shows a simplified representation of a NiCad battery cell. Ni-MH or nickel-metal-hydride batteries show some promise for electric vehicle use.

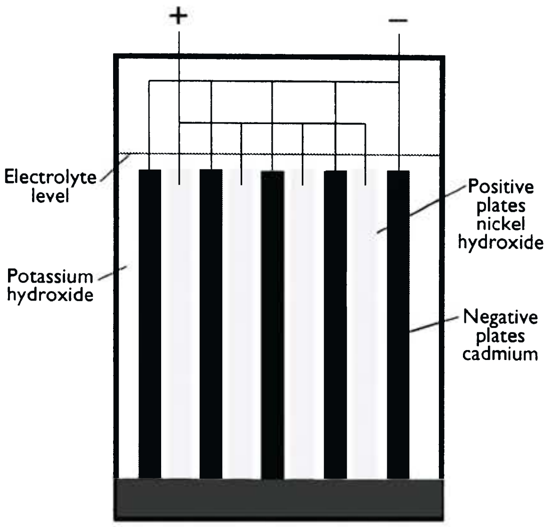


Figure 5.19 Simplified representation of a NiCad alkaline battery cell

5.5.3 ZEBRA

The Zero Emissions Battery Research Activity (ZEBRA) has adopted a sodium-nickel-chloride battery for use in its electric vehicle programme. This battery functions on an electrochemical principle. The base materials are nickel and sodium chloride. When the battery is charged, nickel chloride is produced on one side of a ceramic electrolyte and sodium is produced on the other. Under discharge, the electrodes change back to the base materials. Each cell of the battery has a voltage of 2.58 V.

The battery operates at an internal temperature of 270–350 °C, requiring a heat-insulated enclosure. The whole unit is 'vacuum packed' to ensure that the outer surface never exceeds 30 °C. The ZEBRA battery has a gravimetric energy density of 90 Wh/kg, which is more than twice that of a lead-acid type.

When in use on an electric vehicle (EV), the battery pack consists of 448 individual cells rated at 289 V. The energy density is 81 Wh/kg; it has a mass of 370 kg (over 1/4 of the total vehicle mass) and measures 993 x 793 x 280 mm^3. The battery pack can be recharged in just one hour using an external power source.

5.5.4 Sodium sulphur

Much research is underway to improve on current battery technology in order to provide a greater energy density for electric vehicles. A contender is the sodium sulphur battery, which has now reached production stage.

The sodium-sulphur or NaS battery consists of a cathode of liquid sodium into which is placed a current collector. This is a solid electrode of alumina. A metal can that is in contact with the anode (a sulphur electrode) surrounds the whole assembly. The major problem with this system is that the running temperature needs to be 300–350 °C. A heater rated at a few hundred watts forms part of the charging circuit. This maintains the battery temperature when the vehicle is not running. Battery temperature is maintained when in use due to I2 R losses in the battery.

Each cell of this battery is very small, using only about 15 g of sodium. This is a safety feature because, if the cell is damaged, the sulphur on the outside will cause the potentially dangerous sodium to be converted into polysulphides – which are comparatively harmless. Small cells also have the advantage that they can be distributed around the car. The capacity of each cell is about 10 Ah. These cells fail in an open circuit condition and hence this must be taken into account, as the whole string of cells used to create the required voltage would be rendered inoperative. The output voltage of each cell is about 2 V. Figure 5.20 shows a representation of a sodium-sulphur battery cell.

A problem still to be overcome is the casing material, which is prone to fail due to the very corrosive nature of the sodium. At present, an expensive chromized coating is used.

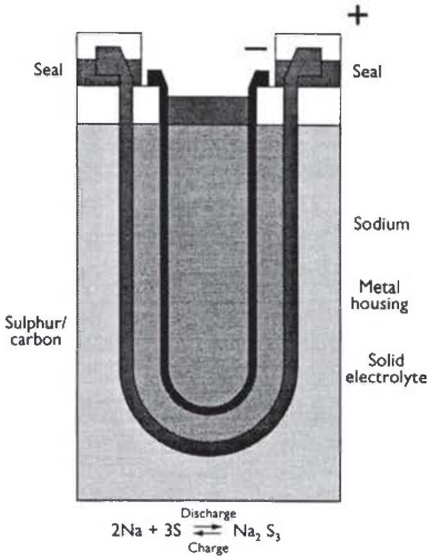


Figure 5.20 Sodium-sulphur battery

5.5.5 Swing

Some potential developments in battery technology are major steps in the right direction but many new methods involve high temperatures. One major aim of battery research is to develop a high performance battery, that works at a normal operating temperature. One new idea is called the 'Swing battery'.

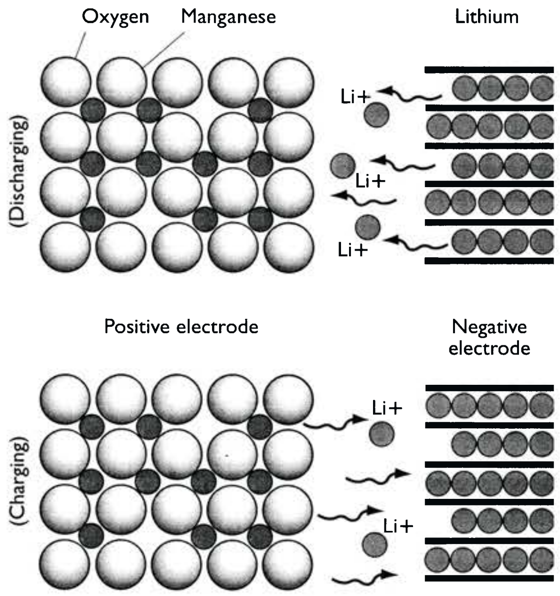


Figure 5.21 Chemical process of the 'Swing' battery

Key fact

The Swing concept batteries use lithium ions.

The Swing concept batteries use lithium ions. These batteries have a carbon anode and a cathode made of transition metal oxides. Lithium ions are in constant movement between these very thin electrodes in a non-aqueous electrolyte.

The Swing process takes place at normal temperatures and gives a very high average cell voltage of 3.5 V, compared with cell voltages of approximately 1.2 V for nickel-cadmium and about 2.1 V for lead-acid or sodium-sulphur batteries. The technology has now developed such that the following energy densities are achievable:

- Gravimetric power density of 180 Wh/kg.
- Volumetric power density of 420 Wh/L.

The complexity of the electrical storage system increases with higher operating temperatures, an increased number of cells and with the presence of agitated or recycled electrolytes. To ensure reliable and safe operation, higher and higher demands will be made on the battery management system. This will clearly introduce more cost to the vehicle system as a whole. Consideration must be given not only to specific energy storage but also to system complexity and safety. Figure 5.22 is a comparison of batteries considering energy density and safety factors.

The high temperature systems have proved their viability for use in vehicles and they have already passed a series of abuse tests. A sodium-sulphur battery when fully charged, which is rated at 20 kWh, contains about 10 kg of liquid sodium. Given 100 000 vehicles, 1000 tonnes of liquid sodium will be in use. These quantities have to be encapsulated in two hermetically sealed containers. For this reason, and safety factors, other battery technologies are now more popular.

The Swing concept offers a potentially safe system for use in the future.

5.5.6 Fuel cells

The energy of oxidation of conventional fuels, which is usually manifested as heat, can be converted directly into electricity in a fuel cell. All oxidations involve a transfer of electrons between the fuel and oxidant, and this is employed in a fuel cell to convert the energy directly into electricity. All battery cells involve an

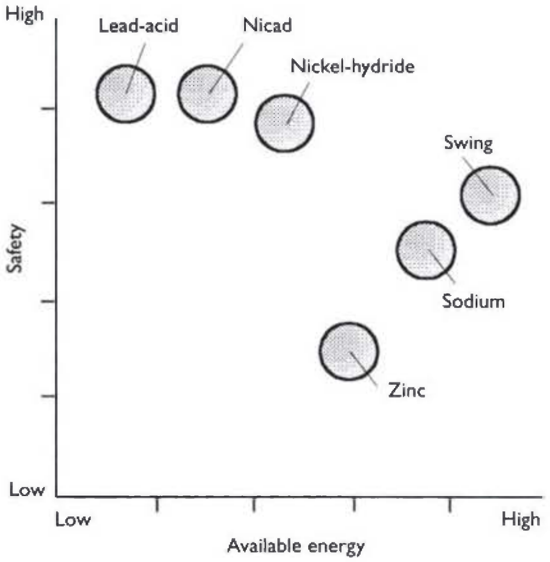


Figure 5.22 Comparison of battery technologies

oxide reduction at the positive pole and an oxidation at the negative during some part of their chemical process. To achieve the separation of these reactions in a fuel cell, an anode, a cathode and electrolyte are required. The electrolyte is fed directly with the fuel.

It has been found that a fuel of hydrogen when combined with oxygen proves to be a most efficient design. Fuel cells are very reliable and silent in operation, but are quite expensive to construct.

Operation of a fuel cell is such that as hydrogen is passed over an electrode (the anode), which is coated with a catalyst, the hydrogen diffuses into the electrolyte. This causes electrons to be stripped off the hydrogen atoms. These electrons

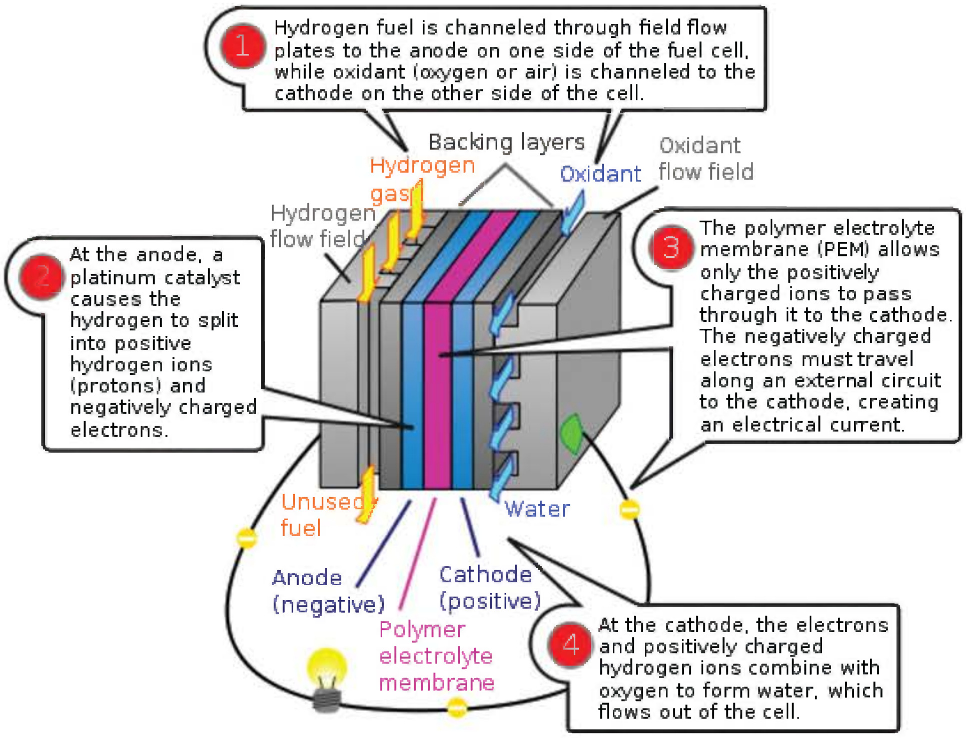


Figure 5.23 Fuel cell operation

then pass through the external circuit. Negatively charged hydrogen anions (OH) are formed at the electrode over which oxygen is passed such that it also diffuses into the solution. These anions move through the electrolyte to the anode. Water is formed as the by-product of a reaction involving the hydrogen ions, electrons and oxygen atoms. If the heat generated by the fuel cell is used, an efficiency of over 80% is possible, together with a very good energy density figure. A unit consisting of many individual fuel cells is often referred to as a stack.

Safety first

The working temperature of fuel cells varies but about 200 °C is typical. High pressure is also used and this can be of the order of 30 bar.

The working temperature of these cells varies but about 200 °C is typical. High pressure is also used and this can be of the order of 30 bar. It is the pressures and storage of hydrogen that are the main problems to be overcome before the fuel cell will be a realistic alternative to other forms of storage for the mass market.

Many combinations of fuel and oxidant are possible for fuel cells. Though hydrogen–oxygen is conceptually simple, hydrogen has some practical difficulties, including that it is a gas at standard temperature and pressure and that there does not currently exist an infrastructure for distributing hydrogen to domestic users. More readily usable, at least in the short term, would be a fuel cell powered by a more easily handled fuel. To this end, fuel cells have been developed which run on methanol. There are two types of fuel cell that use methanol:

- Reformed methanol fuel cell (RMFC).
- Direct methanol fuel cell (DMFC).

In the RMFC, a reaction is used to release hydrogen from the methanol, and then the fuel cell runs on hydrogen. The methanol is used as a carrier for hydrogen. The DMFC uses methanol directly. RMFCs can be made more efficient in the use of fuel than DMFCs, but are more complex.

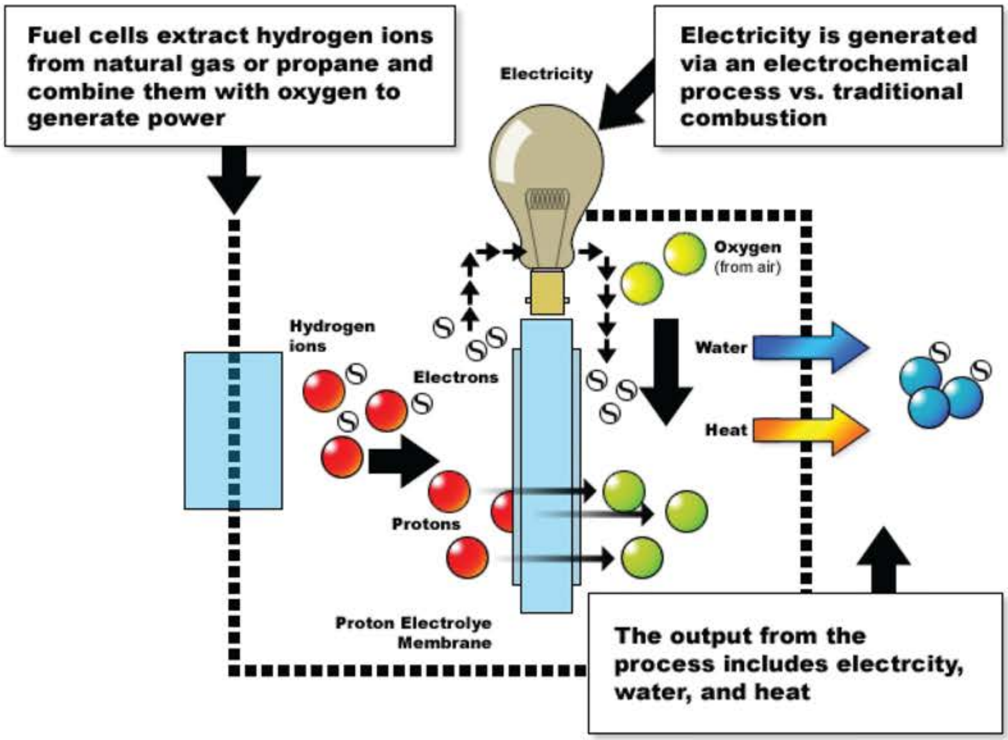


Figure 5.24 Fuel cell operation (Source: Dana)

DMFCs are a type of proton exchange membrane fuel cell (PEMFC). The membrane in a PEMFC fulfils the role of the electrolyte, and the protons (positively charged hydrogen ions) carry electrical charge between the electrodes.

Because the fuel in a DMFC is methanol, not hydrogen, other reactions take place at the anode. Methanol is a hydrocarbon (HC) fuel, which means that its molecules contain hydrogen and carbon (as well as oxygen in the case of methanol). When HCs burn, the hydrogen reacts with oxygen to create water and the carbon reacts with oxygen to create carbon dioxide. The same general process takes place in a DMFC, but in the process the hydrogen crosses the membrane as an ion, in just the same way as it does in a hydrogen-fuelled PEMFC.

The real benefit of methanol is that it can easily fit into the existing fuel infrastructure of filling stations and does not need highly specialized equipment or handling. It is easy to store on-board the vehicle, unlike hydrogen which needs heavy and costly tanks.

5.5.7 Super-capacitors

Super- or ultra-capacitors are very high capacity but (relatively) low size capacitors. This is achieved by employing several distinct electrode materials prepared using special processes. Some state-of-the art ultra-capacitors are based on high surface area, ruthenium dioxide (RuO_2) and carbon electrodes. Ruthenium is extremely expensive and available only in very limited amounts.

Electrochemical capacitors are used for high-power applications such as cellular electronics, power conditioning, industrial lasers, medical equipment, and power electronics in conventional, electric and hybrid vehicles. In conventional vehicles, ultra-capacitors could be used to reduce the need for large alternators for meeting intermittent high peak power demands related to power steering and braking. Ultra-capacitors recover braking energy dissipated as heat and can be used to reduce losses in electric power steering.

One system in use on a hybrid bus uses 30 ultra-capacitors to store 1600 kJ of electrical energy (20 farads at 400 V). The capacitor bank has a mass of 950 kg. Use of this technology allows recovery of energy, such as when braking, that would otherwise have been lost. The capacitors can be charged in a very short space of time. The energy in the capacitors can also be used very quickly, such as for rapid acceleration.

The FIA (governing body for Formula 1) allowed the use of a hybrid drive in 2009, which uses 'superbatteries' made with both batteries and super-capacitors. The system became known as the kinetic energy recovery system (KERS).

Key fact

Capacitors can be charged in a very short space of time (compared to batteries).

5.5.8 Summary

As a summary to this section, Table 5.7 compares the potential energy density of several types of battery. Wh/kg means watt-hours per kilogram. This is a measure of the power it will supply, and for how long, per kilogram.

Some super capacitors can have an energy density of several kWh/kg but are usually used for short term storage and rapid release.

Table 5.7 Voltages and energy densities of batteries and storage devices

Battery type	Nominal open circuit cell voltage (V)	Energy density (Wh/kg)	Specific power (W/kg)
Ambient temperature operation			
Lead-acid	2.1	35	200
Nickel-cadmium	1.35	40	260
Nickel-iron	1.4	30	100
Nickel-hydride	1.5	55	100
Nickel-metal-hydride	1.35	70	200
Nickel-zinc	1.73	60	300
Lithium-ion	4.1	110	
Lithium-polymer electrolyte	4.23	200	160
High temperature operation			
Sodium-sulphur	2.08	100	130
Sodium-nickel-chloride (ZEBRA)	2.58	100	150
Lithium-iron-sulphur	1.33	130	240
Fuel cells (produce high temperatures)			
Hydrogen fuel cell	0.3 to 0.9 (1.23 o/c)	400	650
Direct methanol fuel cell (DMFC)	0.3 to 0.9 (1.23 o/c)	1400	100–500
Capacitors			
Classic capacitor	E.g. 100–300	0.6	100 000
Super capacitor	E.g. 100–300	6	50 000

Sources: Mindl 2003, Steckmann 2009, Flipsen 2006. Note that energy density includes containment and packaging and that some figures are estimated for comparison purposes.

Figure 5.25 Lithium-ion battery (Source: Bosch Media)

CHAPTER 6

Charging

6.1 Requirements of the charging system

6.1.1 Introduction

The ‘current’ demands made by modern vehicles are considerable. The charging system must be able to meet these demands under all operating conditions and still ‘fast charge’ the battery.

The main component of the charging system is the alternator and on most modern vehicles – with the exception of its associated wiring – this is the only component in the charging system. Figure 6.1 shows an alternator in common use. The alternator generates AC but must produce DC at its output terminal as only DC can be used to charge the battery and run electronic circuits. The output of the alternator must be a constant voltage regardless of engine speed and current load.

To summarize, the charging system must meet the following criteria (when the engine is running).

- Supply the current demands made by all loads.
- Supply whatever charge current the battery demands.
- Operate at idle speed.
- Supply constant voltage under all conditions.
- Have an efficient power-to-weight ratio.
- Be reliable, quiet, and have resistance to contamination.
- Require low maintenance.
- Provide an indication of correct operation.

Key fact

The charging system must be able to supply all the required current to the electrical system and still ‘fast charge’ the battery.

Figure 6.1 SRX Alternator (Source: GM Media)

6.1.2 Basic operating principles

A generator, or alternator, is a machine that converts mechanical energy from the engine into electrical energy. The basic principle of an alternator is a magnet (the rotor) rotating inside stationary loops of wire (the stator). Electromagnetic induction caused by the rotating magnet produces an electromotive force in the stator windings.

In order for the output of the alternator to charge the battery and run other vehicle components, it must be converted from alternating current to direct current. The component most suitable for this task is the silicon diode. In order to full-wave rectify the output of a three-phase machine six diodes are needed. These are connected in the form of a bridge in a rectifier pack. Many rectifiers now include two extra diodes that pick up extra power from a centre connection to the stator.

Figure 6.2 Alternator on a vehicle

Key fact

A regulator, which controls rotor magnetic field strength, is used to control the output voltage of an alternator as engine speed and current demand change.

A regulator, which controls rotor magnetic field strength, is used to control the output voltage of an alternator as engine speed and current demand change.

Manufacturers strive to produce ever more efficient machines. A modern alternator's high performance and efficiency are achieved primarily by a very dense winding of the copper wire in the stator grooves. To do so, the wires are first wound onto a flat stator core, which is easier to access, after which it is then bent into the usual rounded form.

6.1.3 Vehicle electrical loads

The loads placed on an alternator can be considered as falling under three separate headings: continuous, prolonged and intermittent. The charging system of a modern vehicle has to cope with high demands under many varied conditions. To give some indication as to the output that may be required, consider the power used by each individual component and add this total to the power required to charge the battery. Table 6.1 lists the typical power requirements of various vehicle systems. The current draw (to the nearest 0.5 A) at 14 and 28 V (nominal; alternator output voltages for 12 and 24 V systems) is also given for comparison.

Figure 6.3 shows how the demands on the alternator have increased over the years, together with a prediction of the future.

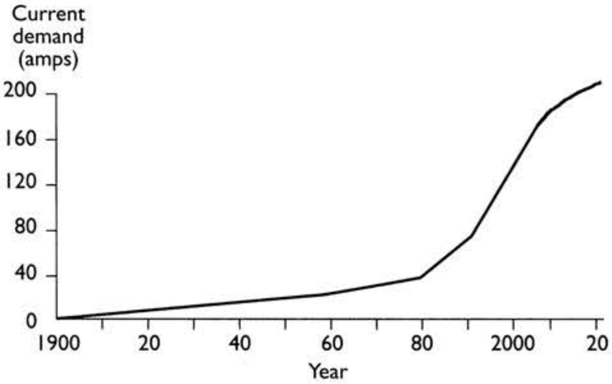


Figure 6.3 How the demands on the alternator have changed

Table 6.1 Typical power requirements of some common vehicle electrical components

Continuous loads	Power	Current at 14 V	28 V
Ignition	30 W	2.0	1.0
Fuel injection	70 W	5.0	2.5
Fuel pump	70 W	5.0	2.5
Instruments	10 W	1.0	0.5
Total	180 W	13.0	6.5
Prolonged loads			
Side and tail lights	30 W	2.0	1.0
Number plate lights	10 W	1.0	0.5
Headlights main beam	200 W	15.0	7.0
Headlights dip beam	160 W	12.0	6.0
Dashboard lights	25 W	2.0	1.0
Radio / Cassette / CD	15 W	1.0	0.5
Total (Av. main & dip)	260 W	18.5	9.5
Intermittent loads	**Power**	**Current at 14 V**	**28 V**
Heater	50 W	3.5	2.0
Indicators	50 W	3.5	2.0
Brake lights	40 W	3.0	1.5
Front wipers	80 W	6.0	3.0
Rear wipers	50 W	3.5	2.0
Electric windows	150 W	11.0	5.5
Radiator cooling fan	150 W	11.0	5.5
Heater blower motor	80 W	6.0	3.0
Heated rear window	120 W	9.0	4.5
Interior lights	10 W	1.0	0.5
Horns	40 W	3.0	1.5
Rear fog lights	40 W	3.0	1.5
Reverse lights	40 W	3.0	1.5
Auxiliary lamps	110 W	8.0	4.0
Cigarette lighter	100 W	7.0	3.5
Headlight wash wipe	100 W	7.0	3.5
Seat movement	150 W	11.0	5.5
Seat heater	200 W	14.0	7.0
Sun-roof motor	150 W	11.0	5.5
Electric mirrors	10 W	1.0	0.5
Total	1.7 kW	125.5	63.5

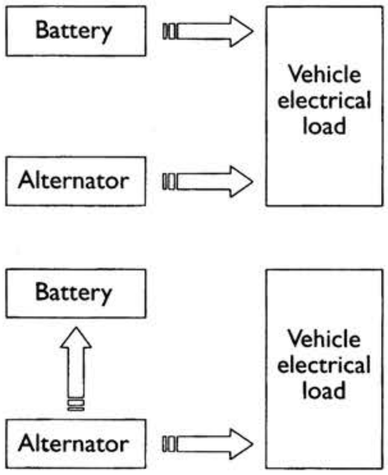


Figure 6.4 Vehicle charging system (the bottom image shows the ideal)

Not shown in Table 6.1 are consumers, such as electrically pre-heated catalytic converters, electrical power assisted steering and heated windscreens, to list just three. Changes will therefore continue to take place in the vehicle electrical system and the charging system will have to keep up!

The intermittent loads are used infrequently and power consumers such as heated rear windows and seat heaters are generally fitted with a timer relay. The factor of 0.1 is therefore applied to the total intermittent power requirement, for the purpose of further calculations. This assumes the vehicle will be used under normal driving conditions.

The consumer demand on the alternator is the sum of the constant loads, the prolonged loads and the intermittent loads (with the factor applied). In this example:

$$180 + 260 + 170 = 610 \text{ W } (43 \text{ A at } 14 \text{ V})$$

The demands placed on the charging system therefore are extensive. This load is in addition to the current required to recharge the battery. Further sections in this chapter discuss how these demands are met.

6.2 Charging system principles

6.2.1 Basic principles

Figure 6.4 shows a representation of the vehicle charging system as three blocks, the alternator, battery and vehicle loads. When the alternator voltage is less than the battery (engine slow or not running for example), the direction of current flow is from the battery to the vehicle loads. The alternator diodes prevent current flowing into the alternator. When the alternator output is greater than the battery voltage, current will flow from the alternator to the vehicle loads and the battery.

From this simple example it is clear that the alternator output voltage must be greater than the battery voltage at all times when the engine is running. The actual voltage used is critical and depends on a number of factors.

6.2.2 Charging voltages

The main consideration for the charging voltage is the battery terminal voltage when fully charged. If the charging system voltage is set to this value then there can be no risk of overcharging the battery. This is known as the constant voltage charging technique. The chapter on batteries discusses this issue in greater detail. The figure of 14.2 ± 0.2 V is the accepted charging voltage for a 12 V system. Commercial vehicles generally employ two batteries in series at a nominal voltage of 24 V, the accepted charge voltage would therefore be doubled. These voltages are used as the standard input for all vehicle loads. For the purpose of clarity the text will just consider a 12 V system.

The other areas for consideration when determining the charging voltage are any expected voltage drops in the charging circuit wiring and the operating temperature of the system and battery. The voltage drops must be kept to a minimum, but it is important to note that the terminal voltage of the alternator may be slightly above that supplied to the battery.

Key fact

Ideally, the alternator output voltage must be greater than the battery voltage at all times when the engine is running.

Definition

The figure of 14.2 ± 0.2 V is the accepted charging voltage for a 12 V system.

6.2.3 Charging circuits

For many applications, the charging circuit is one of the simplest on the vehicle. The main output is connected to the battery via a suitably sized cable (or in some cases two cables to increase reliability and flexibility), and the warning light is connected to an ignition supply on one side and to the alternator terminal at the other. A wire may also be connected to the phase terminal if it is utilized. Figure 6.5 shows two typical wiring circuits. Note that the output of the alternator is often connected to the starter main supply simply for convenience of wiring. If the wires are kept as short as possible this will reduce voltage drop in the circuit. The voltage drop across the main supply wire when the alternator is producing full output current, should be less than 0.5 V.

Key fact

The voltage drop across the main supply wire when the alternator is producing full output current, should be less than 0.5 V.

Some systems have an extra wire from the alternator to ‘sense’ battery voltage directly. An ignition feed may also be found and this is often used to ensure instant excitement of the field windings. A number of vehicles link a wire from the engine management ECU to the alternator. This is used to send a signal to increase engine idle speed if the battery is low on charge.

6.2.4 Generation of electricity

Figure 6.6 shows the basic principle of a three-phase alternator together with a representation of its output. Electromagnetic induction is caused by a rotating magnet inside a stationary loop or loops of wire. In a practical alternator, the rotating magnet is an electromagnet that is supplied via two

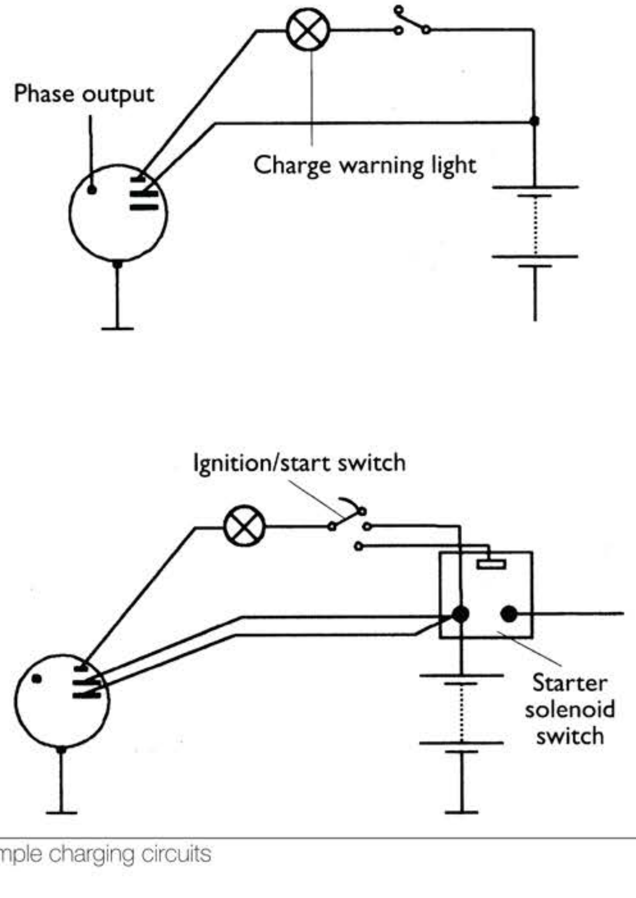


Figure 6.5 Example charging circuits

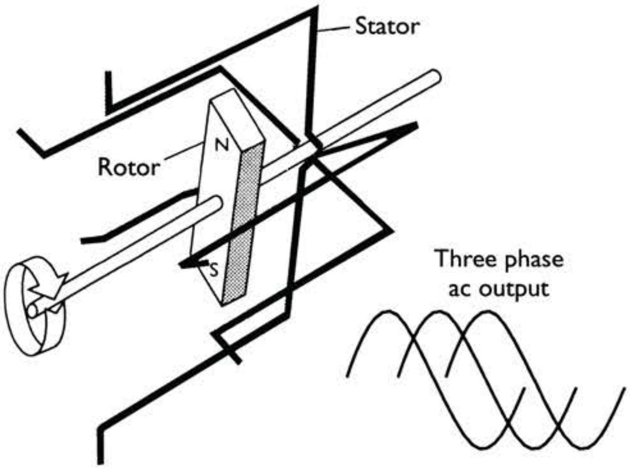


Figure 6.6 Principle of a three-phase alternator

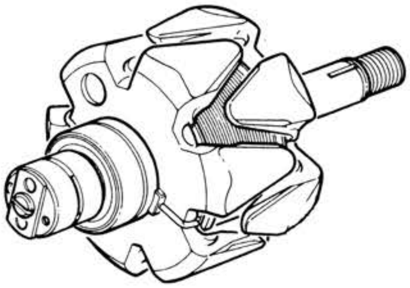


Figure 6.7 Rotor

Key fact

The three-phase windings of the stator can be connected in two ways, known as star or delta.

slip rings. Figure 6.7 shows the most common design, which is known as a claw pole rotor. Each end of the rotor will become a north or south pole and hence each claw will be alternately north and south. It is common practice, due to reasons of efficiency, to use claw pole rotors with 12 or 16 poles.

The stationary loops of wire are known as the stator and consist of three separate phases, each with a number of windings. The windings are mechanically spaced on a laminated core (to reduce eddy currents), and must be matched to the number of poles on the rotor. Figure 6.8 shows a typical example.

The three-phase windings of the stator can be connected in two ways, known as star or delta windings – as shown in Figure 6.9. The current and voltage output characteristics are different for star-and delta-wound stators.

Star connection can be thought of as a type of series connection of the phases and, to this end, the output voltage across any two phases will be the vector sum of the phase voltages. Current output will be the same as the phase current. Star-wound stators therefore produce a higher voltage, whereas delta-wound

Figure 6.8 Stator and other alternator components (Source: Bosch Media)

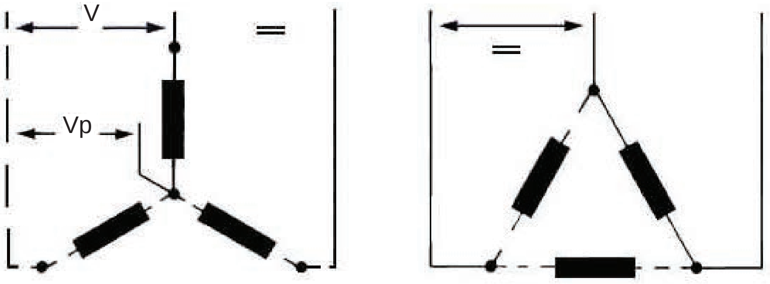


Figure 6.9 Star and delta stator windings

stators produce a higher current. The voltage and current in three-phase stators can be calculated as follows.

Star-wound stators can be thought of as a type of series circuit.

$$V = Vp\sqrt{3}$$

$$I = IP$$

Delta connection can be similarly thought of as a type of parallel circuit. This means that the output voltage is the same as the phase voltage but the output current is the vector sum of the phase currents.

$$V = VP$$

$$I = Ip\sqrt{3}$$

where: V = output voltage, V_p = phase voltage, I = output current, I_p = phase current.

Most vehicle alternators use the star windings but some heavy-duty machines have taken advantage of the higher current output of the delta windings. The majority of modern alternators using star windings incorporate an eight-diode rectifier as to maximise output. This is discussed in a later section.

The frequency of an alternator output can be calculated. This is particularly important if an AC tapping from the stator is used to run a vehicle rev-counter.

$$f = \frac{pn}{60}$$

where: f = frequency in Hz, n = alternator speed in rpm, p = number of pole pairs (a 12 claw rotor has 6 pole pairs).

An alternator when the engine is at idle, will have a speed of about 2000 rpm, which, with a 12 claw rotor will produce a frequency of 6 x 2000/60 = 200 Hz.

A terminal provided on many alternators for this output, is often marked W. The output is half wave rectified and is used in particular on diesel engine to drive a revcounter. It is also used on some earlier petrol-engine applications to drive an electric choke.

6.2.5 Rectification of AC to DC

In order for the output of the alternator to charge the battery and run other vehicle components it must be converted from alternating current (AC) to direct current (DC). The component most suitable for this task is the silicon diode.
If single-phase AC is passed through a diode, its output is half-wave rectified as shown in Figure 6.10. In this example, the diode will only allow the positive half

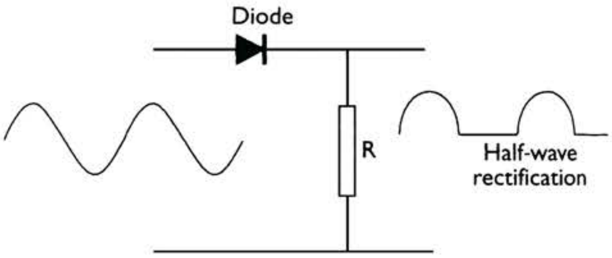


Figure 6.10 Half wave rectification

cycles to be conducted towards the positive of the battery. The negative cycles are blocked.

Figure 6.11 shows a four-diode bridge rectifier to full-wave rectify single phase AC. A diode is often considered to be a one-way valve for electricity. While this is a good analogy it is important to remember that while a good quality diode will block reverse flow up to a pressure of about 400 V, it will still require a small voltage pressure of about 0.6 V to conduct in the forward direction.

Key fact

In order to full-wave rectify the output of a three-phase machine, six diodes are required.

In order to full-wave rectify the output of a three-phase machine, six diodes are required. These are connected in the form of a bridge, as shown in Figure 6.12.

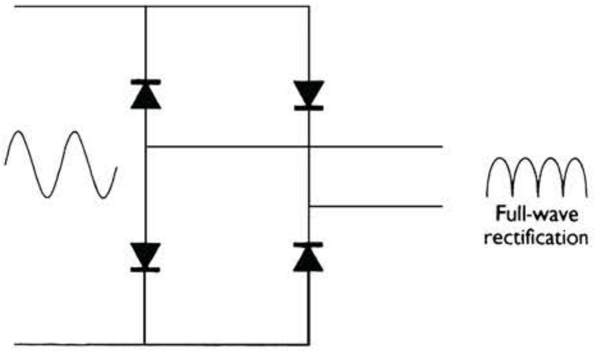


Figure 6.11 Full wave bridge rectifier (single phase)

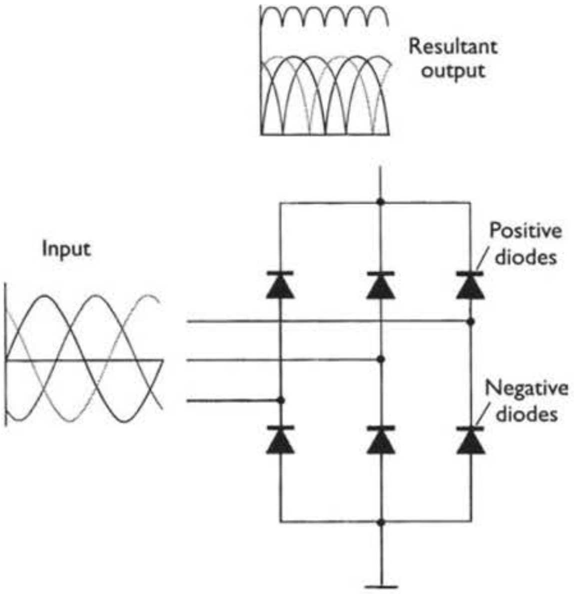


Figure 6.12 Three-phase bridge rectifier

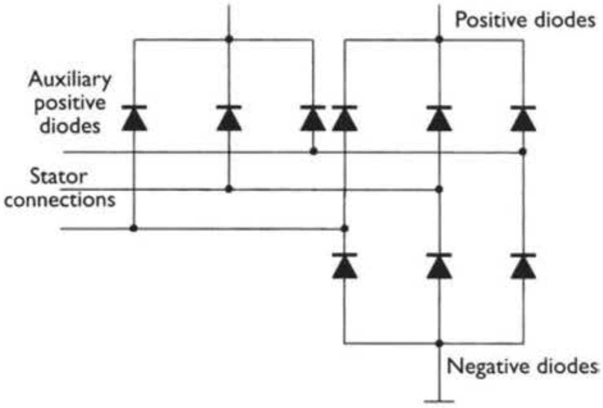


Figure 6.13 Nine-diode rectifier

The 'bridge' consists of three positive diodes and three negative diodes. The output produced by this configuration is shown compared with the three-phase signals.

A further three positive diodes are often included in a rectifier pack. These are usually smaller than the main diodes and are only used to supply a small current back to the field windings in the rotor. The extra diodes are known as the auxiliary, field or excitation diodes. Figure 6.13 shows the layout of a nine-diode rectifier.

Owing to the considerable currents flowing through the main diodes, some form of heat sink is required to prevent thermal damage. In some cases diodes are connected in parallel to carry higher currents without damage. Diodes in the rectifier pack also serve to prevent reverse current flow from the battery to the alternator. This also allows alternators to be run in parallel without balancing, as equalizing current cannot flow from one to the other. Figure 6.14 shows an example of a rectifier pack.

When a star-wound stator is used, the addition of the voltages at the neutral point of the star is, in theory, 0 V. In practice, however, due to slight inaccuracies in the construction of the stator and rotor, a potential develops at this point. This potential (voltage) is known as the third harmonic and is shown in Figure 6.15. Its frequency is three times the fundamental frequency of the phase windings. By employing two extra diodes, one positive and one negative connected to the star point, the energy can be collected. This can increase the power output of an alternator by up to 15%.

Key fact

By fitting two extra diodes, one positive and one negative connected to the star point, the third harmonic energy can be collected.

Figure 6.14 Rectifier pack and stator

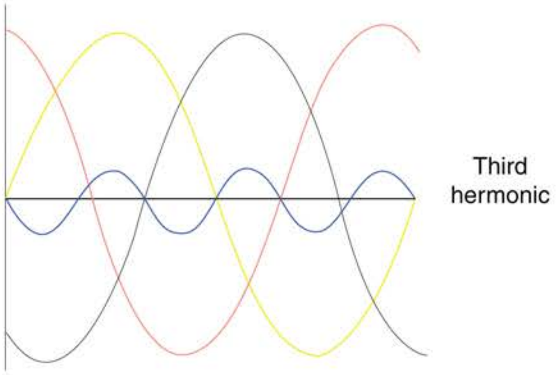


Figure 6.15 The third harmonic

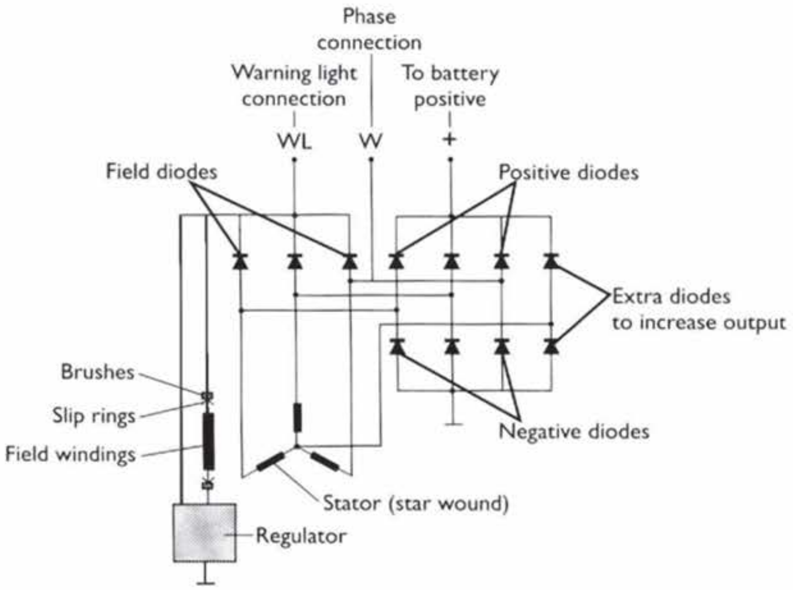


Figure 6.16 Complete internal alternator circuit

Figure 6.16 shows the full circuit of an alternator using an eight-diode main rectifier and three field diodes. The voltage regulator, which forms the starting point for the next section, is also shown in this diagram. The warning light in an alternator circuit, in addition to its function of warning of charging faults, also acts to supply the initial excitation to the field windings. An alternator will not always self-excite as the residual magnetism in the fields is not usually enough to produce a voltage that will overcome the 0.6 or 0.7 V needed to forward bias the rectifier diodes. A typical wattage for the warning light bulb is 2 W. Many manufacturers also connect a resistor in parallel with the bulb to assist in excitation and allow operation if the bulb blows. The charge warning light bulb is extinguished when the alternator produces an output from the field diodes as this causes both sides of the bulb to take on the same voltage (a potential difference across the bulb of 0 V).

6.2.6 Regulation of output voltage

To prevent the vehicle battery from being overcharged the regulated system voltage should be kept below the gassing voltage of the lead-acid battery. A figure of 14.2 ± 0.2 V is used for all 12 V charging systems. Accurate voltage

Figure 6.17 Voltage regulators

control is vital with the ever-increasing use of electronic systems. It has also enabled the wider use of sealed batteries, as the possibility of over-charging is minimal. Figure 6.17 shows two common voltage regulators. Voltage regulation is a difficult task on a vehicle alternator because of the constantly changing engine speed and loads on the alternator. The output of an alternator without regulation would rise linearly in proportion with engine speed. Alternator output is also proportional to magnetic field strength and this, in turn, is proportional to the field current. It is the task of the regulator to control this field current in response to alternator output voltage.

Key fact

The output of an alternator without regulation would rise linearly in proportion with engine speed.

Figure 6.18 shows a flow chart which represents the action of the regulator, showing how the field current is switched off as output voltage increases and then back on again as output voltage falls. The abrupt switching of the field current does not cause abrupt changes in output voltage due to the very high inductance of the field (rotor) windings. In addition, the whole switching process only takes a few milliseconds. Many regulators also incorporate some temperature compensation to allow a higher charge rate in colder conditions and to reduce the rate in hot conditions.

When working with regulator circuits, care must be taken to note 'where' the field circuit is interrupted. For example, some alternator circuits supply a constant

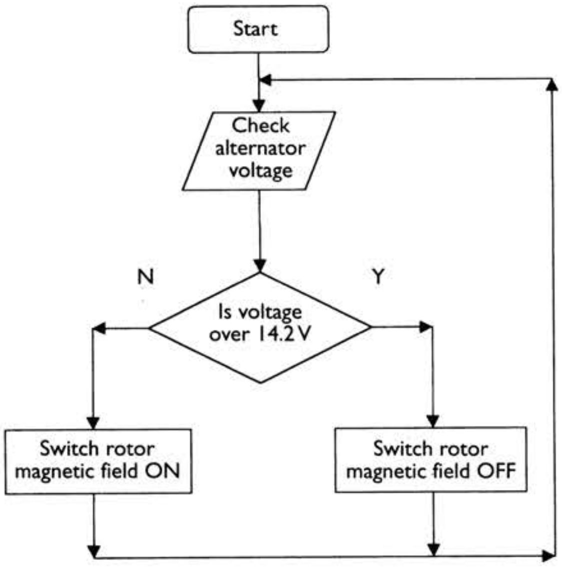


Figure 6.18 Action of the voltage regulator

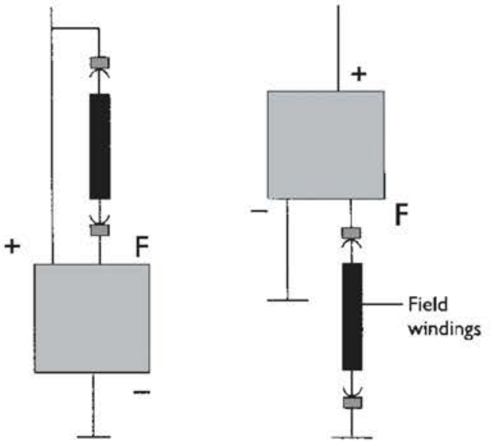


Figure 6.19 How the voltage regulator is incorporated in the supply or earth/ground of the field circuit.

feed to the field windings from the excitation diodes and the regulator switches the earth side. In other systems, one side of the field windings is constantly earthed and the regulator switches the supply side. Figure 6.19 shows these two methods.

Key fact

Alternators do not require any extra form of current regulation.

Alternators do not require any extra form of current regulation. This is because if the output voltage is regulated the voltage supplied to the field windings cannot exceed the pre-set level. This in turn will only allow a certain current to flow due to the resistance of the windings and hence a limit is set for the field strength. This will then limit the maximum current the alternator can produce.

Regulators can be mechanical or electronic, and the latter are now universal on modern cars. The mechanical type uses a winding connected across the output of the alternator. The magnetism produced in this winding is proportional to the output voltage. A set of normally closed contacts is attached to an armature, which is held in position by a spring.

The supply to the field windings is via these contacts. When the output voltage rises beyond a pre-set level, say 14 V, the magnetism in the regulator winding will overcome spring tension and open the contacts. This switches off the field current and causes the alternator output to fall. As the output falls below a pre-set level, the spring will close the regulator contacts again and so the process continues. Figure 6.20 shows a simplified circuit of a mechanical regulator. This principle has not changed from the very early voltage control of dynamo output.

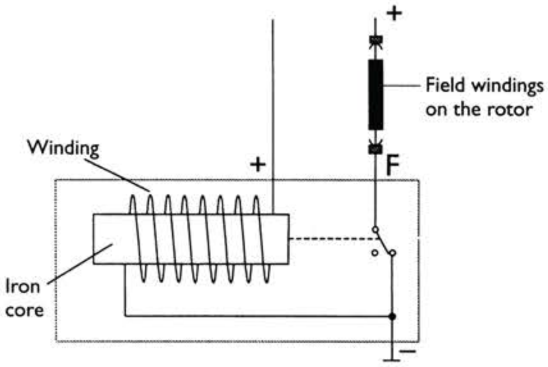


Figure 6.20 Mechanical regulator principle

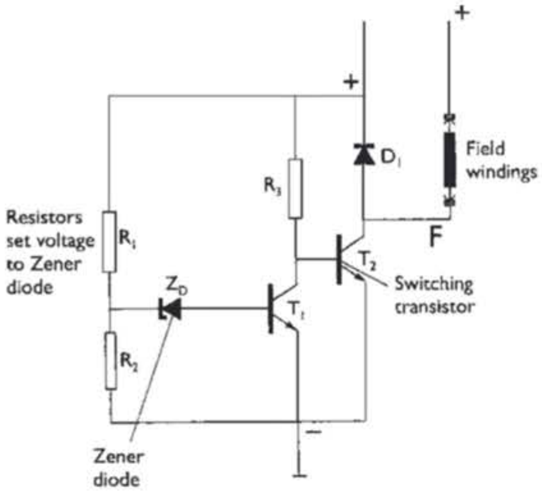


Figure 6.21 Electronic voltage regulator

The problem with mechanical regulators is the wear on the contacts and other moving parts. This has been overcome with the use of electronic regulators which, due to more accurate tolerances and much faster switching, are far superior, producing a more stable output. Due to the compactness and vibration resistance of electronic regulators they are now fitted almost universally on the alternator, reducing the number of connecting cables required.

Key fact

The key to electronic voltage regulation is the Zener diode.

The key to electronic voltage regulation is the Zener diode. As discussed in Chapter 2, this diode can be constructed to break down and conduct in the reverse direction at a precise level. This is used as the sensing element in an electronic regulator. Figure 6.21 shows a simplified electronic voltage regulator.

This regulator operates as follows: When the alternator first increases in speed the output will be below the pre-set level. Under these circumstances transistor T_2 will be switched on by a feed its base via Resistor R_3. This allows full field current to flow thus increasing voltage output. When the pre-set voltage is reached the Zener diode will conduct. Resistors R_1 and R_2 are a simple series circuit to set the voltage appropriate to the value of the diode when the supply is say 14.2 V. Once Z_D conducts transistor T_1 will switch on and pull the base of T_2 down to ground. This switches T_2 off and so the field current is interrupted causing output voltage to fall. This will cause Z_D to stop conducting, T1 will switch off allowing T_2 to switch back on and so the cycle will continue. The conventional diode D_1 is to absorb the back EMF from the field windings damaging the other components.

Electronic regulators can be made to sense either the battery voltage, the machine voltage (alternator), or a combination of the two. Most systems in use at present tend to be machine sensed as this offers some protection against over-voltage in the event of the alternator being driven with the battery disconnected.

Figure 6.22 shows the circuit of a hybrid integrated circuit (IC) voltage regulator. The hybrid system involves the connection of discrete components on a ceramic plate using film techniques. The main part of the regulator is an integrated circuit containing the sensing elements and temperature compensation components. The IC controls an output stage such as a Darlington pair. This technique produces a very compact device and, because of the low number of components and connections, is very reliable.

Figure 6.23 is a graph showing how the IC regulator response changes with temperature. This change is important to ensure correct charging under 'summer'

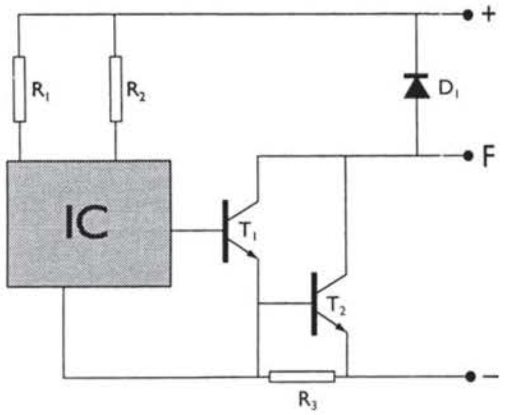


Figure 6.22 Hybrid IC regulator circuit

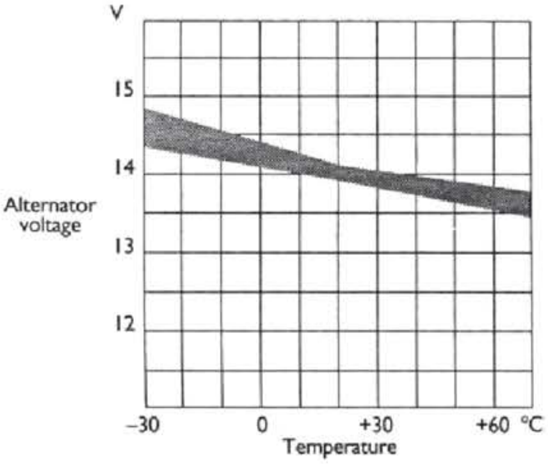


Figure 6.23 How the regulator response changed with temperature

and 'winter' conditions. When a battery is cold, the electrolyte resistance increases. This means a higher voltage is necessary to cause the correct recharging current.

Over-voltage protection is required in some applications in order to prevent damage to electronic components. When an alternator is connected to a vehicle battery system, the voltage, even in the event of regulator failure, will not often exceed about 20 V due to the low resistance and swamping effect of the battery. If an alternator is run with the battery disconnected (which is not recommended), a heavy duty Zener diode connected across the output of the WL/field diodes will offer some protection as, if the system voltage exceeds its breakdown figure, it will conduct and cause the system voltage to be kept within reasonable limits.

6.3 Alternators

6.3.1 Bosch compact alternator

The Bosch compact alternator is becoming very popular with a number of European manufacturers and others. Figure 6.24 shows a cut-away picture of this machine.

Figure 6.24 Bosch compact alternator

The key points are as follows:

- 20–70% more power than conventional units.
- 15–35% better power-to-weight ratio.
- Maximum speed up to 20 000 rpm.
- Twin interior cooling fans.
- Precision construction for reduced noise.
- Versions available: 70, 90 and up to 170 A.

The compact alternator follows the well-known claw pole design. Particular enhancements have been made to the magnetic circuit of the rotor and stator. This was achieved by means of modern ‘field calculation’ programmes. The optimization reduces the iron losses and hence increases efficiency.

A new monolithic circuit regulator is used that reduces the voltage drop across the main power transistor from 1.2 V to 0.6 V. This allows a greater field current to flow, which again will improve efficiency.

The top speed of an alternator is critical as it determines the pulley ratio between the engine and alternator. The main components affected by increased speed are the ball bearings and the slip rings. The bearings have been replaced with a type that uses a plastic cage instead of the conventional metal type. Higher melting point grease is also used. The slip rings are now mounted outside the two bearings and therefore the diameter is not restricted by the shaft size. Smaller diameter slip rings give a much lower peripheral velocity, and thus greater shaft speed can be tolerated.

Key fact

The top speed of an alternator is critical as it determines the pulley ratio between the engine and alternator.

Increased output results in increased temperature so a better cooling system was needed. The machine uses twin internal asymmetric fans, which pull air through central slots front and rear, and push it out radially through the drive and slip ring end brackets over the stator winding heads.

High vibration is a problem with alternators as with all engine mounted components. Cars with four-valve engines can produce very high levels of vibration. The alternator is designed to withstand up to 80 g. New designs are thus required for the mounting brackets.

Key fact

High vibration is a problem with alternators as with all engine mounted components.

Figure 6.25 Alternator and stator

6.3.2 Efficient alternators

A combination of experience, modern development methods and innovative production processes has enabled development engineers to achieve dramatic gains in alternator performance compared to conventional models – a 35% increase in power density to 1.43 W/cc, a rise in maximum operating temperature from 105 °C to 120 °C, and an increase in the maximum degree of efficiency to 76% (VDA average 72%). An example is shown as Figures 6.25 and 6.26.

The improved performance parameters offer automobile manufacturers a reduction in fuel consumption of up to 0.2 litres per hundred kilometres, a saving in space of up to 400 cubic centimetres, and finally an increase in power output of as much as 1 kilowatt. The improvements are largely due to the so-called ‘Flat Pack’ technique, which achieves a very high density of the copper wires in the stator windings.

Bosch supplies its 14 V LI-X alternators in three different sizes: ‘Compact’, ‘Medium’ and ‘High Line’, with outputs ranging from 1.9 to 3.8 kilowatts. The model range is designed to be extremely flexible and the power outputs can easily be adjusted for use in both diesel and gasoline engines. Bosch is also planning a 42 volt version with a peak power output of 4 kilowatts.

The alternator regulator is multifunctional and can be operated through a variety of interfaces (for smart charging, etc.), such as BSS, LIN or RVC, in line with the manufacturer’s preference.

The company recently started production of a new, enhanced-efficiency generator series. An improved electrical configuration and optimized materials have enabled the manufacturer’s engineers to increase the efficiency of all variants to over 70% (according to VDA). The Efficiency Line (EL) generators consequently need less mechanical power to generate the energy for the vehicle’s electrical consumers. This reduces fuel requirements – and therefore CO_2 emissions, too.

Figure 6.26 LI-X alternator (Source: Bosch Media)

The demand for electricity in cars has been rising constantly in recent years due to the increasing number of comfort and safety systems in vehicles. As a result, more and more attention is being devoted to the efficiency of all consumption components and the generator. With ultra-efficient diodes, which are available as an option, the efficiency of the new generators can even be improved by as much as 77%.

At engine idle, which for this device is a generator speed of about 1800 rpm, the EL generators produce about 10% more power. They are therefore an ideal complement to start-stop systems, as they ensure the fastest possible battery charging. As a consequence, the start-stop function can be used more frequently.

An additional analogue or digital communications interface such as local interconnected network (LIN) creates the conditions for intelligent generator regulation. This in turn allows most of the electricity to be generated in coasting mode, which in turn means a fuel saving of up to 2%. The EL series is available in various sizes and covers a performance range of 130 to 210 A (at 6000 rpm). Designed for especially quiet operation, they can also be used at high ambient temperatures of up to 125 °C.

Key fact

The demand for electricity in cars has been rising constantly in recent years due to the increasing number of comfort and safety systems in vehicles.

Figure 6.27 EL Alternator

6.3.3 Water-cooled alternators

An interesting technique involves running the engine coolant through the alternator. A 120–190 A output range is available. Compared with conventional air-cooled alternators the performance of these new machines has been enhanced more particularly in the following areas:

- Improved efficiency (10–25%).
- Increased output at engine idle speed.
- Noise reduction (10–12 dB due to fan elimination).
- Resistance to corrosion (machine is enclosed).
- Resistance to high ambient temperature (130 °C).

Additional heating elements can be integrated into the alternator to form a system that donates an additional 2–3 kW to the coolant, enabling faster engine

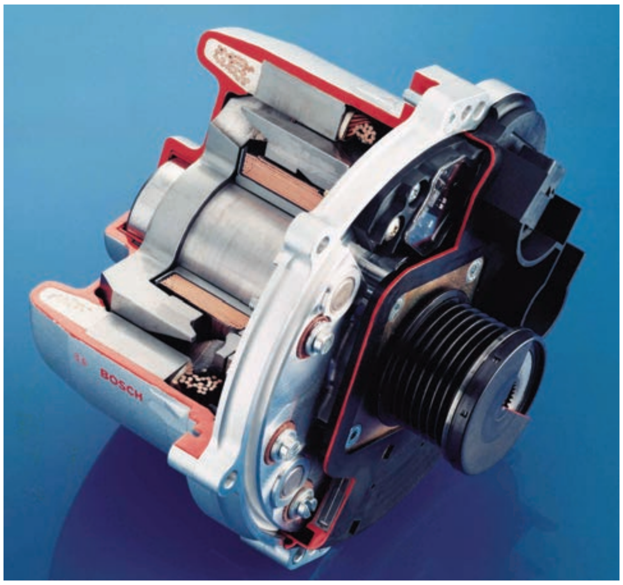


Figure 6.28 Water-cooled alternator (Source: Bosch Media)

Figure 6.29 Denso 220 A alternator (Source: Denso Media)

warm up after a cold start. This contributes to reduced pollution and increased driver comfort.

6.3.4 Denso high-output alternators

The Denso Corporation has developed alternators that provide output currents of 165, 180, 200 and 220 amps; higher than the company's traditional alternators, which supply up to 150 amps. These new products were the smallest and lightest in the world for their output.

In 2000, Denso developed the world's first SC (segment conductor) alternator using a rectangular conductor for its stator coil, reducing coil resistance by 50%. In addition to the rectangular conductor, the SC alternator adopted dual windings and rectifiers, achieving smaller size, lighter weight, higher efficiency, and lower noise.

To develop the smaller, higher power alternators, they further improved the stator coil connection method of the SC alternator. To solve the problem of higher heat production due to higher output, the surface area of the rectifier cooling fins was increased to nearly twice the size of conventional fins, thereby improving the cooling ability of the rectifiers.

6.3.5 Charging system testing procedure

After connecting a voltmeter across the battery and an ammeter in series with the alternator output wire(s), as shown in Figure 6.30.

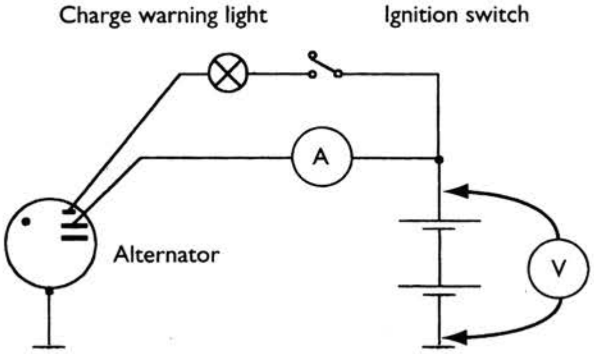
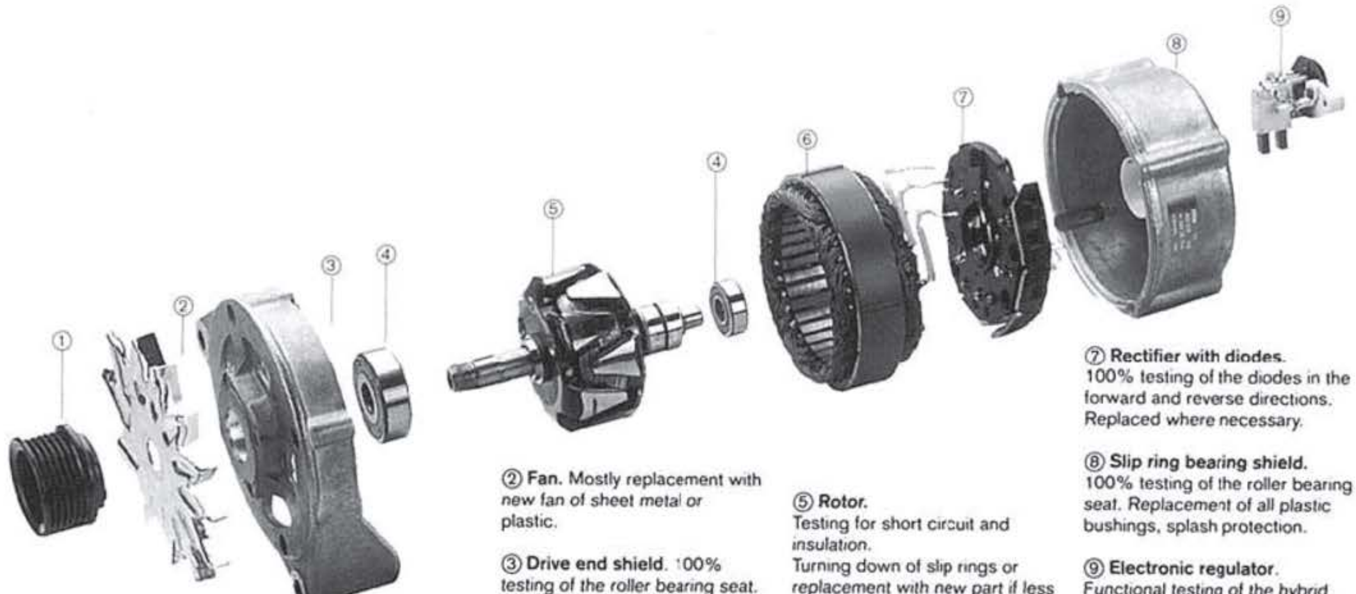


Figure 6.30 Alternator testing



① **Pulley.** Examination of the running surfaces of the pulley for wear and dents.

② **Fan.** Mostly replacement with new fan of sheet metal or plastic.

③ **Drive end shield.** 100% testing of the roller bearing seat.

④ **Roller bearings.** 100% replacement. Lubrication to original equipment specification.

⑤ **Rotor.** Testing for short circuit and insulation. Turning down of slip rings or replacement with new part if less than minimum acceptable size.

⑥ **Stator.** 100% testing for short circuit and insulation.

⑦ **Rectifier with diodes.** 100% testing of the diodes in the forward and reverse directions. Replaced where necessary.

⑧ **Slip ring bearing shield.** 100% testing of the roller bearing seat. Replacement of all plastic bushings, splash protection.

⑨ **Electronic regulator.** Functional testing of the hybrid regulator. 100% replacement of other regulator types. Replacement of carbon brush where less than the minimum acceptable size.

Figure 6.31 Alternator overhaul procedure (Bosch)

The process of checking the charging system operation is as follows.

1. Hand and eye checks (drive belt and other obvious faults) – belt at correct tension, all connections clean and tight.
2. Check battery (see Chapter 5) – must be 70% charged.
3. Measure supply voltages to alternator – battery volts.
4. Maximum output current (discharge battery slightly by leaving lights on for a few minutes, leave lights on and start engine) – ammeter should read within about 10% of rated maximum output.
5. Regulated voltage (ammeter reading 10 A or less) – 14.2 ± 0.2 V.
6. Circuit volt drop – 0.5 V maximum.

If the alternator is found to be defective then a quality replacement unit is the normal recommendation. Figure 6.31 explains the procedure used by Bosch to ensure quality exchange units. Repairs are possible but only if the general state of the alternator is good.

Table 6.2 lists some common symptoms of a charging system malfunction together with suggestions for the possible fault.

6.4 Smart charging

6.4.1 Introduction and closed loop regulation

To prevent the vehicle battery from being overcharged the regulated system voltage should be kept below the gassing voltage of the lead-acid battery. A figure of 14.2 ± 0.2 V was traditionally used for all 12 V (nominal) charging systems. Accurate voltage control is vital with the ever-increasing use of

Table 6.2 Charging system symptoms and faults

Symptom	Possible fault
Battery loses charge	Defective battery Slipping alternator drive belt Battery terminals loose or corroded Alternator internal fault (diode open circuit, brushes worn or regulator fault etc.) Open circuit in alternator wiring, either main supply, ignition or sensing wires if fitted Short circuit component causing battery drain even when all switches are off High resistance in the main charging circuit
Charge warning light stays on when engine is running	Slipping or broken alternator drive belt Alternator internal fault (diode open circuit, brushes worn or regulator fault etc.) Loose or broken wiring/connections
Charge warning light does not come on at any time	Alternator internal fault (brushes worn open circuit or regulator fault etc.) Blown warning light bulb Open circuit in warning light circuit

electronic systems. It has also enabled the widespread use of sealed batteries, as the possibility of overcharging is minimal.

Traditionally the regulator base plate or heat sink temperature was used as a reference to estimate battery temperature. This is because the ideal maximum charge rate for a battery varies with its temperature. Further, if the regulator senses a significant change in voltage, a function is employed to quickly recover this to the normal set regulation point. In normal regulators this function is integrated into the regulator itself.

This method of closed loop control (regulator senses the output voltage and increases rotor field strength if the output is low, or decreases it if the output is too high) has worked well – up until now.

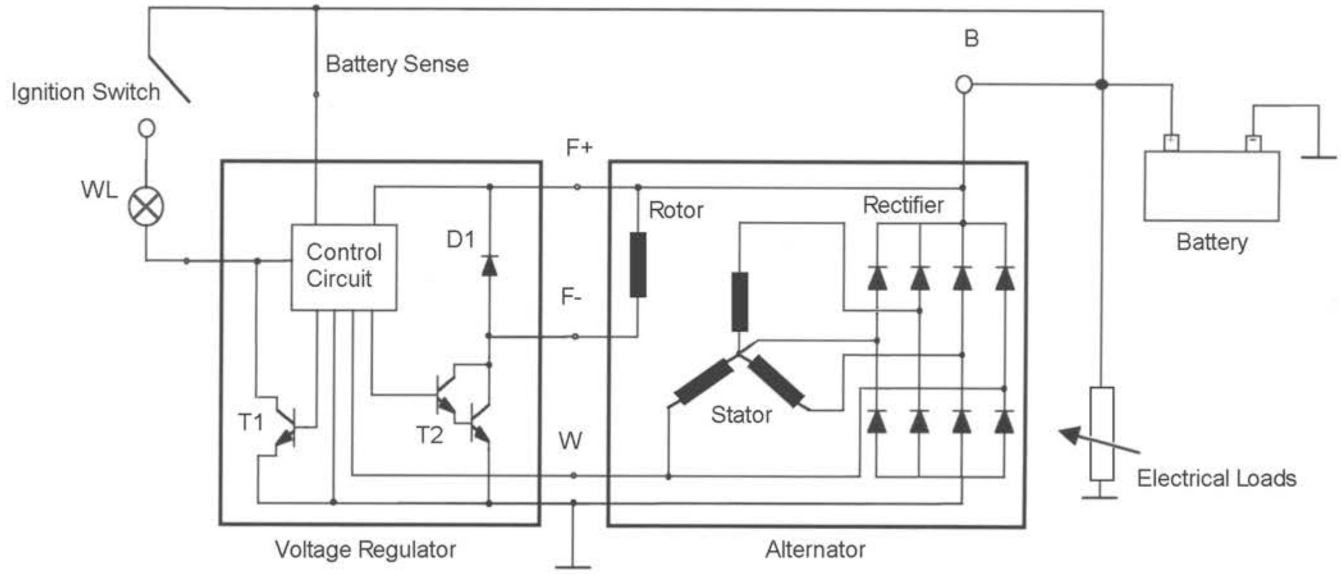


Figure 6.32 Standard closed loop alternator and regulator circuit

6.4.2 Open loop regulation

Some manufacturers are now bringing together alternator output control, electrical power distribution and mechanical power distribution. This is known as intelligent or smart charging.

The principle of open loop control charging is that the alternator regulator and the powertrain control module (PCM) communicate. In simple terms the alternator and the PCM can talk to each other. This allows new features to be developed that benefit the battery and offer other improvements such as:

- Reduced charge times.
- Better idle stability.
- Improved engine performance.
- Increased alternator reliability.
- Better control of electrical load.
- Improved diagnostic functions.

Key fact

The principle of open loop control charging is that the alternator regulator and the powertrain control module (PCM) communicate.

Communication between the regulator and PCM is by signals that are pulse width modulated (PCM). On some systems a LIN bus is used. This signal is used in both directions. It is a constant frequency square wave with a variable on/off ratio or duty cycle.

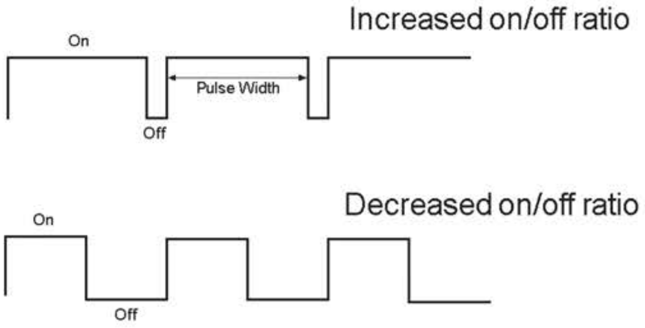


Figure 6.33 Two signals with different pulse width modulation

The PCM determines the set voltage point (regulated voltage) and transmits this to the regulator using a specific duty cycle signal. The regulator responds by transmitting back the field transistor duty cycle (T2 in Figure 6.32, for example). In this way a variety of features can be implemented.

Closed loop regulators estimate the battery temperature based on their own temperature. This does not always result in an accurate figure and hence battery charge rates may not be ideal. With an open loop 'Smart charge' (Figure 6.34) system the PCM can calculate a more accurate figure for battery temperature because it has sensors measuring, for example, coolant temperature, intake air temperature and ambient air temperature. This means a more appropriate charge rate can be set.

Battery recharge times are not only reduced but a significant increase in battery lifetime can be achieved because of this accurate control.

Alternator regulators can also be connected to the local interconnect network (LIN). This is a protocol that allows communication between intelligent actuators and sensors.

6.4.3 Engine performance

Powertrain control modules (PCMs) usually control engine idle speed in two ways. The main method is throttle control, using either a stepper motor or an

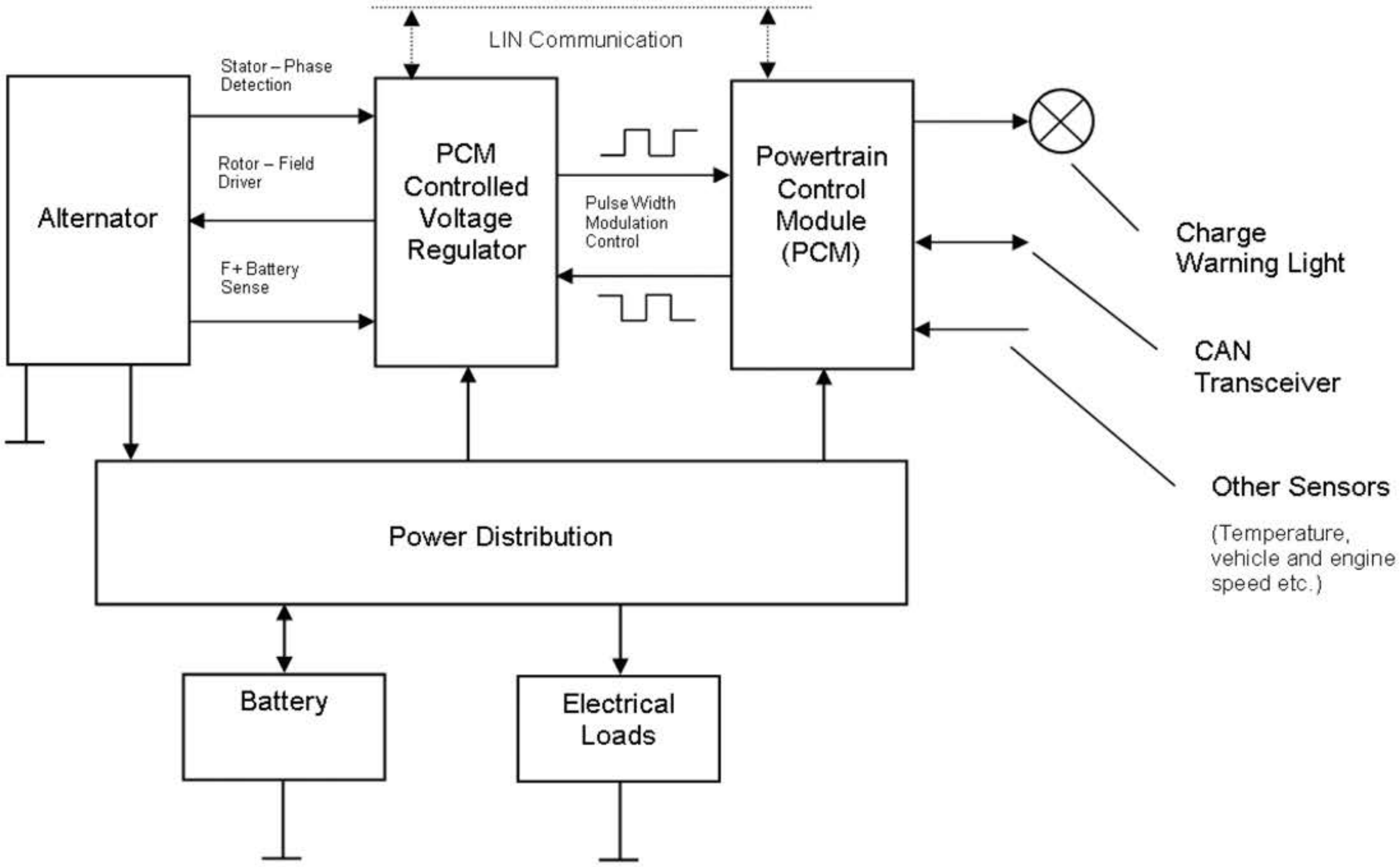


Figure 6.34 Block diagram showing 'Smart charge' system

air bypass valve. This is a good method but can be relatively slow to react. Changes in ignition timing are also used and this results in a good level of control. However, there may be emission implications.

One of the main causes of idle instability is the torque load that the alternator places on the engine. Because a PCM control system is 'aware' of the alternator load, it calculates the corresponding torque load and sets the idle speed accordingly. Overall the idle can be set at a lower value thus reducing fuel consumption and emissions. Equally, when required, the PCM can increase idle speed to increase alternator output and prevent battery drain. This would be likely to occur after a cold start, in the dark, when the screen is frosted over. In these conditions it is likely that, because the driver would switch on lights, interior heaters and screen heaters, there would be an increased electrical load. In addition to the normal electrical load (fuel, ignition, etc.), the battery would also create a high demand after a cold start. The PCM can ensure that it sets an idle speed which results in sufficient alternator output to prevent battery drain.

A dynamic adjustment to the set voltage point is also possible. This may be used during starting to reduce load on the battery. It can also be used during transient engine loads or, in other words, during acceleration. An alternator producing 70 A at 14 V is putting out about 1 kW of power (P = VI). Taking into account the mechanical to electrical energy conversion efficiency of the alternator, the result is a significant torque load on the engine. If the set point (regulated voltage) is reduced during hard acceleration, the 0 to 60 time can be increased by as much as 0.4 s.

Key fact

A dynamic adjustment to the set voltage point is possible during starting and transient loads.

6.4.4 Fault conditions

As well as communicating the load status of the alternator to the PCM, the regulator can also provide diagnostic information. In general the following fault situations can be communicated:

- No communication between regulator and PCM.
- No alternator output due to mechanical fault (drive belt for example).
- Loss of electrical connection to the alternator.
- System over or under voltage due to short or open circuit field driver.
- Failure of rotor or stator windings.
- Failure of a diode.

The PCM can initiate appropriate action in response to these failure conditions, for example, to allow fail-safe operation or at least illuminate the warning light! Suitable test equipment can be used to aid diagnostic work.

6.4.5 Summary

The ability of the alternator regulator and engine control systems to communicate means new possibilities, increased efficiency and improved performance.

New diagnostic equipment may be necessary but new diagnostic techniques certainly are required. However, remember that PWM signals can be examined on a scope or even a duty cycle meter. And, if the voltage you measure across the battery is less than 13 V, it is probably not charging – unless of course you are measuring it during a 0 to 60 acceleration test!

6.5 Advanced charging system technology

6.5.1 Charging system – problems and solutions

The charging system of a vehicle has to cope under many varied conditions. An earlier section gave some indication as to the power output that may be required. Looking at two of the operating conditions that may

The first scenario is the traffic jam, on a cold night, in the rain! This can involve long periods when the engine is just idling, but use of nearly all electrical devices

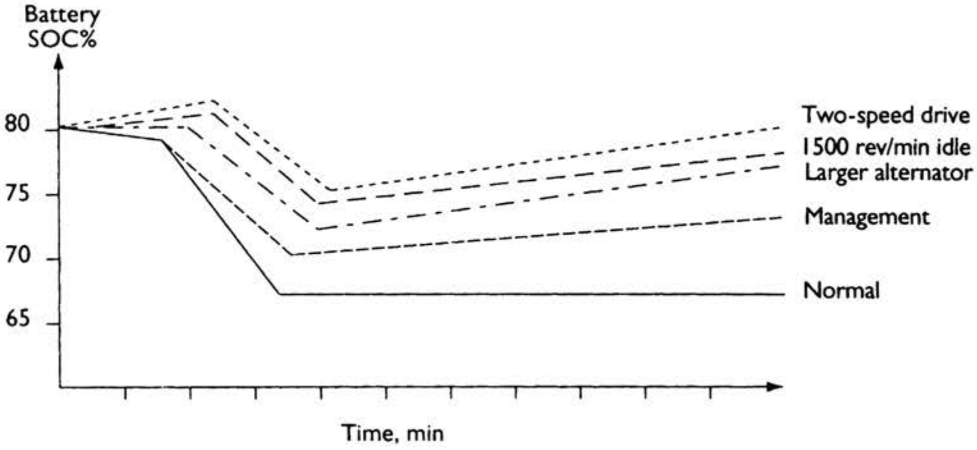


Figure 6.35 Graphical representation comparing various charging techniques when applied to a vehicle used for winter commuting

is still required. The second scenario is that the car has been parked in the open on a frosty night. The engine is started, seat heaters, heated rear window and blower fan are switched on whilst a few minutes are spent scraping the screen and windows. All the lights and wipers are now switched on and a journey of half an hour through busy traffic follows. The seat heaters and heated rear window can generally be assumed to switch off automatically after about 15 minutes.

Tests and simulations have been carried out using the above examples as well as many others. At the end of the first scenario the battery state of charge will be about 35% less than its original level; in the second case the state of charge will be about 10% less. These situations are worst case scenarios, but nonetheless possible. If the situations were repeated without other journeys in between, then the battery would soon be incapable of starting the engine. Combining this with the ever-increasing power demands on the vehicle alternator makes this problem difficult to solve. It is also becoming even more important to ensure the battery remains fully charged, as ECUs with volatile memories and alarm systems make a small but significant drain on the battery when the vehicle is parked.

A number of solutions are available to try and ensure the battery will remain in a state near to full charge at all times. A larger capacity battery could be used to ‘swamp’ variations in electrical use and operating conditions. Some limit, however, has to be set due to the physical size of the battery. Five options for changes to the power supply system are represented graphically in Figure 6.35 and are listed below.

- Fitting a more powerful alternator.
- Power management system.
- Two-stage alternator drive mechanism or increased alternator speed.
- Increased engine idle speed.
- Dual voltage systems.

The options listed above have some things in their favour and some against, not least of which are the technical and economic factors. For the manufacturers, I would predict that a combination of a more powerful alternator, which can be run at a higher speed, together with a higher or dual voltage system, would be the way forward. This is likely to be the most cost effective and technically feasible solution. Each of the suggestions is now discussed in more detail.

The easiest solution to the demand for more power is a larger alternator, and this is, in reality, the only method available as an after-market improvement. It must be remembered, however, that power supplied by an alternator is not ‘free’. For each watt of electrical power produced by the alternator, between 1.5 and 2 W are taken from the engine due to the inefficiency of the energy conversion process. An increase in alternator capacity will also have implications relating to the size of the drive belt, associated pulleys and tensioners.

An intelligent power management system, however, may become more financially attractive as electronic components continue to become cheaper. This technique works by switching off headlights and fog lights when the vehicle is not moving. The cost of this system may be less than increasing the size of the alternator. Figure 6.36 shows the operating principle of this system. A speed sensor signal is used via an electronic processing circuit to trigger a number of relays. The relays can be used to interrupt the chosen lighting circuits. An override switch is provided, for use in exceptional conditions.

A two-speed drive technique which uses a ratio of 5 : 1 for engine speeds under 1200 rpm and usually about 2.5 : 1 at higher speeds shows some promise but

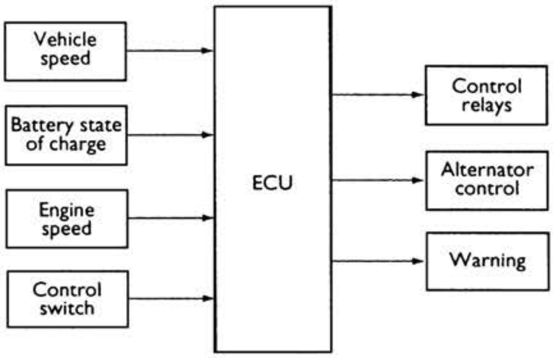


Figure 6.36 Operating principles

adds more complications to the drive system. Due to improvements in design, however, modern alternators are now being produced that are capable of running at speeds up to 20 000 rpm. If the maximum engine speed is considered to be about 6000 rpm, a pulley ratio of about 3.3 : 1 can be used. This will allow the alternator to run as fast as 2300 rpm, even with a low engine idle speed of 700 rpm. The two-speed drive is only at the prototype stage at present.

Increased idle speed may not be practical in view of the potential increase in fuel consumption and emissions. It is nonetheless an option, but may be more suitable for diesel-engined vehicles. Some existing engine management systems, however, are provided with a signal from the alternator when power demand is high. The engine management system can then increase engine idle speed both to prevent stalling and ensure a better alternator output. Figure 6.37 shows the wiring associated with this technique.

Much research has been carried out on dual voltage electrical systems. It has long been known that a 24 V system is better for larger vehicles. This, in the main, is due to the longer lengths of wire used. Double the voltage and the same power can be transmitted at half the current (watts volts amps). This causes less volt drop due to the higher resistance in longer lengths of cable. Wiring harnesses used on passenger cars are becoming increasingly heavy and unmanageable. If a higher supply voltage was used, the cross-section of individual cables could be halved with little or no effect. Because heavy vehicle electrics have been 24 V for a long time, most components (bulbs etc.) are already available if a change in strategy by the vehicle manufacturers takes place. Under discussion is a 12, 0, 12 V technique using three bus bars or rails. High power loads can be connected between 12 and 12 (24 V), and loads which must be supplied by 12 V can be balanced between the 12, 0 and 0, 12 voltage supply rails. A representation of this is shown in Figure 6.38. Note, however,

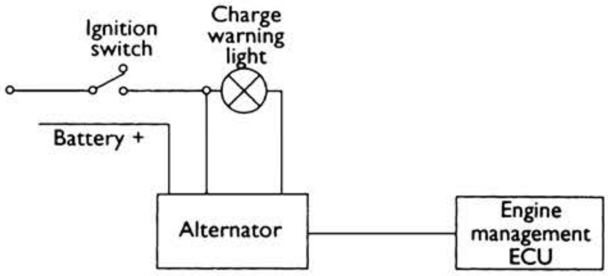


Figure 6.37 Alternator wiring to allow engine management system to sense current demand and control engine idle speed to prevent stalling

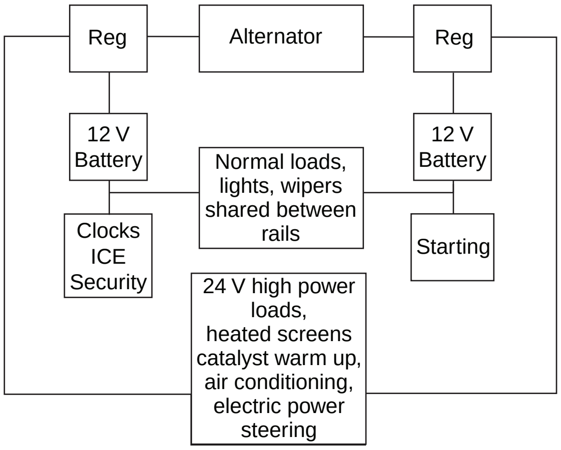


Figure 6.38 Dual rail power supply technique

that running some bulbs (such as for high power headlights) can be a problem because the filament has to be very thin. Some commercial (24 V) vehicles actually use a 12 V supply to the headlights for this reason. To date the dual voltage system has not been used on production vehicles.

6.5.2 Charge balance calculation

The charge balance or energy balance of a charging system is used to ensure that the alternator can cope with all the demands placed on it and still charge the battery. The following steps help to indicate the size of alternator required or to check if the one fitted to a vehicle is suitable.

As a worked example, the figures from Table 6.1 will be used. The calculations relate to a passenger car with a 12 V electrical system. A number of steps are involved.

1. Add the power used by all the continuous and prolonged loads.
2. Total continuous and prolonged power (P1) = 440 W.
3. Calculate the current at 14 V (I = W/V) = 31.5 A.
4. Determine the intermittent power (factored by 0.1) (P2) = 170 W.
5. Total power (P1 + P2) = 610 W.
6. Total current 610/14 = 44 A.

Electrical component manufacturers provide tables to recommend the required alternator, calculated from the total power demand and the battery size. However, as a guide for 12 V passenger cars, the rated output should be about 1.5 times the total current demand (in this example 44 1.5 66 A). Manufacturers produce machines of standard sizes, which in this case would probably mean an alternator rated at 70 A. In the case of vehicles with larger batteries and starters, such as for diesel-powered engines and commercial vehicles, a larger output alternator may be required.

The final check is to ensure that the alternator output at idle is large enough to supply all continuous and prolonged loads (P1) and still charge the battery. Again the factor of 1.5 can be applied. In this example the alternator should be able to supply (31.5 x 1.5) = 47 A, at engine idle. On normal systems this relates to an alternator speed of about 2000 rpm (or less). This can be checked against the characteristic curve of the alternator.

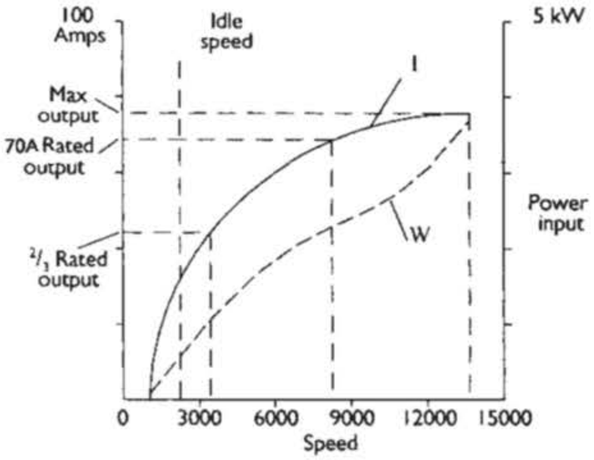


Figure 6.39 Typical alternator characteristic curve

6.5.3 Alternator characteristics

Alternator manufacturers supply 'characteristic curves' for their alternators. These show the properties of the alternator under different conditions. The curves are plotted as output current (at stabilized voltage), against alternator rpm and input power against input rpm. Figure 6.39 shows a typical alternator characteristic curve.

It is common to mark the following points on the graph.

- Cut in speed.
- Idle speed range.
- Speed at which 2/3 of rated output is reached.
- Rated output speed.
- Maximum speed.
- Idle current output range.
- Current 2/3 of rated output.
- Rated output.
- Maximum output.

The graphs are plotted under specific conditions such as regulated output voltage and constant temperature (27 °C is often used). The graph is often used when working out what size alternator will be required for a specific application.

The power curve is used to calculate the type of drive belt needed to transmit the power or torque to the alternator. As an aside, the power curve and the current curve can be used together to calculate the efficiency of the alternator. At any particular speed when producing maximum output for that speed, the efficiency of any machine is calculated from:

Efficiency = Power out/Power in

Efficiency at 8000 rpm is:

(Power out = 14 V x 70 A = 980 W)

980 W/2300 W = 0.43 or about 43%

Efficiency at 2/3 rated output:

(Power out = 14 V x 47 A = 653 W)

653/1100 = 0.59 or about 59%

These figures help to illustrate how much power is lost in the generation process. The inefficiency is mainly due to iron losses, copper losses, windage (air friction) and mechanical friction. The energy is lost as heat.

6.5.4 Mechanical and external considerations

Most light vehicle alternators are mounted in similar ways. This usually involves a pivoted mounting on the side of the engine with an adjuster on the top or bottom to set drive belt tension. It is now common practice to use 'multi-V' belts driving directly from the engine crankshaft pulley. This type of belt will transmit greater torque and can be worked on smaller diameter pulleys or with tighter corners than the more traditional 'V' belt. Figure 6.40 is an extract from information regarding the mounting and drive belt fitting for a typical alternator.

Key fact

The drive ratio between the crank pulley and alternator pulley is important.

The drive ratio between the crank pulley and alternator pulley is important. A typical ratio is about 2.5 : 1. In simple terms, the alternator should be driven as fast as possible at idle speed, but must not exceed the maximum rated speed of the alternator at maximum engine speed. The ideal ratio can therefore be calculated as follows:

Maximum ratio = max alternator speed/max engine speed.

For example: 15 000 rpm / 6000 rpm = 2.5 : 1

During the design stage the alternator will often have to be placed in a position determined by the space available in the engine compartment. However, where possible the following points should be considered:

- Adequate cooling.
- Suitable protection from contamination.
- Access for adjustment and servicing.
- Minimal vibration if possible.
- Recommended belt tension.

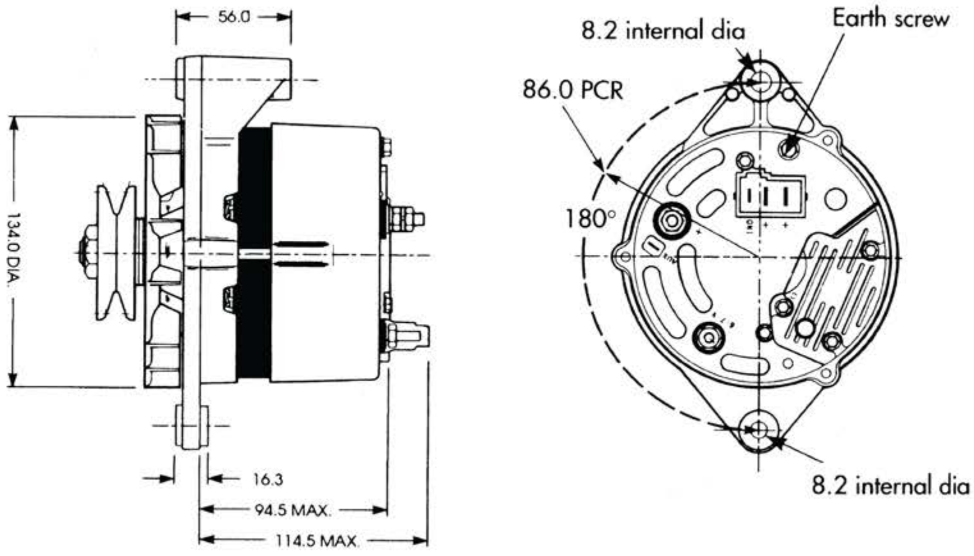


Figure 6.40 Alternator mounting specifications

CHAPTER 7

Starting

7.1 Requirements of the starting system

7.1.1 Engine starting requirements

An internal combustion engine requires the following criteria in order to start and continue running:

- Combustible mixture.
- Compression stroke.
- A form of ignition.
- The minimum starting speed (about 100 rpm).

In order to produce the first three of these, the minimum starting speed must be achieved. This is where the electric starter comes in. The ability to reach this minimum speed is dependent on a number of factors:

- Rated voltage of the starting system.
- Lowest possible temperature at which it must still be possible to start the engine. This is known as the starting limit temperature.
- Engine cranking resistance. In other words the torque required to crank the engine at its starting limit temperature (including the initial stalled torque).
- Battery characteristics.
- Voltage drop between the battery and the starter.
- Starter-to-ring gear ratio.
- Characteristics of the starter.
- Minimum cranking speed of the engine at the starting limit temperature.

The starter is an isolated component within the vehicle electrical system, as Figure 7.2 shows. The battery in particular is of prime importance.

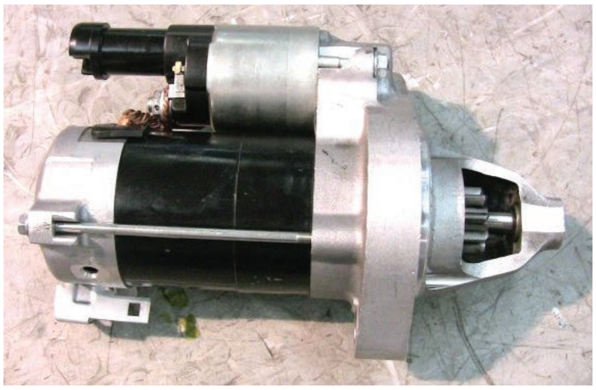


Figure 7.1 Honda starter motor

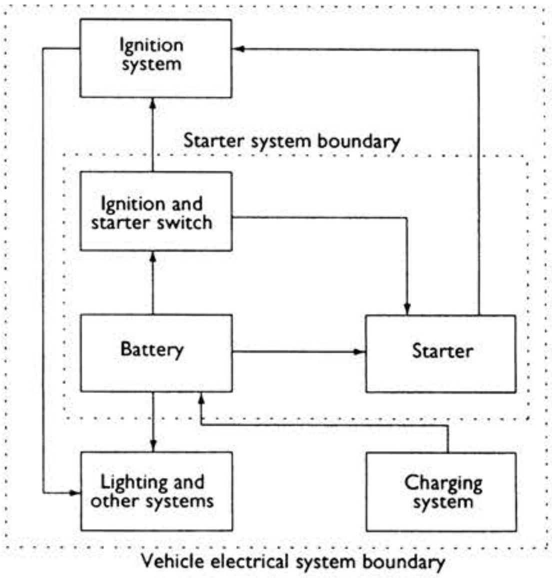


Figure 7.2 Starting system as part of the complete electrical system

Key fact

As temperature decreases, starter torque decreases and the torque required to crank the engine to its minimum speed increases.

Another particularly important consideration in relation to engine starting requirements is the starting limit temperature. Figure 7.3 shows how, as temperature decreases, starter torque also decreases and the torque required to crank the engine to its minimum speed increases.

Typical starting limit temperatures are –18 °C to –25 °C for passenger cars and –15 °C to –20 °C for trucks and buses. Figures from starter manufacturers are normally quoted at both –20 °C and +20 °C.

7.1.2 Starting system design

The starting system of any vehicle must meet a number of criteria in excess of the eight listed above.

- Long service life and maintenance free.
- Continuous readiness to operate.

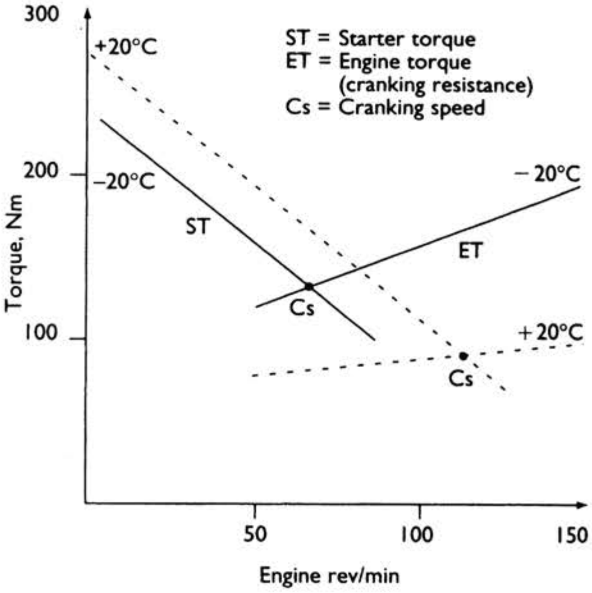


Figure 7.3 Starter torque and engine cranking torque

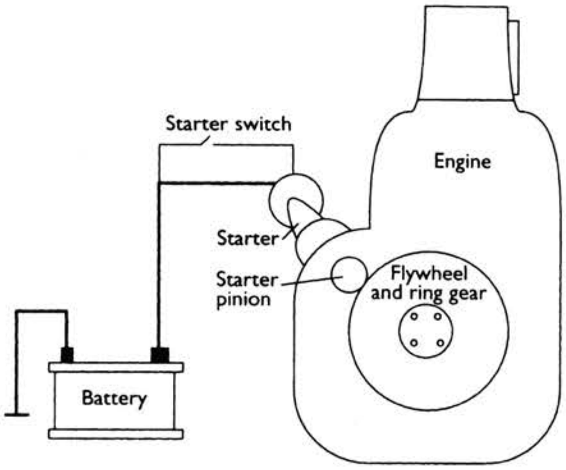


Figure 7.4 Starter system general layout

- Robust, such as to withstand starting forces, vibration, corrosion and temperature cycles.
- Lowest possible size and weight.

Figure 7.4 shows the starting system general layout. It is important to determine the minimum cranking speed for the particular engine. This varies considerably with the design and type of engine. Some typical values are given in Table 7.1 for a temperature of 20 °C.

The rated voltage of the system for passenger cars is, almost without exception, 12 V. Trucks and buses are generally 24 V as this allows the use of half the current that would be required with a 12 V system to produce the same power. It will also considerably reduce the voltage drop in the wiring, as the length of wires used on commercial vehicles is often greater than passenger cars.

The rated output of a starter motor can be determined on a test bench. A battery of maximum capacity for the starter, which has a 20% drop in capacity at 20 °C, is connected to the starter by a cable with a resistance of 1 mΩ. These criteria will ensure the starter is able to operate even under the most adverse conditions. The actual output of the starter can now be measured under typical operating conditions. The rated power of the motor corresponds to the power drawn from the battery less copper losses (due to the resistance of the circuit), iron losses (due to eddy currents being induced in the iron parts of the motor) and friction losses.

Figure 7.5 shows an equivalent circuit for a starter and battery. This indicates how the starter output is very much determined by line resistance and battery internal resistance. The lower the total circuit resistance is, then the higher the output from the starter.

Table 7.1 Typical minimum cranking speeds

Engine	Minimum cranking speed (rpm)
Reciprocating spark ignition	60–90
Rotary spark ignition	150–180
Diesel with glow plugs	60–140
Diesel without glow plugs	100–200

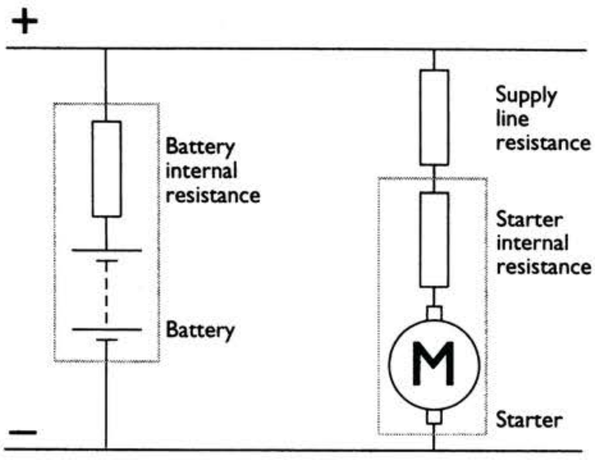


Figure 7.5 Equivalent circuit for starter system

There are two other considerations when designing a starting system. The location of the starter on the engine is usually pre-determined, but the position of the battery must be considered. Other constraints may determine this, but if the battery is closer to the starter the cables will be shorter. A longer run will mean cables with a greater cross-section are needed to ensure a low resistance. Depending on the intended use of the vehicle, special sealing arrangements on the starter may be necessary to prevent the ingress of contaminants. Starters are available designed with this in mind. This may be appropriate for off-road vehicles.

7.1.3 Choosing a starter motor

As a guide, the starter motor must meet all the criteria previously discussed. Referring back to Figure 7.3 (the data showing engine cranking torque compared with minimum cranking speed) will determine the torque required from the starter.

Key fact

Manufacturers of starter motors provide data in the form of characteristic curves.

Manufacturers of starter motors provide data in the form of characteristic curves. These are discussed in more detail in the next section. The data will show the torque, speed, power and current consumption of the starter at –20°C and +20 °C. The power rating of the motor is quoted as the maximum output at 20 °C using the recommended battery.

As a very general guide the stalled (locked) starter torque required per litre of engine capacity at the starting limit temperature is as shown in Table 7.2.

Table 7.2 Torque required for various engine sizes

Engine cylinders	Torque per litre
2	12.5Nm
4	8.0Nm
6	6.5Nm
8	6.0Nm
12	5.5Nm

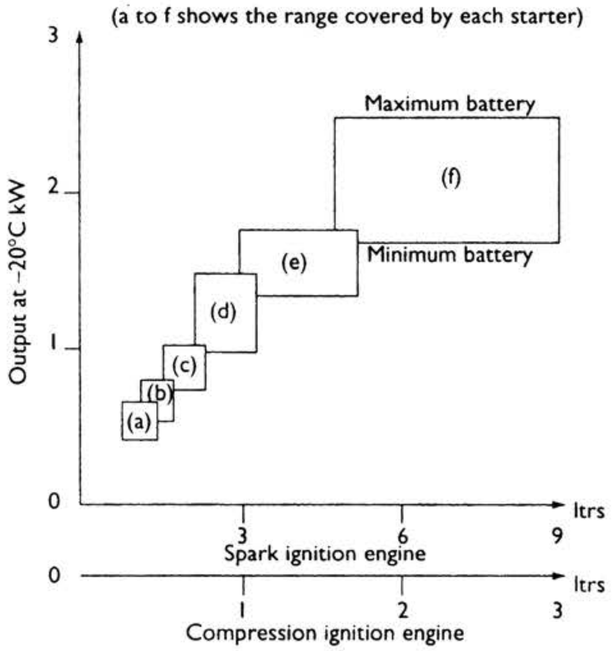


Figure 7.6 Power output of the starter compared with engine size

Key fact

Starter motor stalled torque is generally three to four times the cranking torque.

A greater torque is required for engines with a lower number of cylinders due to the greater piston displacement per cylinder. This will determine the peak torque values. The other main factor is compression ratio.

To illustrate the link between torque and power, we can assume that, under the worst conditions (20 °C), a four-cylinder 2-litre engine requires 480 Nm to overcome static friction and 160 Nm to maintain the minimum cranking speed of 100 rev/ min. With a starter pinion-to-ring gear ratio of 10 : 1, the motor must therefore, be able to produce a maximum stalled torque of 48 Nm and a driving torque of 16 Nm. This is working on the assumption that stalled torque is generally three to four times the cranking torque.

A greater torque is required for engines with a lower number of cylinders due to the greater piston displacement per cylinder. This will determine the peak torque values. The other main factor is compression ratio.

To illustrate the link between torque and power if we assume that under the worst conditions (–20 °C) a four cylinder 2 litre engine requires 240 Nm to overcome static friction and 80 Nm to maintain the minimum cranking speed of 100 rpm. With a starter pinion to ring gear ratio of 10:1 the motor must therefore, be able to produce a maximum stalled torque of 24 Nm and a driving torque of 8 Nm. This is working on the assumption that stalled torque is generally three to four times cranking torque.

Torque is converted to power as follows:

$$P = T\omega$$

where: P = power, T = torque, ω = angular velocity,

$$\omega = \frac{2\pi n}{60}$$

where: n = rpm.

In this example the power developed at 1000 rpm with a torque of 8 Nm (at the starter), is about 840 W. Referring back to Figure 7.6 the ideal choice would appear to be the starter marked (c).

The maximum permissible recommended battery would be 55 Ah and 255 A cold start performance.

7.2 Starter motors and circuits

7.2.1 Starting system circuits

In comparison with most other circuits on the modern vehicle, the starter circuit is very simple. The problem to be overcome, however, is that of volt drop in the main supply wires. The starter is usually operated by a spring-loaded key switch, and the same switch also controls the ignition and accessories. The supply from the key switch, via a relay in many cases, causes the starter solenoid to operate, and this in turn, by a set of contacts, controls the heavy current. In some cases an extra terminal on the starter solenoid provides an output when cranking, which is usually used to bypass a dropping resistor on the ignition or fuel pump circuits. The basic circuit for the starting system is shown in Figure 7.7.

Key fact

The volt drop in the main starter supply wires must be kept to a minimum.

The problem of volt drop in the main supply circuit is due to the high current required by the starter, particularly under adverse starting conditions such as very low temperatures.

A typical cranking current for a light vehicle engine is of the order of 100 to 150 A, but this may peak in excess of 500 A to provide the initial stalled torque. It is generally accepted that a maximum volt drop of only 0.5 V should be allowed between the battery and the starter when operating. An Ohm's law calculation indicates that the maximum allowed circuit resistance is 2.5 m when using a 12 V supply. This is a worst case situation and lower resistance values are used in most applications. The choice of suitable conductors is therefore very important.

Key fact

A typical cranking current for a light vehicle engine is of the order of 100 to 150 A, but this may peak in excess of 500 A to provide the initial stalled torque.

7.2.2 Example circuits

The circuit shown in Figure 7.8 is from a Ford vehicle fitted with manual or automatic transmission. The inhibitor circuits will only allow the starter to operate when the automatic transmission is in 'park' or 'neutral'. Similarly for the manual version, the starter will only operate if the clutch pedal is depressed.

The starter relay coil is supplied with the positive connection by the key switch. The earth path is connected through the appropriate inhibitor switch. To prevent starter operation when the engine is running the power control module (EEC V) controls the final earth path of the relay.

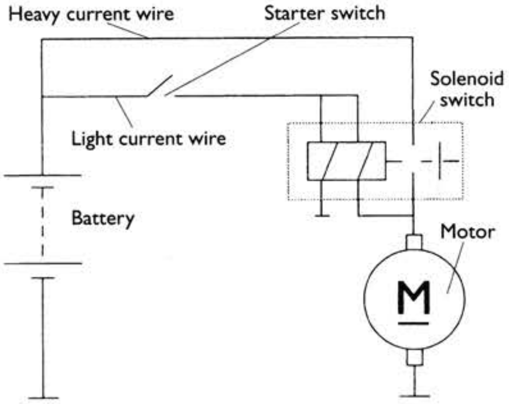


Figure 7.7 basic starting circuit

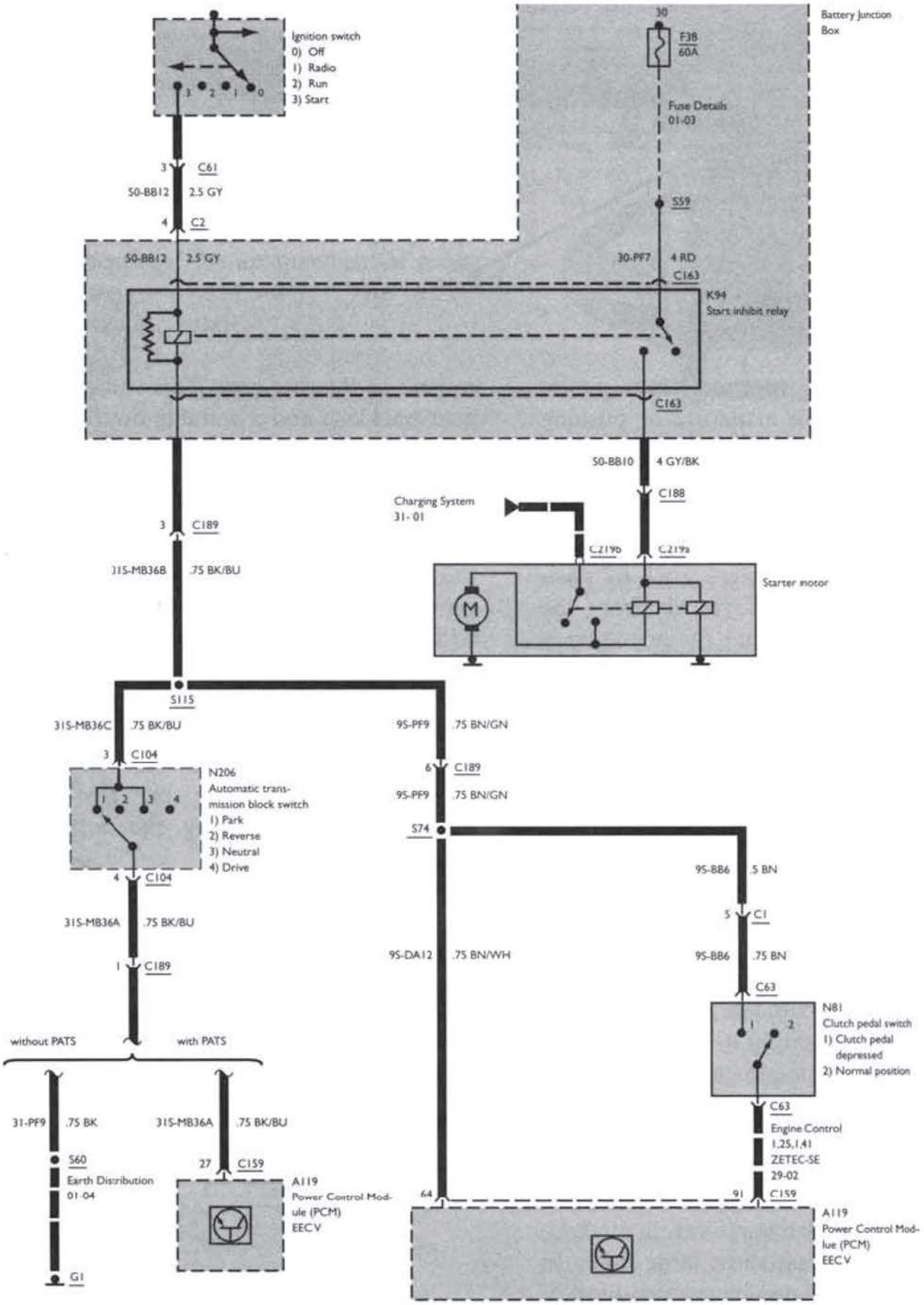


Figure 7.8 Starter circuit as used by Ford (Source: Ford Motor Company)

A resistor fitted across the relay coil reduces back EMF. The starter in current use is a standard pre-engaged, permanent magnet motor.

Starter motor circuits may have additional automatic switching to prevent the engine being started in particular situations. Automatic transmission systems incorporate an inhibitor switch on the gear selector, which allows engine starting in the park and neutral positions only. This prevents the engine being started with the transmission in gear, which could result in the vehicle pulling away unexpectedly. The inhibitor switch must be carefully checked and adjusted so that there is no risk of incorrect operation.

In the diagram shown as Figure 7.9, the powertrain control module (PCM) allows the engine to start, only when the passive anti-theft system (PATS) reads a key which transmits a valid code. On a key-free vehicle, the passive key is recognized by the key-free module and if the key is valid the permission to start is issued directly. On vehicles with a manual transmission it is necessary to depress the clutch pedal; on those with automatic transmission the brake pedal must be pressed. On a key-free system the key-free module switches on the control voltage for the starter relay.

Key fact

On a key-free system the key-free module switches on the control voltage for the starter relay.

The PCM switches the ground in the control circuit of the starter relay which then connects power through to the starter solenoid. As soon as the speed of the

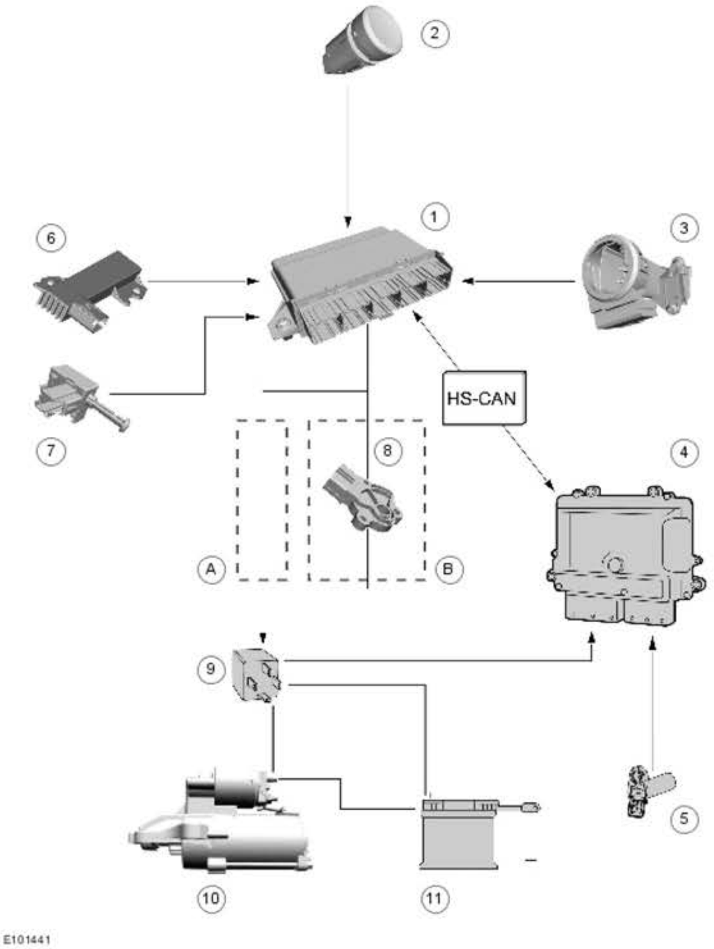


Figure 7.9 Keyless starting system 1-Keylesss vehicle module, 2-Start/stop button, 3-Electronic steering lock, 4-Powertrain control module, 5-Crank sensor, 6-Keyless vehicle antenna, 7-Vehicles with manual transmission: Clutch pedal position switch / vehicles with automatic transmission: Stoplamp switch, 8-The TR sensor, 9-Starter relay, 10-Starter motor, 11-Battery (Source: Ford Motor Company)

engine has reached 750 rpm or the maximum permitted start time of 30 seconds has been exceeded, the PCM switches off the starter relay and therefore the starter motor. This protects the starter. If the engine does not turn or turns only slowly, the starting process is aborted by the PCM.

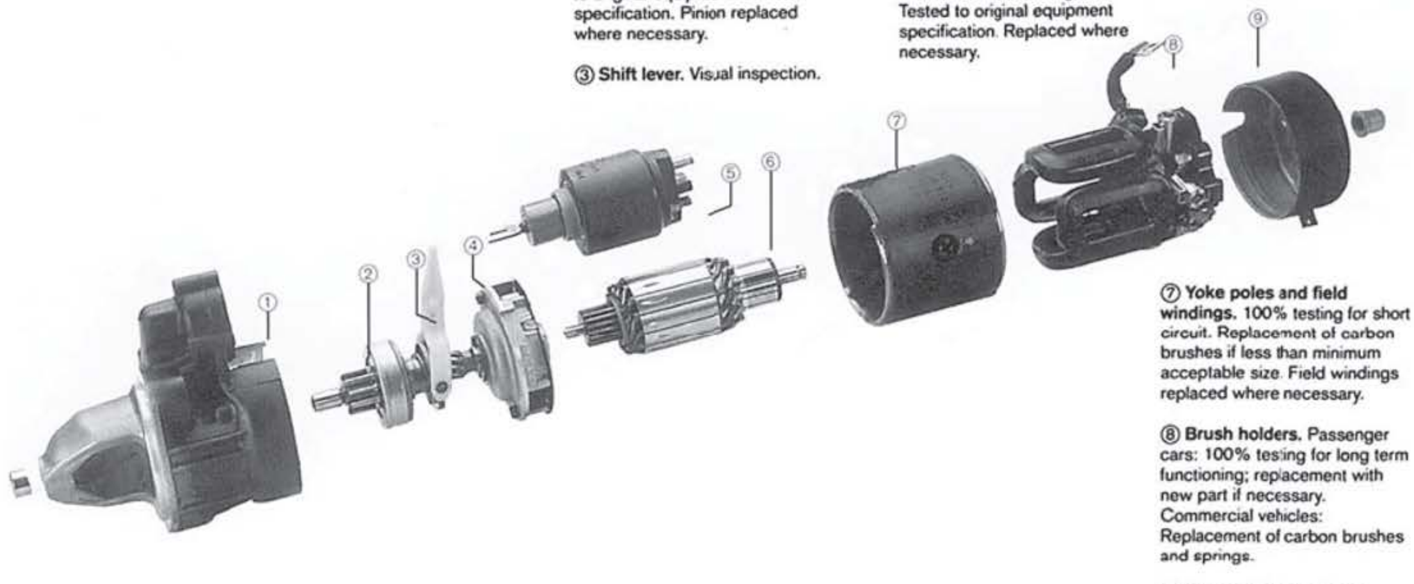
7.2.3 Starter circuit testing

The process of checking a 12 V starting system operation is as follows (tests 3 to 8 are all carried out while trying to crank the engine).

1. Battery (at least 70%).
2. Hand and eye checks.
3. Battery volts (minimum 10 V).
4. Solenoid lead (same as battery).
5. Starter voltage (no more than 0.5 V less than battery).
6. Insulated line volt drop (maximum 0.25 V).
7. Solenoid contacts volt drop (almost 0 V).
8. Earth line volt drop (maximum 0.25 V).

The idea of these tests is to see if the circuit is supplying all the available voltage to the starter. If it is, then the starter is at fault, if not then the circuit is at fault.

If the starter is found to be defective then a replacement unit is the normal recommendation. Figure 7.10 explains the procedure used by Bosch to ensure



① **Drive end bearing and bushes.** 100% replacement of bushes. Remanufacture to original equipment specification.

② **Pinion and overrun clutch drive.** 100% dismantling and cleaning. 100% replacement of bushes, rollers, springs and plastic insert parts. Lubrication to original equipment specification. Pinion replaced where necessary.

③ **Shift lever.** Visual inspection.

④ **Planetary gear train.** Dismantling. Functional testing of the needle bearing in the installed condition. 100% replacement of bearing bushes.

⑤ **Solenoid.** Testing the ease of movement of the armature. Dismantling of the switch cover. 100% replacement of contact bolts and contact bridges. Tested to original equipment specification. Replaced where necessary.

⑥ **Armature with commutator.** Testing for short circuit. Testing the rotation of laminated armature core and commutator. 100% commutator skimming and undercutting. Replacement of commutator if diameter is less than the minimum accepted.

⑦ **Yoke poles and field windings.** 100% testing for short circuit. Replacement of carbon brushes if less than minimum acceptable size. Field windings replaced where necessary.

⑧ **Brush holders.** Passenger cars: 100% testing for long term functioning; replacement with new part if necessary. Commercial vehicles: Replacement of carbon brushes and springs.

⑨ **Commutator bearings.** 100% replacement to meet original equipment specification.

Figure 7.10 Quality starter overhaul procedure

Table 7.3 Common starting system symptoms and faults

Symptom	Possible fault
Engine does not rotate when trying to start	Battery connection loose or corroded Battery discharged or faulty Broken loose or disconnected wiring in the starter circuit Defective starter switch or automatic gearbox inhibitor switch Starter pinion or flywheel ring gear loose Earth strap broken. Loose or corroded
Starter noisy	Starter pinion or flywheel ring gear loose Starter mounting bolts loose Starter worn (bearings etc.) Discharged battery (starter may jump in and out)
Starter turns engine slowly	Discharged battery (slow rotation) Battery terminals loose or corroded Earth strap or starter supply loose or disconnected High resistance in supply or earth circuit Internal starter fault

quality exchange units. Repairs are possible but only if the general state of the motor is good.

Table 7.3 lists some common symptoms of a charging system malfunction together with suggestions for the possible fault.

7.2.4 Principle of operation

Definition

Motor: A machine to convert electrical energy into mechanical energy.

The simple definition of any motor is a machine to convert electrical energy into mechanical energy. The starter motor is no exception. When current flows through a conductor placed in a magnetic field, a force is created acting on the conductor relative to the field. The magnitude of this force is proportional to the field strength, the length of the conductor in the field and the current flowing in the conductor.

In any DC motor, the single conductor is of no practical use and so the conductor is shaped into a loop or many loops to form the armature. A many-segment commutator allows contact via brushes to the supply current.

The force on the conductor is created due to the interaction of the main magnetic field and the field created around the conductor. In a light vehicle starter motor, the main field was traditionally created by heavy duty series windings wound around soft iron pole shoes. Due to improvements in magnet technology, permanent magnet fields allowing a smaller and lighter construction are replacing wire-wound fields. The strength of the magnetic field created around the conductors in the armature is determined by the value of the current flowing. The principle of a DC motor is shown in Figure 7.11.

Key fact

Most starter designs use a four-pole four-brush system.

Most starter designs use a four-pole four-brush system. Using four field poles concentrates the magnetic field in four areas as shown in Figure 7.12. The magnetism is created in one of three ways, permanent magnets, series field windings or series-parallel field windings.

Figure 7.13 shows the circuits of the two methods where field windings are used. The series–parallel fields can be constructed with a lower resistance thereby increasing the current and hence torque of the motor. Four brushes are used

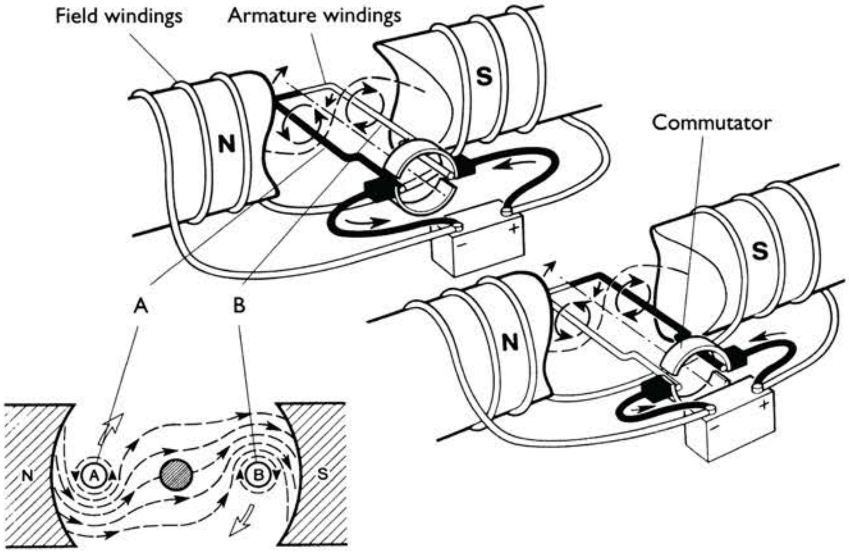


Figure 7.11 Interaction of two magnetic fields results in rotation when a commutator is used to reserve the supply each half turn

to carry the heavy current. The brushes are made of a mixture of copper and carbon, as is the case for most motor or generator brushes. Starter brushes have a higher copper content to minimize electrical losses. Figure 7.14 shows some typical field coils with brushes attached. The field windings on the right are known as wave wound.

The armature consists of a segmented copper commutator and heavy duty copper windings. The windings on a motor armature can, broadly speaking, be wound in two ways. These are known as lap winding and wave winding. Figure 7.15 shows the difference between these two methods. Starter motors tend to use wave winding as this technique gives the most appropriate torque and speed characteristic for a four-pole system.

A starter must also have some method of engaging with, and release from, the vehicle's flywheel ring gear. In the case of light vehicle starters, this is achieved either by an inertia-type engagement or a pre-engagement method. These are both discussed further in subsequent sections.

Key fact

Starter brushes are a mixture of copper and carbon, as is the case for most motor or generator brushes – but with more copper to lower the resistance.

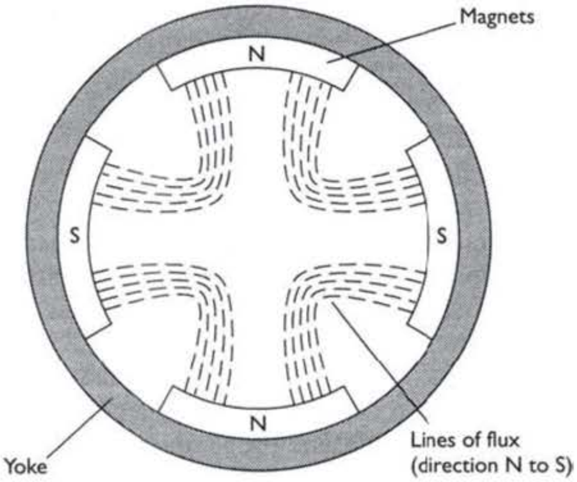


Figure 7.12 Four-pole magnetic field

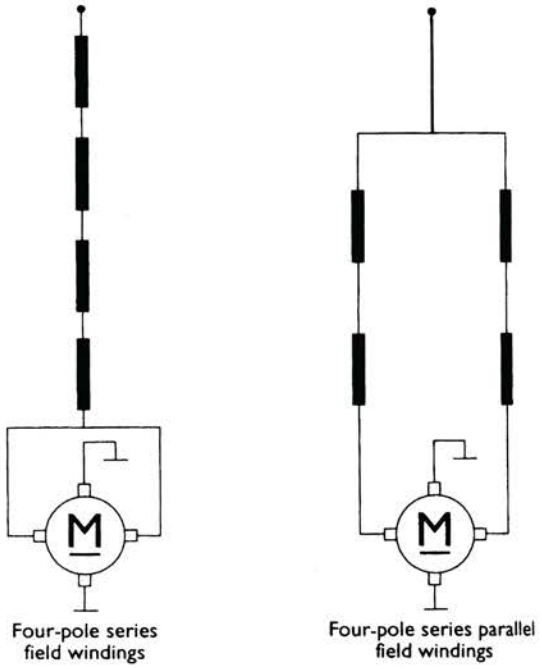


Figure 7.13 Starter internal circuits

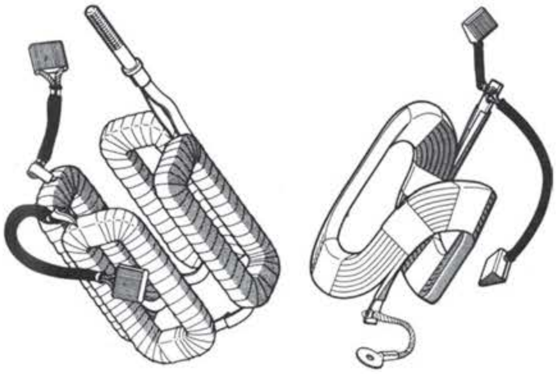


Figure 7.14 Typical field coils and brushes

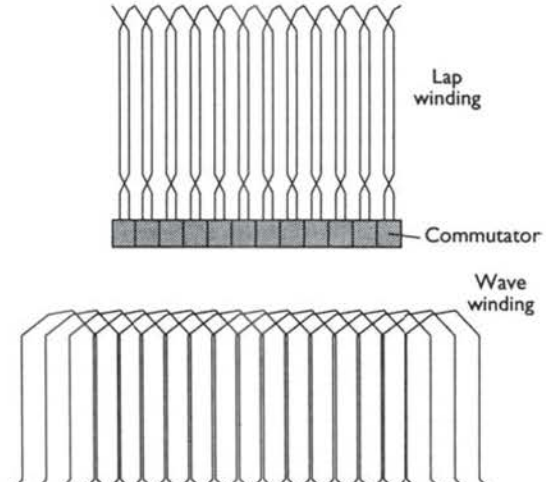


Figure 7.15 Typical lap and wave wound armature circuits

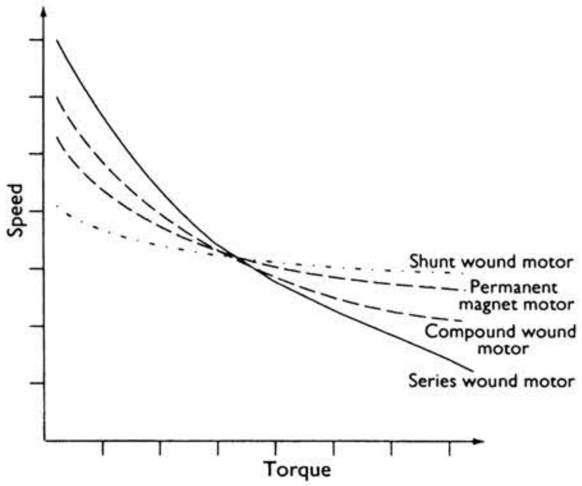


Figure 7.16 Speed and torque characteristics of DC motors

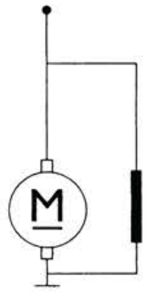


Figure 7.17 Shunt wound motor (parallel wound)

Figure 7.18 Series wound DC motor

7.2.5 DC motor characteristics

It is possible to design a motor with characteristics that are most suitable for a particular task. For a comparison between the main types of DC motor, the speed–torque characteristics are shown in Figure 7.16. The four main types of motor are referred to as shunt wound, series wound, compound wound and permanent magnet excitation.

In shunt wound motors, the field winding is connected in parallel with the armature as shown in Figure 7.17. Due to the constant excitation of the fields, the speed of this motor remains constant, virtually independent of torque.

Series wound motors have the field and armature connected in series. Because of this method of connection, the armature current passes through the fields making it necessary for the field windings to consist usually of only a few turns of heavy wire. When this motor starts under load the high initial current, due to low resistance and no back EMF, generates a very strong magnetic field and therefore high initial torque. This characteristic makes the series wound motor ideal as a starter motor. Figure 7.18 shows the circuit of a series wound motor.

Key fact

Series wound motors have the field and armature connected in series.

The compound wound motor, as shown in Figure 7.19, is a combination of shunt and series wound motors. Depending on how the field windings are connected, the characteristics can vary. The usual variation is where the shunt winding is connected, which is either across the armature or across the armature and series winding. Large starter motors are often compound wound and can be operated in two stages. The first stage involves the shunt winding being connected in series with the armature. This unusual connection allows for low meshing torque due to the resistance of the shunt winding. When the pinion of the starter is fully in mesh with the ring gear, a set of contacts causes the main supply to be passed through the series winding and armature giving full torque. The shunt winding will now be connected in parallel and will act in such a way as to limit the maximum speed of the motor.

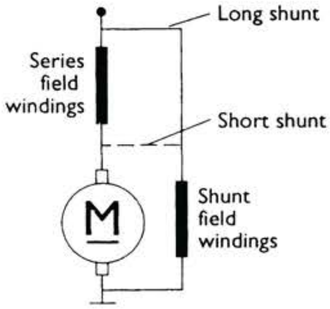


Figure 7.19 Compound wound motor

Permanent magnet motors are smaller and simpler compared with the other three discussed. Field excitation, as the name suggests, is by permanent magnet. This excitation will remain constant under all operating conditions. Figure 7.20 shows the accepted representation for this type of motor.

The characteristics of this type of motor are broadly similar to the shunt wound motors. However, when one of these types is used as a starter motor, the drop

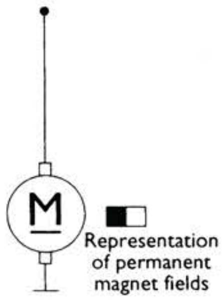


Figure 7.20 Permanent magnet motor

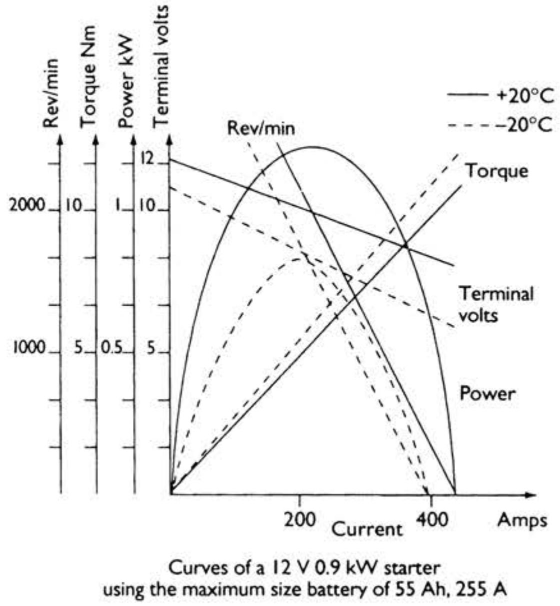


Figure 7.21 Starter motor characteristic curves

Key fact

Note that the maximum power of a starter motor is developed at mid-range speed but maximum torque is at zero speed.

in battery voltage tends to cause the motor to behave in a similar way to a series wound machine. In some cases though, the higher speed and lower torque characteristic are enhanced by using an intermediate transmission gearbox inside the starter motor.

Information on particular starters is provided in the form of characteristic curves. Figure 7.21 shows the details for a typical light vehicle starter motor.

This graph shows how the speed of the motor varies with load. Owing to the very high speeds developed under no load conditions, it is possible to damage this type of motor. Running off load due to the high centrifugal forces on the armature may cause the windings to be destroyed. Note that the maximum power of this motor is developed at midrange speed but maximum torque is at zero speed.

7.3 Types of starter motor

7.3.1 Inertia starters

In all standard motor vehicle applications it is necessary to connect the starter to the engine ring gear only during the starting phase. If the connection remained permanent, the excessive speed at which the starter would be driven by the engine would destroy the motor almost immediately.

The inertia type of starter motor (Figure 7.22) has been the technique used for over 80 years, but is now becoming redundant. It is a four-pole, four-brush machine and was used on small to medium-sized petrol engined vehicles. It is capable of producing 9.6 Nm with a current draw of 350 A. The motor shown uses a face-type commutator and axially aligned brush gear. The fields are wave wound and are earthed to the starter yoke.

The starter engages with the flywheel ring gear by means of a small pinion. The toothed pinion and a sleeve splined on to the armature shaft are threaded such that when the starter is operated, via a remote relay, the armature will cause the sleeve to rotate inside the pinion. The pinion remains still due to its inertia and,

Figure 7.22 Inertia type starter

because of the screwed sleeve rotating inside it, the pinion is moved to mesh with the ring gear.

When the engine fires and runs under its own power, the pinion is driven faster than the armature shaft. This causes the pinion to be screwed back along the sleeve and out of engagement with the flywheel. The main spring acts as a buffer when the pinion first takes up the driving torque and also acts as a buffer when the engine throws the pinion back out of mesh.

One of the main problems with this type of starter was the aggressive nature of the engagement. This tended to cause the pinion and ring gear to wear prematurely. In some applications the pinion tended to fall out of mesh when cranking due to the engine almost, but not quite, running. The pinion was also prone to seizure often due to contamination by dust from the clutch. This was often compounded by application of oil to the pinion mechanism, which tended to attract even more dust and thus prevent engagement.

The pre-engaged starter motor has largely overcome these problems.

7.3.2 Pre-engaged starters

Pre-engaged starters are fitted to the majority of vehicles in use today. A cutaway pre-engaged starter is shown as Figure 7.23. This type of starter provides a positive engagement with the ring gear, as full power is not applied until the pinion is fully in mesh. They prevent premature ejection as the pinion is held into mesh by the action of a solenoid. A one-way clutch is incorporated into the pinion to prevent the starter motor being driven by the engine. One example of a pre-engaged starter similar to many in common use is shown in Figure 7.24.

Figure 7.25 shows the circuit associated with operating this type of pre-engaged starter. The basic operation of the pre-engaged starter is as follows. When the key switch is operated, a supply is made to terminal 50 on the solenoid. This causes two windings to be energized, the hold-on winding and the pull-in winding. Note that the pull-in winding is of very low resistance and hence a high current flows. This winding is connected in series with the motor circuit and the current flowing will allow the motor to rotate slowly to facilitate engagement. At the same time, the magnetism created in the solenoid attracts the plunger and, via an operating lever, pushes the pinion into mesh with the flywheel ring gear.

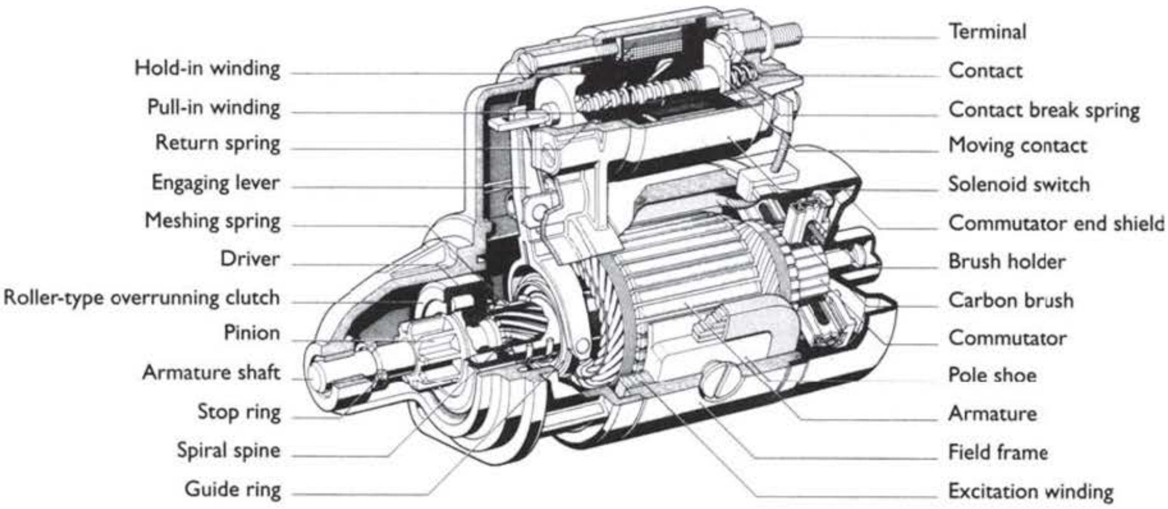


Figure 7.23 Pre-engaged starter

Figure 7.24 Bosch starter

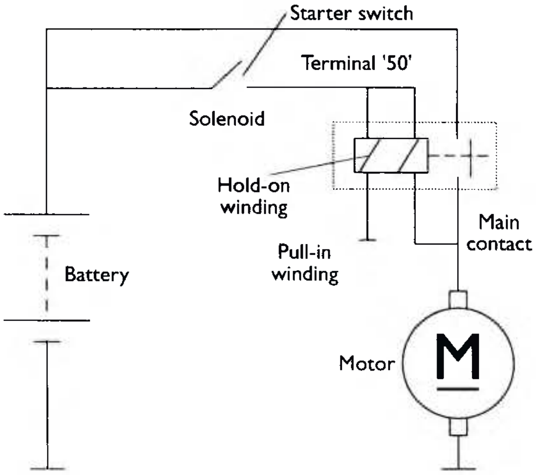


Figure 7.25 Starter circuit

When the pinion is fully in mesh the plunger, at the end of its travel, causes a heavy-duty set of copper contacts to close. These contacts now supply full battery power to the main circuit of the starter motor. When the main contacts are closed, the pull-in winding is effectively switched off due to equal voltage supply on both ends. The hold-on winding holds the plunger in position as long as the solenoid is supplied from the key switch.

When the engine starts and the key is released, the main supply is removed and the plunger and pinion return to their rest positions under spring tension. A lost motion spring located on the plunger ensures that the main contacts open before the pinion is retracted from mesh.

During engagement, if the teeth of the pinion hit the teeth of the flywheel (tooth to tooth abutment), the main contacts are allowed to close due to the engagement spring being compressed. This allows the motor to rotate under power and the pinion will slip into mesh.

Figure 7.26 shows a sectioned view of a one-way clutch assembly. The torque developed by the starter is passed through the clutch to the ring gear. The purpose of this free-wheeling device is to prevent the starter being driven at an excessively high speed if the pinion is held in mesh after the engine has started.

Key fact

The purpose of the free-wheeling device is to prevent the starter being driven at an excessively high speed if the pinion is held in mesh after the engine has started.

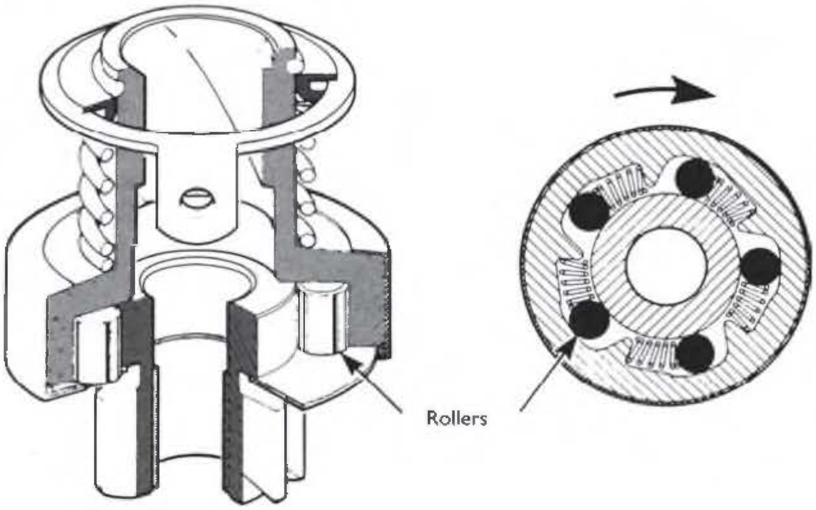


Figure 7.26 One-way roller clutch drive pinion

The clutch consists of a driving and driven member with several rollers between the two. The rollers are spring loaded and either wedge-lock the two members together by being compressed against the springs, or free-wheel in the opposite direction.

Many variations of the pre-engaged starter are in common use, but all work on similar lines to the above description. The wound field type of motor has now largely been replaced by the permanent magnet version.

7.3.3 Permanent magnet starters

Permanent magnet starters began to appear on production vehicles in the late 1980s. The two main advantages of these motors, compared with conventional types, are less weight and smaller size. This makes the permanent magnet starter a popular choice by vehicle manufacturers as, due to the lower lines of today's cars, less space is now available for engine electrical systems. The reduction in weight provides a contribution towards reducing fuel consumption.

The principle of operation is similar in most respects to the conventional pre-engaged starter motor. The main difference being the replacement of field windings and pole shoes with high quality permanent magnets. The reduction in weight is in the region of 15% and the diameter of the yoke can be reduced by a similar factor.

Permanent magnets provide constant excitation and it would be reasonable to expect the speed and torque characteristic to be constant.

However, due to the fall in battery voltage under load and the low resistance of the armature windings, the characteristic is comparable to series wound motors. In some cases, flux concentrating pieces or interpoles are used between the main magnets. Due to the warping effect of the magnetic field, this tends to make the characteristic curve very similar to that of the series motor.

Development by some manufacturers has also taken place in the construction of the brushes. A copper and graphite mix is used but the brushes are made in two parts allowing a higher copper content in the power zone and a higher graphite content in the commutation zone. This results in increased service life and a reduction in voltage drop, giving improved starter power. Figure 7.28 shows a modern permanent magnet (PM) starter.

For applications with a higher power requirement, permanent magnet motors with intermediate transmission have been developed. These allow the armature

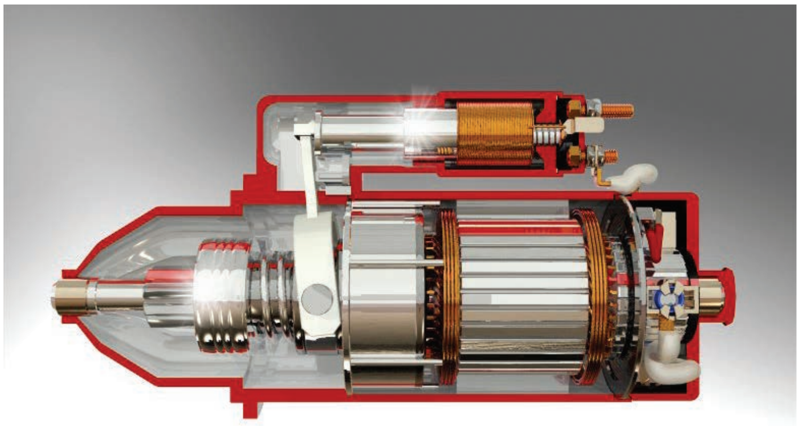


Figure 7.27 Cutaway view of a permanent magnet starter

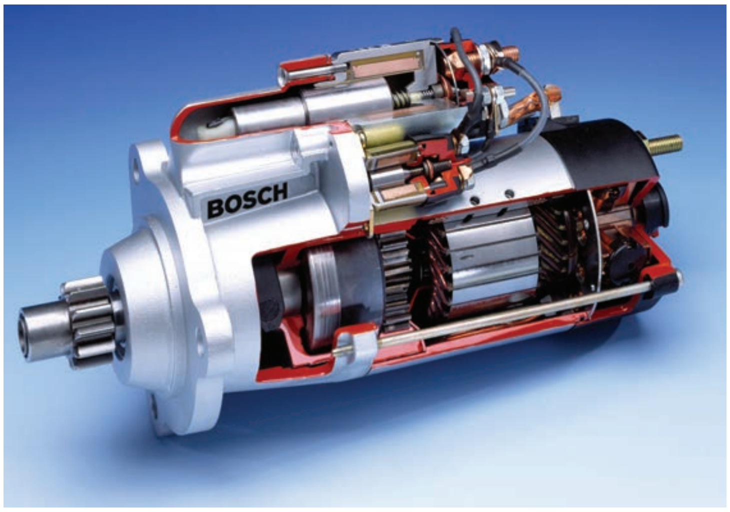


Figure 7.28 Permanent magnet starter (Source: Bosch Media)

Key fact

For applications with a higher power requirement, permanent magnet motors with intermediate transmission have been developed.

to rotate at a higher and more efficient speed whilst still providing the torque, due to the gear reduction. Permanent magnet starters with intermediate transmission are available with power outputs of about 1.7 kW and are suitable for spark ignition engines up to about 3 litres, or compression ignition engines up to about 1.6 litres. This form of permanent magnet motor can give a weight saving of up to 40%. The principle of operation is again similar to the conventional pre-engaged starter. The intermediate transmission, as shown in Figures 7.29 and 7.30, is of the epicyclic type.

The sun gear is on the armature shaft and the planet carrier drives the pinion. The ring gear or annulus remains stationary and also acts as an intermediate bearing. This arrangement of gears gives a reduction ratio of about 5 : 1. This can be calculated by the formula:

$$\text{Ratio} = AS/S$$

where: A = number of teeth on the annulus, S = number of teeth on the sun gear.

The annulus gear in some types is constructed from a high grade polyamide compound with mineral additives to improve strength and wear resistance. The sun and planet gears are conventional steel. This combination of materials gives a quieter and more efficient operation.

Figure 7.29 Starter motor intermediate transmission

Figure 7.30 Pre-engaged starter internal components

7.3.4 Integrated starters

A device called a 'dynastart' was used on a number of vehicles from the 1930s through to the 1960s. This device was a combination of the starter and a dynamo. The device, directly mounted on the crankshaft, was a compromise and hence not very efficient.

The method is now known as an Integrated Starter Alternator Damper (ISAD). It consists of an electric motor, which functions as a control element between the engine and the transmission, and can also be used to start the engine and deliver electrical power to the batteries and the rest of the vehicle systems. The electric motor replaces the mass of the flywheel.

The motor transfers the drive from the engine and is also able to act as a damper/ vibration absorber unit. The damping effect is achieved by a rotation capacitor. A change in relative speed between the rotor and the engine due to the vibration, causes one pole of the capacitor to be charged. The effect of this is to take the energy from the vibration.

Using ISAD to start the engine is virtually noiseless, and cranking speeds of 700 rpm are possible. Even at –25 °C it is still possible to crank at about 400 rpm. A good feature of this is that a stop/start function is possible as an economy and emissions improvement technique. Because of the high speed cranking, the engine will fire up in about 0.1–0.5 seconds.

The motor can also be used to aid with acceleration of the vehicle. This feature could be used to allow a smaller engine to be used or to enhance the performance of a standard engine. It is effectively a light hybrid.

When used in alternator mode, the ISAD can produce up to 2 kW at idle speed. It can supply power at different voltages as both AC and DC. Through the application of intelligent control electronics, the ISAD can be up to 80% efficient.

Figure 7.31 Integrated starter alternator damper (Source: Bosch Media)

7.3.5 Electronic starter control

The electronic starter incorporates a static relay on a circuit board integrated into the solenoid switch. This will prevent cranking when the engine is running. Starter control can be supported by an ECU and 'smart' features can be added to improve comfort, safety and service life.

- Starter torque can be evaluated in real time to tell the precise instant of engine start. The starter can be simultaneously shut off to reduce wear and noise generated by the free-wheel phase.
- Thermal protection of the starter components allows optimization of the components to save weight and to give short circuit protection.
- Electrical protection also reduces damage from misuse or system failure.
- Modulating the solenoid current allows redesign of the mechanical parts allowing a softer operation and weight reduction.

7.3.6 Starter installation

Starters are generally mounted in a horizontal position next to the engine crankcase with the drive pinion in a position for meshing with the flywheel or drive plate ring gear.

The starter can be secured in two ways: either by flange or cradle mounting. Flange mounting is the most popular technique used on small and medium-sized vehicles and, in some cases, it will incorporate a further support bracket at the

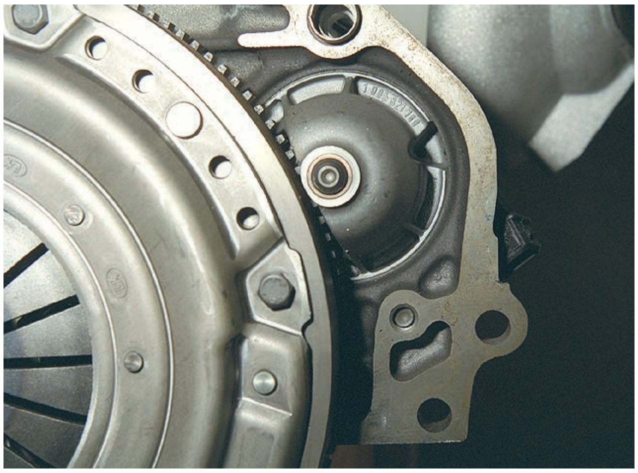


Figure 7.32 Flange mounting is used for most light vehicle starter motors

rear of the starter to reduce the effect of vibration. Larger vehicle starters are often cradle mounted but again also use the flange mounting method, usually fixed with at least three large bolts. In both cases the starters must have some kind of pilot, often a ring machined on the drive end bracket, to ensure correct positioning with respect to the ring gear. This will ensure correct gear backlash and a suitable out of mesh clearance. Figure 7.32 shows the flange mountings method used for most light vehicle starter motors.

Clearly the main load on the vehicle battery is the starter and this is reflected in the size of supply cable required. Any cable carrying a current will experience power loss known as I^2R loss. In order to reduce this power loss, the current or the resistance must be reduced. In the case of the starter the high current is the only way of delivering the high torque. This is the reason for using heavy conductors to the starter to ensure low resistance, thus reducing the volt drop and power loss. The maximum allowed volt drop is 0.5 V on a 12 V system and 1 V on a 24 V system. The short circuit (initial) current for a typical car starter is 500 A and for very heavy applications can be 3000 A.

Key fact

Any cable carrying a current will experience power loss known as I^2R loss.

Control of the starter system is normally by a spring-loaded key switch. This switch will control the current to the starter solenoid, in many cases via a relay. On vehicles with automatic transmission, an inhibitor switch to prevent the engine being started in gear will also interrupt this circuit.

Diesel engined vehicles may have a connection between the starter circuit and a circuit to control the glow plugs. This may also incorporate a timer relay. On some vehicles the glow plugs are activated by a switch position just before the start position.

7.3.7 Belt-driven starter-generator

Gates, well known as manufacturers of drive belts, created a starter-generator concept that is belt driven. It is an electromechanical system made up of a high efficiency induction motor, long-life belt-drive system and sophisticated electronic controls. The belt-driven starter-generator replaces the current alternator and has a similar space requirement.

One of the key components of this system, in addition to the starter-generator is a hydraulic tensioner. This must be able to prevent significant movement during starting but also control system dynamics during acceleration and deceleration

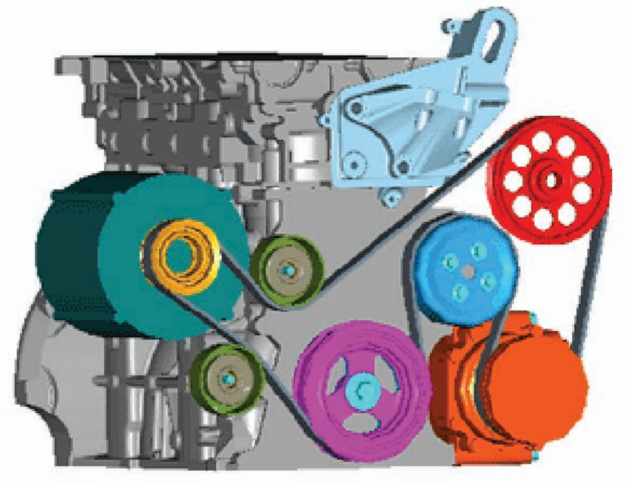


Figure 7.33 Belt-driven starter-generator concept (Source: Gates)

of the engine. The starter-generator is driven by a multi-vee belt, which has been specially designed for the extra load.

The main features and potential benefits of this system are as follows:

- Generating capability of 6 kW at 42 V and a brushless design for 10-year life.
- Regenerative braking and electric torque assist.
- Power for increased feature content and silent cranking at a lower system cost than in-line starter-alternator systems.
- Allows implementation of fuel-saving strategies and emission reduction through hybrid electric strategies, increased cranking speed and start–stop systems.

The starter-generator concept is not new but until recently it could not meet the requirements of modern vehicles. These requirements relate to the starting torque and the power generation capabilities. The biggest advantage of the system under development is that it can be fitted to existing engine designs with only limited modifications. It may, therefore, become a 'stepping stone technology' that allows manufacturers to offer new features without the expense of development and extensive redesigning.

7.3.8 Summary

The overall principle of starting a vehicle engine with an electric motor has changed little in over 80 years. Of course, the motors have become far more reliable and longer lasting. It is interesting to note that, assuming average mileage, the modern starter is used about 2000 times a year in city traffic! This level of reliability has been achieved by many years of research and development.

Key fact

The modern starter is used about 2000 times a year in city traffic.

7.4 Advanced starting system technology

7.4.1 Speed, torque and power

To understand the forces acting on a starter motor let us first consider a single conducting wire in a magnetic field. The force on a single conductor in a magnetic field can be calculated by the formula:

$$F = Bll$$

where: F = force in N, l = length of conductor in the field in m, B = magnetic field strength in Wb/m², I = current flowing in the conductor in amps.

Fleming's left hand rule will serve to give the direction of the force (the conductor is at 90° to the field).

This formula may be further developed to calculate stalled torque of a motor with a number of armature windings as follows:

$$T = BIlrZ$$

where: T = torque in Nm, r = armature radius in m, Z = number of active armature conductors.

This will only produce a result for stalled or lock torque because when a motor is running a back emf is produced in the armature windings. This opposes the applied voltage and hence reduces the current flowing in the armature winding. In the case of a series wound starter motor this will also reduce the field strength B. Armature current in a motor is given by the equation:

$$I = \frac{V - e}{R}$$

where: I = armature current in amps, V = applied voltage in volts, R = resistance of the armature in ohms, e = total back emf in volts.

From the above is should be noted that, at the instant of applying a voltage to the terminals of a motor the armature current will be at a maximum since the back emf is zero. As soon as the speed increases so will back emf and hence armature current will decrease. This is why a starter motor produces 'maximum torque at zero rpm'.

For any DC machine the back emf is given by:

$$e = \frac{2\,p\phi nZ}{c}$$

where: e = back emf in volts, p = number of pairs of poles, ϕ = flux per pole in Webers, n = speed in revs/second, Z = number of armature conductors, c = 2.p for lap wound and 2 for a wave wound machine.

The formula can be re-written for calculating motor speed:

$$n = \frac{ce}{2\,p\phi Z}$$

If the constants are removed from this formula it shows clearly the relationship between field flux, speed and back emf.

$$n \propto \frac{e}{\phi}$$

To consider the magnetic flux (ϕ) it is necessary to differentiate between permanent magnet starters and those using excitation via windings. Permanent magnetism remains reasonably constant. The construction and design of the magnet determine its strength. Flux density can be calculated as follows:

$$B = \frac{\phi}{A}$$

(Units: T (tesla) or Wb/m²)

where: A = area of the pole perpendicular to the flux.

Pole shoes with windings are more complicated as the flux density depends on the material of the pole shoe as well as the coil and the current flowing. The magneto-motive force (MMF) of a coil is determined as follows:

$$\text{MMF} = NI \text{ Ampere turns}$$

where: N *= the number of turns on the coil,* I *= the current flowing in the coil.*

Magnetic field strength H, requires the active length of the coil to be included:

$$H = \frac{NI}{l}$$

where: l *= active length of the coil,* H *= magnetic field strength.*

In order to convert this to flux density B, the permeability of the pole shoe must be included:

$$B = H\mu_0\mu_r$$

where: μ_0 *= permeability of free space (4 x* 10^{-7} *Henry/meter),* μ_r *= relative permeability of the core to free space.*

To calculate power consumed is a simple task using the formula:

$$P = T\omega$$

where: P *= power in watts,* T *= torque in Nm,* ω *= angular velocity in rad/s.*

Here is a simple example of the use of this formula. An engine requires a minimum cranking speed of 100 rpm and the required torque to achieve this is 9.6 Nm.

At a 10:1 ring gear to pinion ratio this will require a 1000 rpm starter speed (n). To convert this to rad/s:

$$\omega = \frac{2\pi n}{60}$$

This works out to 105 rad/s

$$P = T\omega$$

$$9.6 \times 105 = 1000 \text{ W or } 1 \text{ kW}$$

7.4.2 Efficiency

The efficiency of most starter motors is in the order of 60%.

Efficiency = Power out / Power in (× 100%)

So, to calculate the required input power: 1 kW/60% = 1.67 kW

The main losses, which cause this, are iron losses, copper losses and mechanical losses.

Iron losses are due to hysteresis loss caused by changes in magnetic flux, and also due to induced eddy currents in the iron parts of the motor. Copper losses are caused by the resistance of the windings; sometimes called I^2R losses. Mechanical losses include friction and windage (air) losses.

Using the previous example of a 1 kW starter it can be seen that at an efficiency of 60% this motor will require a supply of about 1.7 kW.

From a nominal 12 V supply and allowing for battery volt drop, a current in the order of 170 A will be required to achieve the necessary power.

CHAPTER 8

Ignition

8.1 Ignition system fundamentals

8.1.1 Functional requirements

The fundamental purpose of the ignition system is to supply a spark inside the cylinder, near the end of the compression stroke, to ignite the compressed charge of air-fuel vapour.

For a spark to jump across an air gap of 1 mm under normal atmospheric conditions (1 bar), a voltage of 3–4 kV is required. For a spark to jump across a similar gap in an engine cylinder, having a compression ratio of 8 : 1, approximately 8–10 kV is required. For higher compression ratios and weaker mixtures, a voltage up to 30 kV may be necessary. The ignition system has to transform the normal battery voltage of 12 V to approximately 8–20 kV and, in addition, has to deliver this high voltage to the right cylinder, at the right time. Some ignition systems will supply up to 40 kV to the spark plugs.

Key fact

For higher compression ratios and weaker mixtures, a voltage up to 30 kV may be necessary to jump a 1mm gap.

Conventional ignition is the forerunner of the more advanced systems controlled by electronics. It is worth mentioning at this stage that the fundamental operation of most ignition systems is very similar. One winding of a coil is switched on and off causing a high voltage to be induced in a second winding. A coil-ignition system is composed of various components and sub-assemblies, the actual design and construction of which depend mainly on the engine with which the system is to be used.

When considering the design of an ignition system many factors must be taken into account, the most important of these being:

- Combustion chamber design.
- Air–fuel ratio.
- Engine speed range.
- Engine load.
- Engine combustion temperature.
- Intended use.
- Emission regulations.

8.1.2 Generation of high tension

If two coils (known as the primary and secondary) are wound on to the same iron core then any change in magnetism of one coil will induce a voltage into the other. This happens when a current is switched on and off to the primary coil. If the number of turns of wire on the secondary coil is more than the primary, a higher voltage can be produced. This is called transformer action and is the principle of the ignition coil.

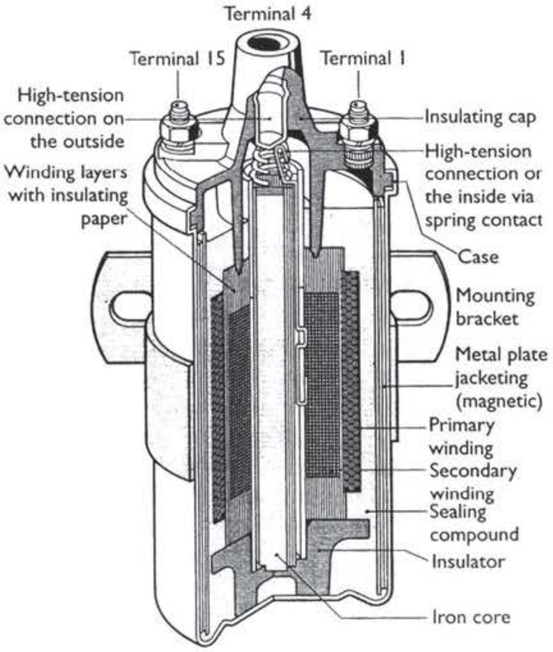


Figure 8.1 Earlier type ignition coil

The value of this 'mutually induced' voltage depends upon:

- The primary current.
- The turns ratio between the primary and secondary coils.
- The speed at which the magnetism changes.

Figure 8.1 shows a typical ignition coil in section. The two windings are wound on a laminated iron core to concentrate the magnetism. Some coils are oil filled to assist with cooling.

8.1.3 Advance angle (timing)

Key fact

The ideal ignition timing is dependent on two main factors, engine speed and engine load.

For optimum efficiency the ignition advance angle should be such as to cause the maximum combustion pressure to occur about 10° after top dead centre (TDC). The ideal ignition timing is dependent on two main factors, engine speed and engine load. An increase in engine speed requires the ignition timing to be advanced. The cylinder charge, of air–fuel mixture, requires a certain time to burn (normally about 2 ms). At higher engine speeds the time taken for the piston to travel the same distance reduces. Advancing the time of the spark ensures full burning is achieved.

A change in timing due to engine load is also required as the weaker mixture used on low load conditions burns at a slower rate. In this situation, further ignition advance is necessary. Greater load on the engine requires a richer mixture, which burns more rapidly. In this case some retardation of timing is necessary. Overall, under any condition of engine speed and load an ideal advance angle is required to ensure maximum pressure is achieved in the cylinder just after top dead centre. The ideal advance angle may be further refined by engine temperature and any risk of detonation.

Spark advance is achieved in a number of ways. The simplest of these being the mechanical system comprising a centrifugal advance mechanism and a vacuum (load sensitive) control unit. Manifold vacuum is almost inversely proportional to the engine load. I prefer to consider manifold pressure, albeit less than atmospheric pressure, as the manifold absolute pressure (MAP) is proportional to engine load.

Digital ignition systems may adjust the timing in relation to the temperature as well as speed and load. The values of all ignition timing functions are combined either mechanically or electronically in order to determine the ideal ignition point.

The energy storage takes place in the ignition coil. The energy is stored in the form of a magnetic field. To ensure the coil is charged before the ignition point a dwell period is required. Ignition timing is at the end of the dwell period.

8.1.4 Fuel consumption and exhaust emissions

The ignition timing has a significant effect on fuel consumption, torque, drivability and exhaust emissions. The three most important pollutants are hydrocarbons (HC), carbon monoxide (CO) and nitrogen oxides (NOx).

The HC emissions increase as timing is advanced. NOx emissions also increase with advanced timing due to the higher combustion temperature. CO changes very little with timing and is mostly dependent on the air–fuel ratio.

As is the case with most alterations of this type, a change in timing to improve exhaust emissions will increase fuel consumption. With the leaner mixtures now prevalent, a larger advance is required to compensate for the slower burning rate. This will provide lower consumption and high torque but the mixture must be controlled accurately to provide the best compromise with regard to the emission problem. Figure 8.2 shows the effect of timing changes on emissions, performance and consumption.

Key fact

The ignition timing has a significant effect on fuel consumption, torque, drivability and exhaust emissions.

8.1.5 Contact breaker ignition

Contact breaker ignition (Figure 8.3) is not now used on new vehicles but there are many early cars out their still! The main components of this system, together with some that are still used, are outlined in Table 8.1.

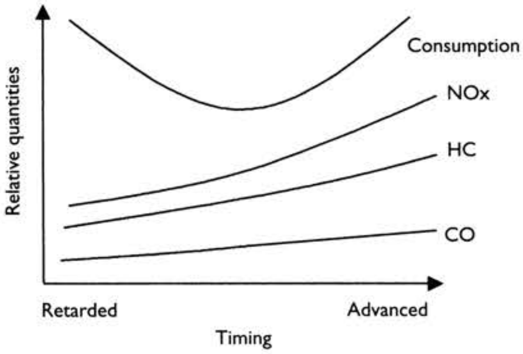


Figure 8.2 Effect of changes in ignition timing at a fixed engine speed

Figure 8.3 Contact breaker system

Table 8.1 Traditional ignition components

Spark plug	Seals electrodes for the spark to jump across in the cylinder. Must withstand very high voltages, pressures and temperatures.
Ignition coil	Stores energy in the form of magnetism and delivers it to the distributor via the HT lead. Consists of primary and secondary windings.
Ignition switch	Provides driver control of the ignition system and is usually also used to cause the starter to crank.
Contact breakers (breaker points)	Switches the primary ignition circuit on and off to charge and discharge the coil. The contacts are operated by a rotating cam in the distributor.
Capacitor (condenser)	Suppresses most of the arcing as the contact breakers open. This allows for a more rapid break of primary current and hence a more rapid collapse of coil magnetism which produces a higher voltage output.
Distributor	Directs the spark from the coil to each cylinder in a pre-set sequence.
Plug leads	Thickly insulated wires to connect the spark from the distributor to the plugs.
Centrifugal advance	Changes the ignition timing with engine speed. As speed increases the timing is advanced.
Vacuum advance	Changes timing depending on engine load. On conventional systems the vacuum advance is most important during cruise conditions.

8.1.6 Plug leads

Definition

High tension (HT): High voltage.

High tension (HT) is just an old fashioned way of saying high voltage. HT components and systems, must meet or exceed stringent ignition product requirements, such as:

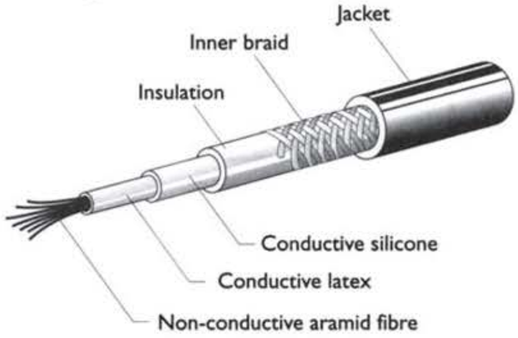
- Insulation to withstand 50 000 V.
- Temperatures from 40 °C to 260 °C (40 °F to 500 °F).
- Radio frequency interference suppression.
- 160 000 km (100 000 mile) product life.
- Resistance to ozone, corona, and fluids.
- 10-year durability.

Delphi produces a variety of cable types that meet the increased energy needs of lean-burn engines without emitting electromagnetic interference (EMI). The cable products offer metallic and non-metallic cores, including composite, high-temperature resistive and wire-wound inductive cores. Conductor construction includes copper, stainless steel, Delcore, CHT, and wire-wound. Jacketing materials include organic and inorganic compounds, such as CPE, EPDM and silicone. Figure 8.4 shows the construction of these leads. Table 8.2 summarizes some of the materials used for different temperature ranges.

8.1.7 Ignition coil cores

Most ignition coil cores are made of laminated iron. The iron is ideal as it is easily magnetized and demagnetized. The laminations reduce eddy currents, which

Resistive ignition cable



Wire-wound ignition cable

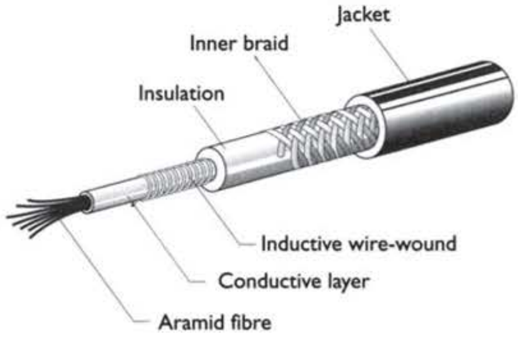


Figure 8.4 Ignition plug leads

Table 8.2 Materials used for ignition HT leads/cables

Operating Temperature (Continuous)	110 °C	175 °C	232 °C
Terminals	zinc plated	phosphor bronze or stainless	stainless
Boot Material	EPDM or silicone	silicone	high-temperature silicone
Jacket	CPE	silicone	silicone
Insulation	EPDM	EPDM	silicone
Conductor	Delcore copper or stainless	Delcore or CHT	CHT or wire-wound core

cause inefficiency due to the heating effect (iron losses). If thinner laminations or sheets are used, then the better the performance.

Powder metal is now in use as coil cores. This reduces eddy currents to a minimum but the density of the magnetism is decreased. Overall, however, this produces a more efficient and higher output ignition coil. Developments are continuing and the flux density problem is about to be solved, giving rise to even more efficient components.

8.2 Electronic ignition

8.2.1 Introduction

Electronic ignition is now fitted to almost all spark ignition vehicles. This is because the conventional mechanical system has some major disadvantages.

- Mechanical problems with the contact breakers, not the least of which is the limited lifetime.
- Current flow in the primary circuit is limited to about 4 A or damage will occur to the contacts – or at least the lifetime will be seriously reduced.
- Legislation requires stringent emission limits, which means the ignition timing must stay in tune for a long period of time.
- Weaker mixtures require more energy from the spark to ensure successful ignition, even at very high engine speed.

These problems can be overcome by using a power transistor to carry out the switching function and a pulse generator to provide the timing signal. Very early forms of electronic ignition used the existing contact breakers as the signal provider. This was a step in the right direction but did not overcome all the mechanical limitations, such as contact bounce and timing slip. All modern systems are constant energy, ensuring high performance ignition even at high engine speed. Figure 8.5 shows the circuit of a traditional electronic ignition system.

Definition

Dwell: When applied to ignition is a measure of the time during which the ignition coil is charging, in other words when the primary current is flowing.

8.2.2 Constant dwell systems

The term 'dwell' when applied to ignition is a measure of the time during which the ignition coil is charging, in other words when the primary current is flowing. The dwell in conventional systems was simply the time during which the contact breakers were closed. This is now often expressed as a percentage of one charge–discharge cycle. Constant dwell electronic ignition systems have now been replaced almost without exception by constant energy systems discussed in the next section.

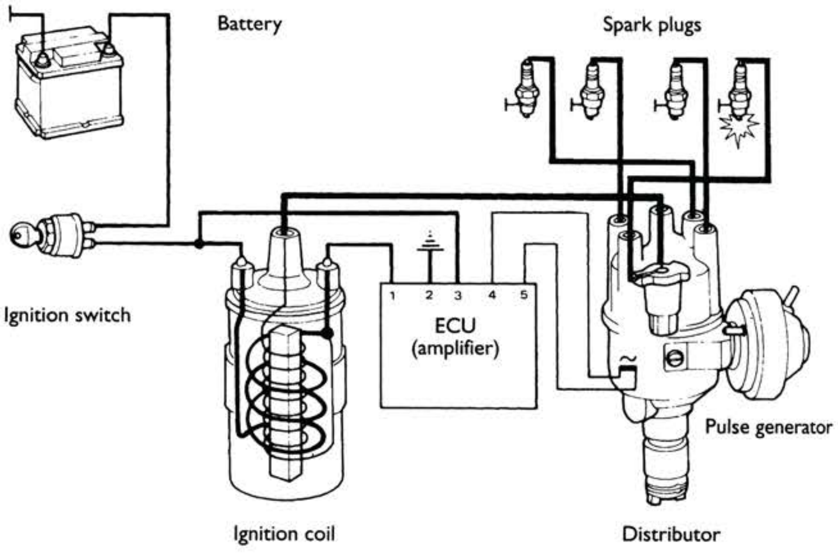


Figure 8.5 Early electronic ignition system

Whilst this was a very good system in its time, constant dwell still meant that at very high engine speeds, the time available to charge the coil could only produce a lower power spark. Note that as engine speed increases, the dwell angle or dwell percentage remains the same but the actual time is reduced.

8.2.3 Constant energy systems

In order for a constant energy electronic ignition system to operate, the dwell must increase with engine speed. This will only be of benefit, however, if the ignition coil can be charged up to its full capacity, in a very short time (the time available for maximum dwell at the highest expected engine speed). To this end, constant energy coils are very low resistance and low inductance. Typical resistance values are less than 1 (often 0.5). Constant energy means that, within limits, the energy available to the spark plug remains constant under all operating conditions.

An energy value of about 0.3 mJ is all that is required to ignite a static stoichiometric mixture. In the case of lean or rich mixtures together with high turbulence, energy values in the region of 3–4 mJ are necessary. This has made constant energy ignition essential on all of today's vehicles in order to meet the expected emission and performance criteria. Figure 8.6 is a block diagram of a closed loop constant energy ignition system. The earlier open loop systems are the same but without the current detection feedback section.

Due to the high energy nature of constant energy ignition coils, the coil cannot be allowed to remain switched on for more than a certain time. This is not a problem when the engine is running, as the variable dwell or current limiting circuit prevents the coil overheating. Some form of protection must be provided for, however, when the ignition is switched on but the engine is not running. This is known as the 'stationary engine primary current cut off'.

8.2.4 Hall Effect pulse generator

The operating principle of the Hall Effect is discussed in Chapter 2. The Hall Effect distributor has become very popular with many manufacturers. Figure 8.7 shows a typical distributor with a Hall Effect sensor.

As the central shaft of the distributor rotates, the vanes attached under the rotor arm alternately cover and uncover the Hall chip. The number of vanes corresponds to the number of cylinders. In constant dwell systems the dwell is determined by the width of the vanes. The vanes cause the Hall chip to be alternately in and out of a magnetic field. The result of this is that the device will produce almost a square wave output, which can then easily be used to switch further electronic circuits. The three terminals on the distributor are marked '–, 0, + ', the terminals –

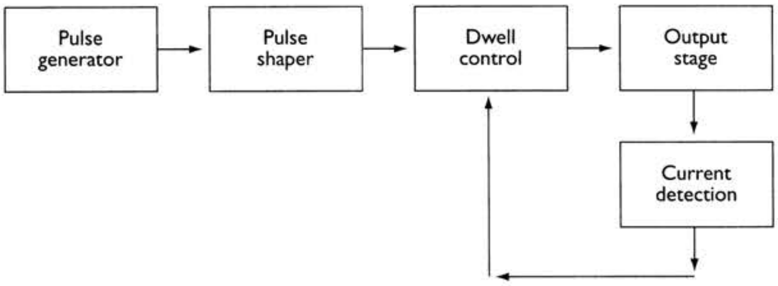


Figure 8.6 Constant energy ignition

Figure 8.7 Ignition distributor with Hall generator

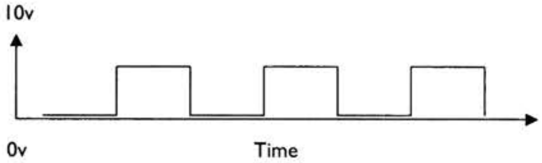


Figure 8.8 A Hall effective sensor output will switch between 0 V and 7 V

and +, are for a voltage supply and terminal '0' is the output signal. Typically, the output from a Hall Effect sensor in a distributor will switch between 0 V and about 7 V as shown in Figure 8.8. The supply voltage is taken from the ignition ECU and, on some systems, is stabilized at about 10 V to prevent changes to the output of the sensor when the engine is being cranked.

Hall Effect distributors are very common due to the accurate signal produced and long term reliability. They are suitable for use on both constant dwell and constant energy systems. Operation of a Hall effect pulse generator can easily be tested with a DC voltmeter or a logic probe. Note that tests must not be carried out using an ohmmeter as the voltage from the meter can damage the Hall chip.

8.2.5 Inductive pulse generator

Inductive pulse generators use the basic principle of induction to produce a signal typical of the one shown in Figure 8.9. Many forms exist but all are based around a coil of wire and a permanent magnet.

The example distributor shown in Figure 8.10 has the coil of wire wound on the pick-up and, as the reluctor rotates, the magnetic flux varies due to the peaks on the reluctor. The number of peaks, or teeth, on the reluctor corresponds to the number of engine cylinders. The gap between the reluctor and pick-up can be important and manufacturers have recommended settings.

8.2.6 Other pulse generators

An early system known as transistor assisted contacts (TAC) was used, where the contact breakers were used as the trigger. The only other technique, which has been used on a reasonable scale, is the optical pulse generator. This involved a focused beam of light from a light emitting diode (LED) and a photo-transistor.

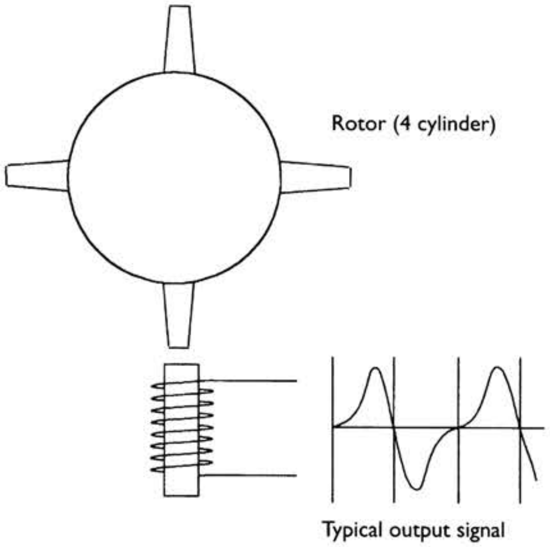


Figure 8.9 Inductive pulse generators use the basic principle of induction to produce a signal

Figure 8.10 Inductive pulse generator in a distributor

The beam of light is interrupted by a rotating vane, which provides a switching output in the form of a square wave. The most popular use for this system is in the after-market as a replacement for conventional contact breakers. Figure 8.11 shows the basic principle of an optical pulse generator; note how the beam is focused to ensure accurate switching.

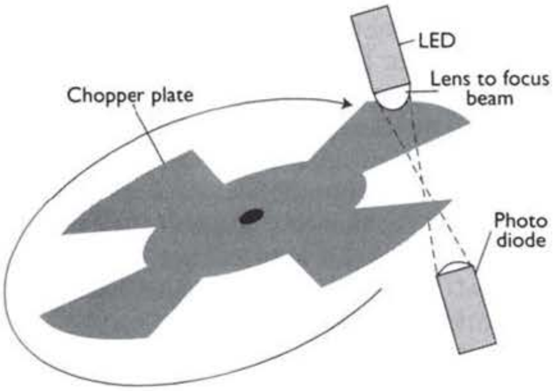


Figure 8.11 Basic principles of an optical pulse generator

8.2.7 Dwell angle control (open loop)

Figure 8.12 shows a circuit diagram of a transistorized ignition module. For the purposes of explaining how this system works, the pulse generator is the inductive type. To understand how the dwell is controlled, an explanation of the whole circuit is necessary.

The first part of the circuit is a voltage stabilizer to prevent damage to any components and to allow known voltages for charging and discharging the capacitors. This circuit consists of ZD_1 and R_1.

The alternating voltage coming from the inductive-type pulse generator must be reshaped into square-wave type pulses in order to have the correct effect in the trigger box. The reshaping is done by an electronic threshold switch known as a Schmitt trigger. This circuit is termed a pulse shaping circuit because of its function in the trigger box.

The pulse shaping circuit starts with D_4, a silicon diode which, due to its polarity, will only allow the negative pulses of the alternating control voltage to reach the base of transistor T_1. The induction-type pulse generator is loaded only in the negative phase of the alternating control voltage because of the output of energy. In the positive phase, on the other hand, the pulse generator is not loaded. The negative voltage amplitude is therefore smaller than the positive amplitude.

As soon as the alternating control voltage, approaching from negative values, exceeds a threshold at the pulse shaping circuit input, transistor T_1 switches off and prevents current passing. The output of the pulse-shaping circuit is 'currentless' for a time (anti-dwell?). This switching state is maintained until the alternating control voltage, now approaching from positive values, drops below the threshold voltage. Transistor T_1 now switches off. The base of T_2 becomes

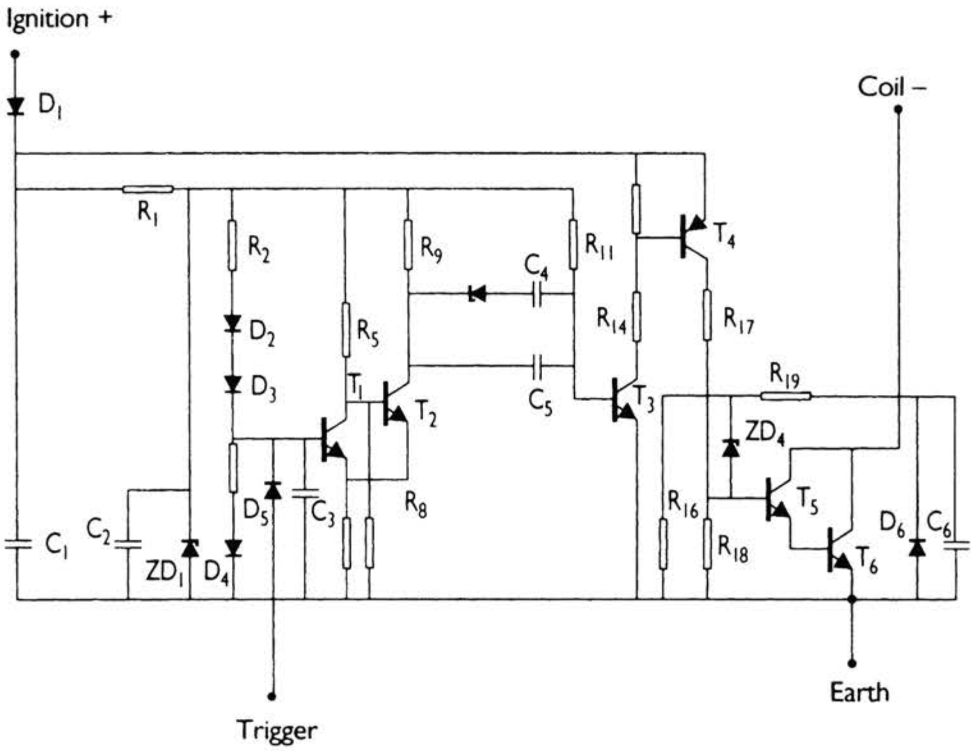


Figure 8.12 Circuit diagram of a transistorized ignition module

positive via R_5 and T_2 is on. This alternation – T_1 on/T_2 off or T_1 off/T_2 on – is typical of the Schmitt trigger and the circuit repeats this action continuously. Two series-connected diodes, D_2 and D_3, are provided for temperature compensation. The diode D1 is for reverse polarity protection.

The energy stored in the ignition coil can be put to optimum use with the help of the dwell section in the trigger box. The result is that sufficient high voltage is available for the spark at the spark plug under any operating condition of the engine. The dwell control specifies the start of the dwell period. The beginning of the dwell period (when T_3 switches on), is also the beginning of a rectangular current pulse that is used to trigger the transistor T_4, which is the driver stage. This in turn switches on the output stage.

A timing circuit using RC elements is used to provide a variable dwell. This circuit alternately charges and discharges capacitors by way of resistors. This is an open loop dwell control circuit because the combination of the resistors and capacitors provides a fixed time relationship as a function of engine speed.

The capacitor C_5 and the resistors R_9 and R_{11} form the RC circuit. When transistor T_2 is switched off, the capacitor C_5 will charge via R_9 and the base emitter of T_3. At low engine speed the capacitor will have time to charge to almost 12 V. During this time T_3 is switched on and, via T_4, T_5 and T_6, so is the ignition coil. At the point of ignition T_2 switches on and capacitor C5 can now discharge via R_{11} and T_2. T_3 remains switched off all the time C_5 is discharging. It is this discharge time (which is dependent on how much C_5 had been charged), that delays the start of the next dwell period. Capacitor C_5 finally begins to be charged, via R11 and T_2, in the opposite direction and, when it reaches about 12 V, T_3 will switch back on. T_3 remains on until T_2 switches off again. As the engine speed increases, the charge time available for capacitor C_5 decreases. This means it will only reach a lower voltage and hence will discharge more quickly. This results in T_3 switching on earlier and hence a longer dwell period is the result.

The current from this driver transistor drives the power output stage (a Darlington pair). In this Darlington circuit the current flowing into the base of transistor T_5 is amplified to a considerably higher current, which is fed into the base of the transistor T_6. The high primary current can then flow through the ignition coil via transistor T_6. The primary current is switched on the collector side of this transistor. The Darlington circuit functions as one transistor and is often described as the power stage.

Components not specifically mentioned in this explanation are for protection against back emf (ZD_4, D_6) from the ignition coil and to prevent the dwell becoming too small (ZD_2 and C_4). A trigger box for Hall Effect pulse generators functions in a similar manner to the above description. The hybrid ignition trigger boxes are considerably smaller than those utilizing discrete components. Figure 8.13 is a picture of a typical complete unit.

8.2.8 Current limiting and closed loop dwell

Primary current limiting ensures no damage can be caused to the system by excessive primary current, but also forms a part of a constant energy system.

The primary current is allowed to build up to its pre-set maximum as soon as possible and then be held at this value. The value of this current is calculated and then pre-set during construction of the amplifier module. This technique, when combined with dwell angle control, is known as closed loop control as the actual value of the primary current is fed back to the control stages.

Figure 8.13 Transistorized ignition module

A very low resistance, high power precision resistor is used in this circuit. The resistor is connected in series with the power transistor and the ignition coil. A voltage sensing circuit connected across this resistor will be activated at a pre-set voltage (which is proportional to the current), and will cause the output stage to hold the current at a constant value. Figure 8.14 shows a block diagram of a closed loop dwell control system.

Stationary current cut-off is for when the ignition is on but the engine is not running. This is achieved in many cases by a simple timer circuit, which will cut the output stage after about one second.

8.2.9 Capacitor discharge ignition

Capacitor discharge ignition (CDI) has been in use for many years by Saab, on some models of the Porsche 911 and some Ferrari models. It is now still used by some coil on plug (COP) ignition systems.

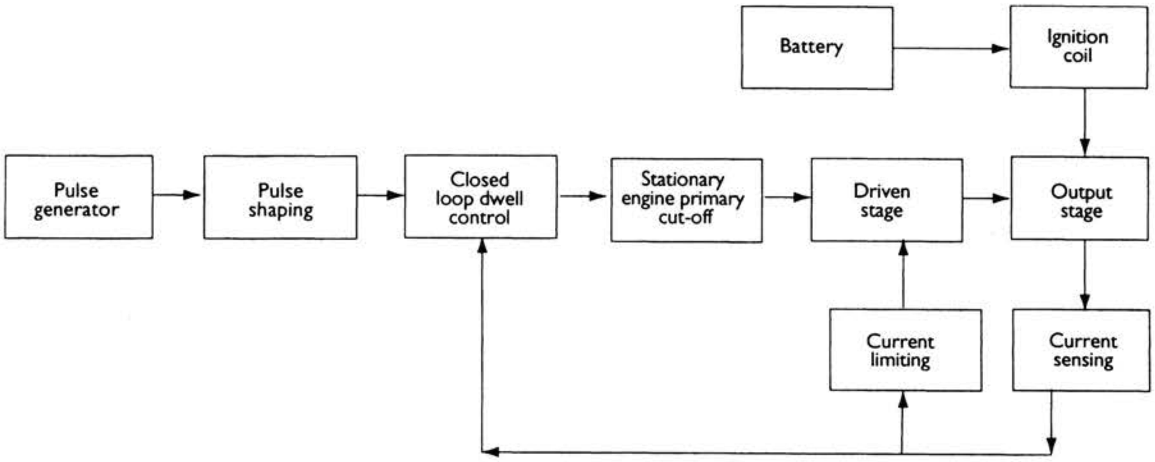


Figure 8.14 Closed loop dwell control system

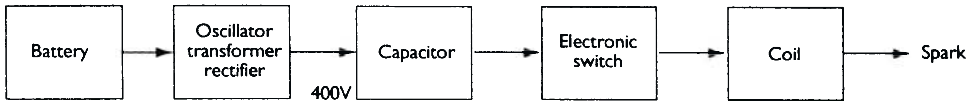


Figure 8.15 CDI system

Figure 8.15 shows a block diagram of the CDI system. The CDI works by first stepping up the battery voltage to about 400 V (DC), using an oscillator and a transformer, followed by a rectifier. This high voltage is used to charge a capacitor. At the point of ignition the capacitor is discharged through the primary winding of a coil, often by use of a thyristor. This rapid discharge through the coil primary will produce a very high voltage output from the secondary winding. This voltage has a very fast rise time compared with a more conventional system. Typically, the rise time for CDI is 3–10 kV/s as compared with the pure inductive system, which is 300–500 V/s. This very fast rise time and high voltage will ensure that even a carbon- or oil-fouled plug will be fired. The disadvantage, however, is that the spark duration is short, which can cause problems particularly during starting. This is often overcome by providing the facility for multi-sparking. However, when used in conjunction with direct ignition (one coil for each plug) the spark duration is acceptable.

8.3 Electronic spark advance

8.3.1 Overview

Constant energy electronic ignition was a major step forwards and is still used on most vehicles. However, limitations lay in still having to rely upon mechanical components for speed and load advance characteristics. In many cases these did not match ideally the requirements of the engine.

Electronic spark advance (ESA) ignition systems have a major difference compared with earlier systems, in that they operate digitally. Information about the operating requirements of a particular engine is programmed into the memory inside the electronic control unit. The data for storage in ROM are obtained from rigorous testing on an engine dynamometer and from further development work on the vehicle under various operating conditions.

Key fact

Electronic spark advance (ESA) ignition systems have data on the operating requirements of a particular engine programmed into memory.

ESA ignition has several advantages.

- The ignition timing can be accurately matched to the individual application under a range of operating conditions.
- Other control inputs can be utilized such as coolant temperature and ambient air temperature.
- Starting is improved and fuel consumption is reduced, as are emissions, and idle control is better.
- Other inputs can be taken into account such as engine knock.
- The number of wearing components in the ignition system is considerably reduced.

ESA (also referred to as programmed ignition), can be a separate system but is now most likely to be included as part of the full engine management system.

8.3.2 Sensors and input information

A typical early ESA system is shown as Figure 8.16. In order for the ECU to calculate suitable timing and dwell outputs, certain input information is required.

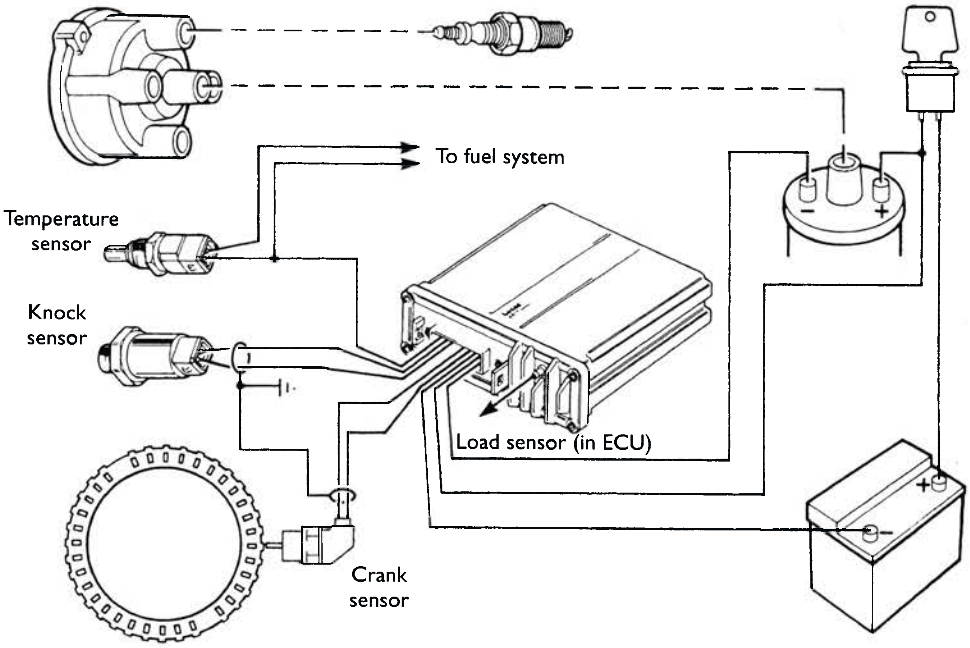


Figure 8.16 Programmed ignition systems

Engine speed and position – crankshaft sensor

This sensor is a reluctance sensor positioned as shown in Figure 8.17. The device consists of a permanent magnet, a winding and a soft iron core. It is mounted in proximity to a reluctor disc. The disc has 34 teeth, spaced at 10° intervals around the periphery of the disc. It has two teeth missing, 180° apart, at a known position before TDC (BTDC). Many manufacturers use this technique with minor differences. As a tooth from the reluctor disc passes the core of the sensor, the reluctance of the magnetic circuit is changed. This induces a voltage in the winding, the frequency of the waveform being proportional to the engine speed. The missing tooth causes a ‘missed’ output wave and hence the engine position can be determined.

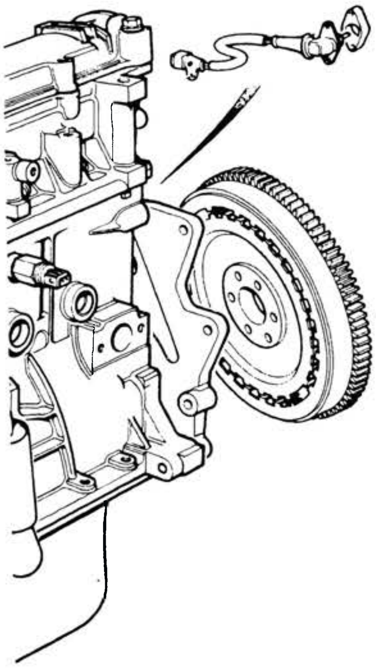


Figure 8.17 Position of a programmed ignition crankshaft sensor

Engine load – manifold absolute pressure sensor

Engine load is proportional to manifold pressure in that high load conditions produce high pressure and lower load conditions – such as cruise – produce lower pressure. Load sensors are therefore pressure transducers. They are either mounted in the ECU or as a separate unit, and are connected to the inlet manifold with a pipe. The pipe often incorporates a restriction to damp out fluctuations and a vapour trap to prevent petrol fumes reaching the sensor.

Engine temperature – coolant sensor

Coolant temperature measurement is carried out by a simple thermistor, and in many cases the same sensor is used for the operation of the temperature gauge and to provide information to the fuel control system. A separate memory map is used to correct the basic timing settings. Timing may be retarded when the engine is cold to assist in more rapid warm up.

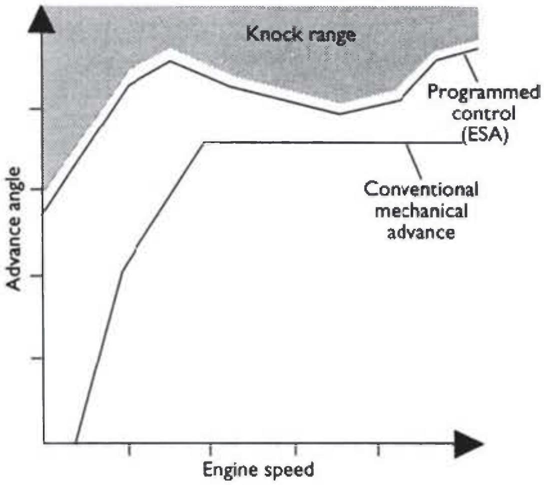


Figure 8.18 Ideal timing angle for an engine is close to the knock limit

Detonation – knock sensor

Combustion knock can cause serious damage to an engine if sustained for long periods. This knock, or detonation, is caused by over-advanced ignition timing. At variance with this is that an engine will, in general, run at its most efficient when the timing is advanced as far as possible. To achieve this, the data stored in the basic timing map will be as close to the knock limit of the engine as possible (see Figure 8.18). The knock limit is also known as the detonation border line (DBL). The knock sensor provides a margin for error. The sensor itself is an accelerometer often of the piezoelectric type. It is fitted in the engine block between cylinders two and three on in-line four-cylinder engines. Vee engines require two sensors, one on each side. The ECU responds to signals from the knock sensor in the engine's knock window for each cylinder – this is often just a few degrees each side of TDC. This prevents clatter from the valve mechanism being interpreted as knock. The signal from the sensor is also filtered in the ECU to remove unwanted noise. If detonation is detected, the ignition timing is retarded on the fourth ignition pulse after detection (four-cylinder engine) in steps until knock is no longer detected.

Key fact

Combustion knock can cause serious damage to an engine if sustained for long periods.

The steps vary between manufacturers, but about 2° is typical. The timing is then advanced slowly in steps of, say 1°, over a number of engine revolutions, until the advance required by memory is restored. This fine control allows the engine to be run very close to the knock limit without risk of engine damage.

Battery voltage

Correction to dwell settings is required if the battery voltage falls, as a lower voltage supply to the coil will require a slightly larger dwell figure. This information is often stored in the form of a dwell correction map.

8.3.3 Electronic control unit

As the sophistication of systems has increased, the information held in the memory chips of the ECU has also increased. The earlier versions of a programmed ignition system achieved accuracy in ignition timing of 1.8° whereas a conventional distributor is 8°. The information, which is derived from dynamometer tests as well as running tests in the vehicle, is stored in ROM. The basic timing map consists of the correct ignition advance for 16 engine speeds and 16 engine load conditions. This is shown in Figure 8.19 using a cartographic representation.

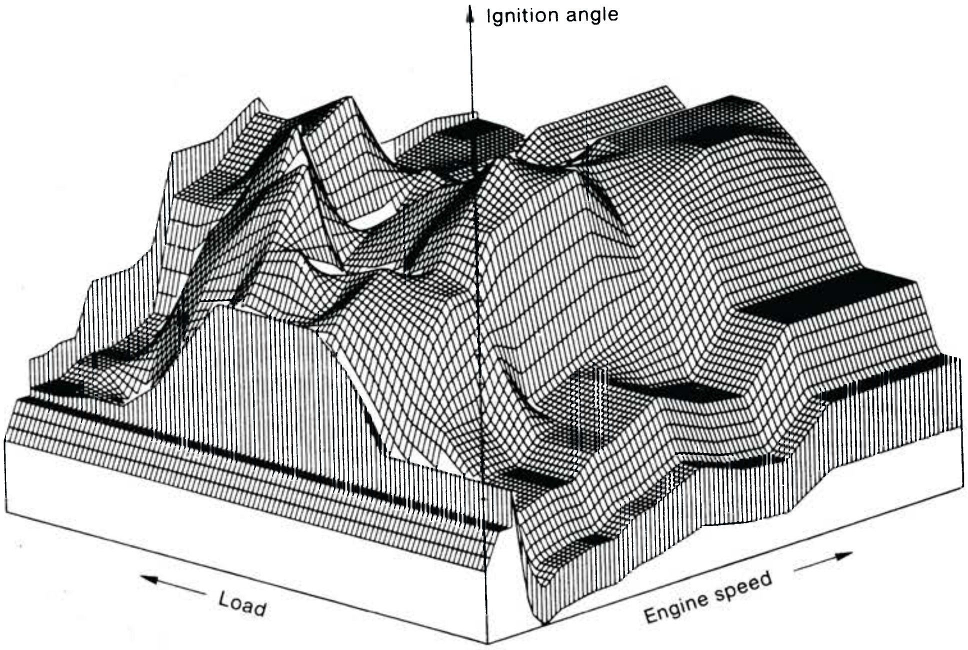


Figure 8.19 Cartographic map representing how ignition timing is stored in the ECU

A separate three-dimensional map is used that has eight (or more on modern systems) speed and eight temperature sites. This is used to add corrections for engine coolant temperature to the basic timing settings. This improves drivability and can be used to decrease the warm-up time of the engine. The data are also subjected to an additional load correction below 70 °C. Figure 8.20 shows a flow chart representing the logical selection of the optimum ignition setting. Note that the ECU will also make corrections to the dwell angle, both as a function of engine speed to provide constant energy output and corrections due to changes in battery voltage. A lower battery voltage will require a slightly longer dwell and a higher voltage a slightly shorter dwell.

Typical of most ‘computer’ systems, a block diagram as shown in Figure 8.21 can represent the programmed ignition ECU. Input signals are processed and the data provided are stored in RAM. The program and pre-set data are held in ROM. In these systems a microcontroller is used to carry out the fetch-execute sequences demanded by the program. Information, which is collected from the sensors, is converted to a digital representation in an A/D circuit.

A flow chart used to represent the program held in ROM, inside the ECU, is shown in Figure 8.22. A Windows® shareware program that simulates the ignition system (as well as many other systems) is available for downloading from the book web site (details in Preface).

Ignition output

The output of a system, such as this programmed ignition, is very simple. The output stage, in common with most electronic ignitions, consists of a heavy-duty transistor that forms part of, or is driven by, a Darlington pair. This is simply to allow the high ignition primary current to be controlled. The switch off point of the coil will control ignition timing and the switch on point will control the dwell period.

HT distribution

The high tension distribution is similar to a more conventional system. The rotor arm however is mounted on the end of the camshaft with the distributor

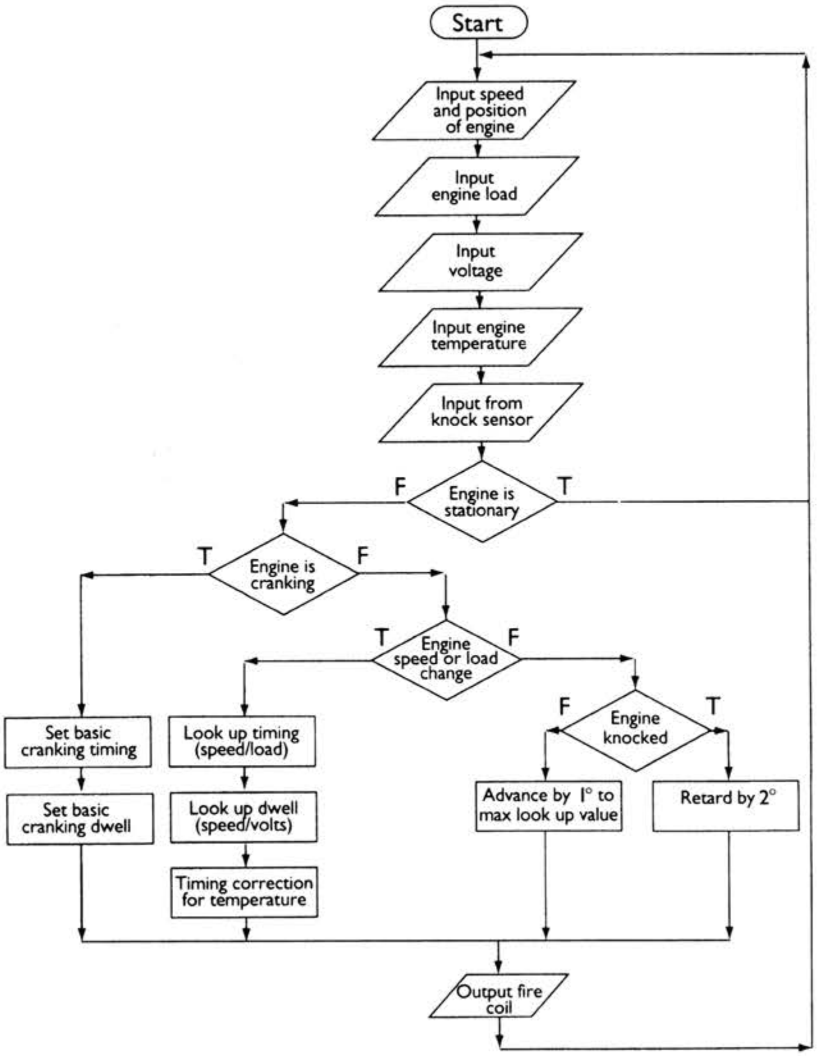


Figure 8.20 Ignition calculation flow diagram

cap positioned over the top. The material used for the cap is known as Velox, which is similar to the epoxy type but has better electrical characteristics – it is less prone to tracking, for example. The distributor cap is mounted on a base plate made of Crasline which, as well as acting as the mounting point, prevents any oil that leaks from the camshaft seal fouling the cap and rotor arm. Another important function of the mounting plate is to prevent the

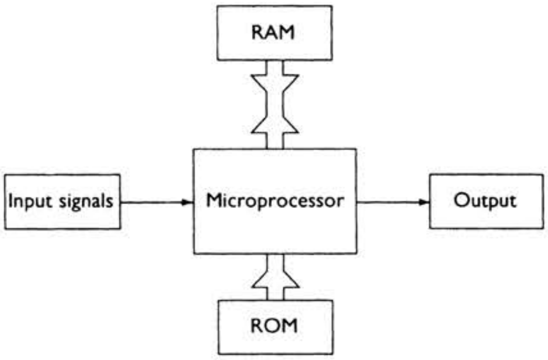


Figure 8.21 Typical of most 'computer' systems the programmed ignition ECU can be represented by a block diagram

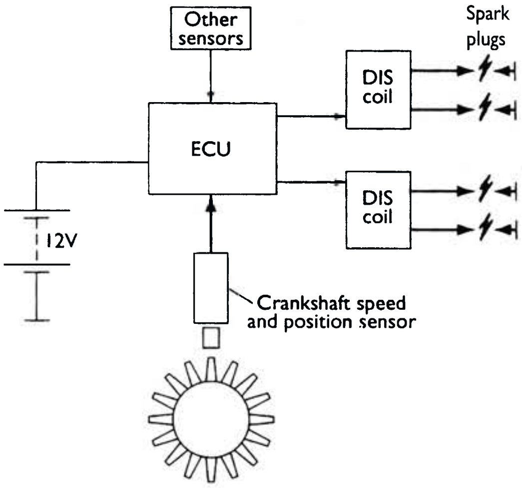


Figure 8.22 DIS ignition system

build-up of harmful gases such as ozone and nitric oxide by venting them to the atmosphere. These gases are created by the electrolytic action of the spark as it jumps the air gap between the rotor arm and the cap segment. The rotor arm is also made of Crasline and is reinforced with a metal insert to relieve fixing stresses.

8.4 Distributorless ignition

8.4.1 Principle of operation

Key fact

Distributorless ignition outputs to the spark plugs without the need for an HT distributor.

Distributorless ignition has all the features of programmed ignition systems but, by using a special type of ignition coil, outputs to the spark plugs without the need for an HT distributor.

The system is generally only used on four-cylinder engines because the control system becomes more complex for higher numbers. The basic principle is that of the ‘lost spark’. The distribution of the spark is achieved by using two double-ended coils, which are fired alternately by the ECU. The timing is determined from a crankshaft speed and position sensor as well as load and other corrections. When one of the coils is fired, a spark is delivered to two engine cylinders, either 1 and 4, or 2 and 3. The spark delivered to the cylinder on the compression stroke will ignite the mixture as normal. The spark produced in the other cylinder will have no effect, as this cylinder will be just completing its exhausted stroke.

Key fact

Because of the low compression and the exhaust gases in the ‘lost spark’ cylinder, the voltage used for the spark to jump the gap is only about 3 kV.

Because of the low compression and the exhaust gases in the ‘lost spark’ cylinder, the voltage used for the spark to jump the gap is only about 3 kV. This is similar to the more conventional rotor arm to cap voltage. The spark produced in the compression cylinder is therefore not affected.

An interesting point here is that the spark on one of the cylinders will jump from the earth electrode to the spark plug centre. Many years ago this would not have been acceptable, as the spark quality when jumping this way would not have been as good as when it jumps from the centre electrode. However, the energy available from modern constant energy systems will produce a spark of suitable quality in either direction. Figure 8.23 shows the layout of the distributorless ignition system (DIS) system.

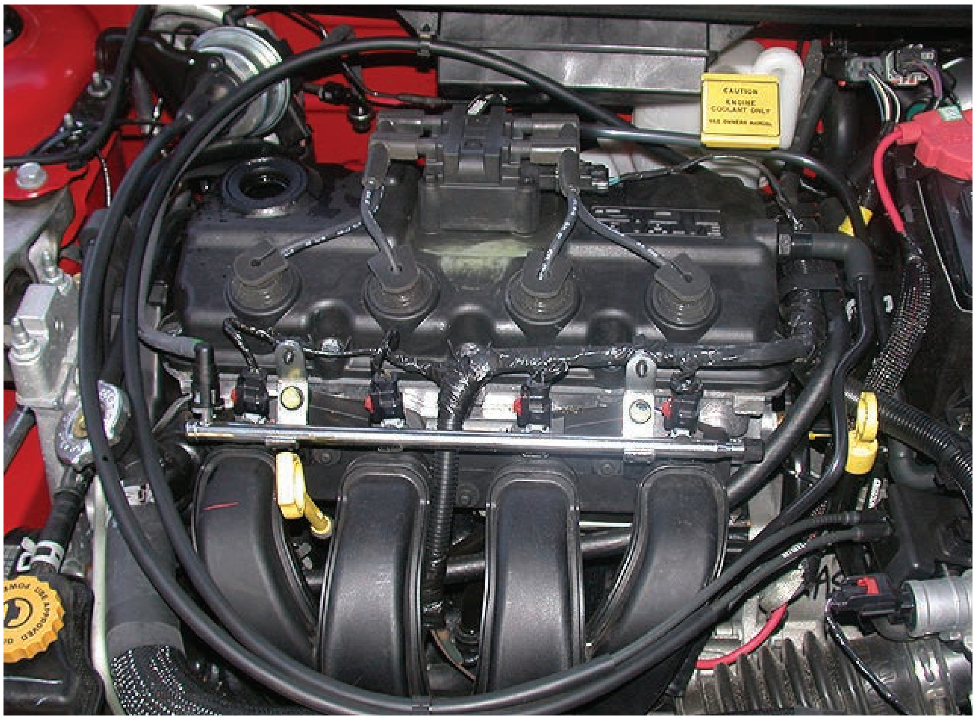


Figure 8.23 DIS coil on a car

8.4.2 System components

The DIS system consists of three main components: the electronic module, a crankshaft position sensor and the DIS coil. In many systems a manifold absolute pressure sensor is integrated in the module. The module functions in much the same way as has been described for the previously described electronic spark advance system.

The crankshaft position sensor is similar in operation to the one described in the previous section. It is again a reluctance sensor and is positioned against the front of the flywheel or against a reluctor wheel just behind the front crankshaft pulley. The tooth pattern consists of 35 teeth. These are spaced at 10° intervals with a gap where the 36th tooth would be. The missing tooth is positioned at 90° BTDC for cylinders number 1 and 4. This reference position is placed a fixed number of degrees before top dead centre, in order to allow the timing or ignition point to be calculated as a fixed angle after the reference mark.

The low tension winding is supplied with battery voltage to a centre terminal. The appropriate half of the winding is then switched to earth in the module. The high tension windings are separate and are specific to cylinders 1 and 4, or 2 and 3. Figure 8.24 shows a typical DIS coil.

8.5 Coil on plug (COP) ignition

8.5.1 General description

Coil on plug (COP), or direct ignition, is the natural follow-on from distributorless ignition. This system utilizes an inductive coil for each cylinder. These coils are mounted directly on the spark plugs. Figure 8.25 shows a cross-section of the direct ignition coil. The use of an individual coil for each plug ensures that the

Figure 8.24 DIS coil and plug leads

Key fact

The use of an individual coil for each plug ensures that the rise time for the low inductance primary winding is very fast.

rise time for the low inductance primary winding is very fast. This ensures that a very high voltage, high energy spark is produced. This voltage, which can be in excess of 40 kV, provides efficient initiation of the combustion process under cold starting conditions and with weak mixtures. Some direct ignition systems use capacitor discharge ignition.

In order to switch the ignition coils, igniter units are used. These can control up to three coils and are simply the power stages of the control unit but in a

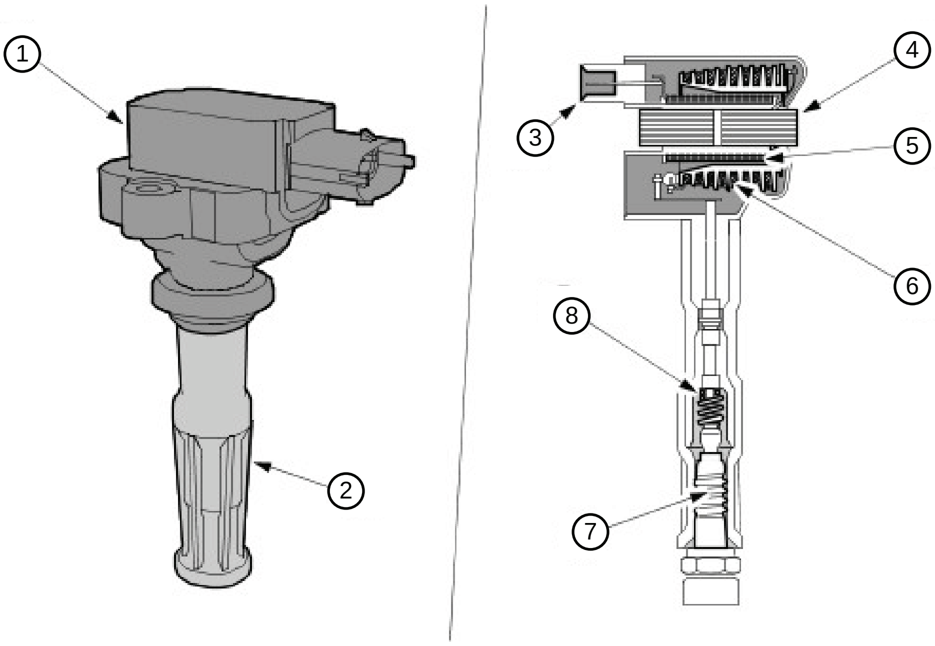


Figure 8.25 Direct ignition coil features: 1-Direct ignition coil, 2-Spark plug connector, 3-Low voltage connection, outer, 4-Laminated iron core, 5-Primary winding, 6-Secondary winding, 7-Spark plug, 8-High voltage connection, inner, via spring contact (Source: Ford Motor Company)

Figure 8.26 Six direct ignition coils in position

separate container. This allows less interference to be caused in the main ECU due to heavy current switching and shorter runs of wires carrying higher currents.

8.5.2 Control of ignition

Ignition timing and dwell are controlled in a manner similar to the previously described programmed system. The one important addition to this on some systems is a camshaft sensor to provide information as to which cylinder is on the compression stroke. A system that does not require a sensor to determine which cylinder is on compression (engine position is known from a crank sensor) determines the information by initially firing all of the coils. The voltage across the plugs allows measurement of the current for each spark and will indicate which cylinder is on its combustion stroke. This works because a burning mixture has a lower resistance. The cylinder with the highest current at this point will be the cylinder on the combustion stroke.

A further feature of some systems is the case when the engine is cranked over for an excessive time, making flooding likely. The plugs are all fired with multisparks for a period of time after the ignition is left in the on position for 5 seconds. This will burn away any excess fuel.

During difficult starting conditions, multisparking is also used by some systems during 70° of crank rotation before TDC. This assists with starting and then, once the engine is running, the timing will return to its normal calculated position.

8.6 Spark plugs

8.6.1 Functional requirements

The simple requirement of a spark plug is that it must allow a spark to form within the combustion chamber, to initiate burning. In order to do this the plug has to withstand a number of severe conditions. Consider, as an example, a

Figure 8.27 Spark plugs

four-cylinder four-stroke engine with a compression ratio of 9 : 1, running at speeds up to 5000 rpm. The following conditions are typical. At this speed the four-stroke cycle will repeat every 24 ms.

- End of induction stroke –0.9 bar at 65 °C.
- Ignition firing point –9 bar at 350 °C.
- Highest value during power stroke –45 bar at 3000 °C.
- Power stroke completed –4 bar at 1100 °C.

Besides the above conditions, the spark plug must withstand severe vibration and a harsh chemical environment. Finally, but perhaps most important, the insulation properties must withstand voltage pressures up to 40 kV.

8.6.2 Construction

Figure 8.28 shows a standard and a resistor spark plug. The centre electrode is connected to the top terminal by a stud. The electrode is constructed of a nickel-based alloy. Silver and platinum are also used for some applications. If a copper core is used in the electrode this improves the thermal conduction properties.

The insulating material is ceramic-based and of a very high grade. Aluminium oxide, Al2O3 (95% pure), is a popular choice, it is bonded into the metal parts and glazed on the outside surface. The properties of this material, which make it most suitable, are as follows:

- Young’s modulus: 340 kN/mm^2.
- Coefficient of thermal expansion: 7.8 10 K1.
- Thermal conductivity: 15–5 W/m K (Range 200–900 °C).
- Electrical resistance: 1013 /m.

The above list is intended as a guide only, as actual values can vary widely with slight manufacturing changes. The electrically conductive glass seal between the electrode and terminal stud is also used as a resistor. This resistor has two

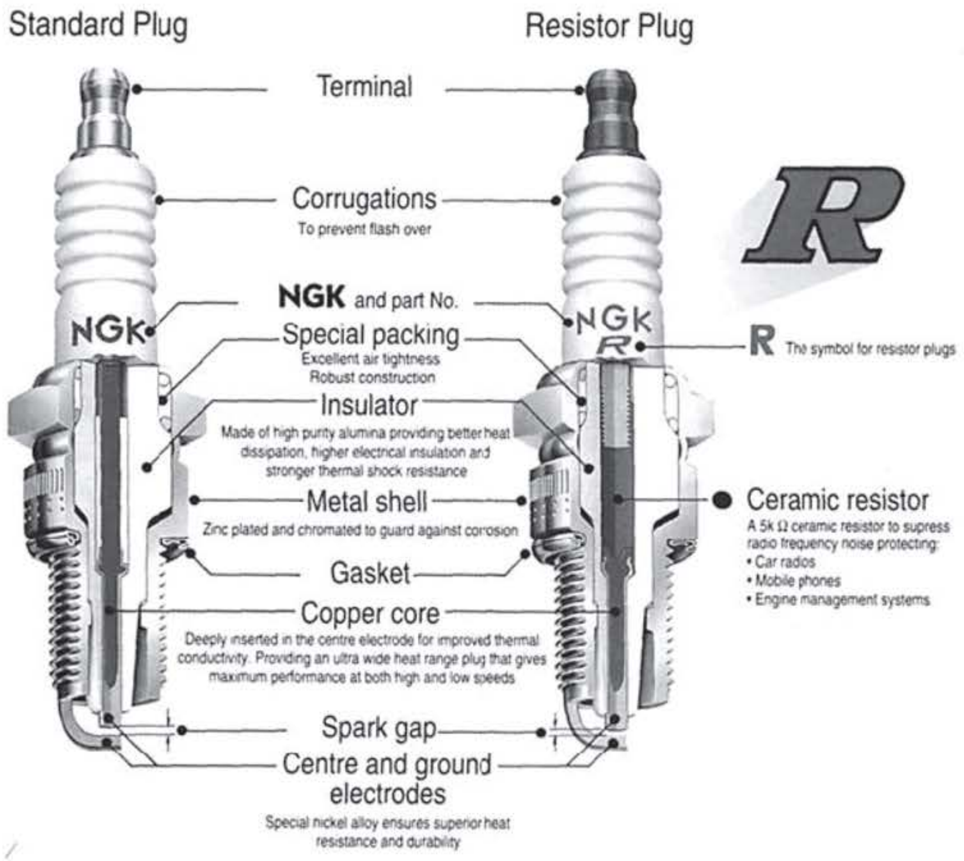


Figure 8.28 Spark-plug construction (Source: NGK)

functions. First, to prevent burn-off of the centre electrode, and secondly to reduce radio interference. In both cases the desired effect is achieved because the resistor damps the current at the instant of ignition.

Flash-over, or tracking down the outside of the plug insulation, is prevented by ribs that effectively increase the surface distance from the terminal to the metal fixing bolt, which is of course earthed to the engine.

8.6.3 Heat range

Due to the many and varied constructional features involved in the design of an engine, the range of temperatures in which a spark plug is exposed to, can vary significantly. The operating temperature of the centre electrode of a spark plug is critical. If the temperature becomes too high then pre-ignition may occur as the fuel–air mixture may become ignited due to the incandescence of the plug electrode. On the other hand, if the electrode temperature is too low then carbon and oil fouling can occur as deposits are not burnt off. Fouling of the plug nose can cause shunts (a circuit in parallel with the spark gap). It has been shown through experimentation and experience that the ideal operating temperature of the plug electrode is between 400 and 900 °C. Figure 8.29 shows how the temperature of the electrode changes with engine power output.

The heat range of a spark plug then is a measure of its ability to transfer heat away from the centre electrode. A hot running engine will require plugs with a higher thermal loading ability than a colder running engine. Note that hot and cold running of an engine in this sense refers to the combustion temperature and not to the efficiency of the cooling system.

Key fact

The operating temperature of the centre electrode of a spark plug is critical.

Key fact

The heat range of a spark plug then is a measure of its ability to transfer heat away from the centre electrode.

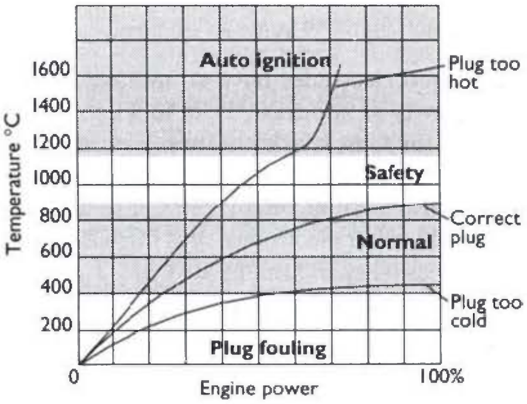


Figure 8.29 Temperature of a spark plug electrode changes with engine power output

The following factors determine the thermal capacity of a spark plug.

- Insulator nose length.
- Electrode material.
- Thread contact length.
- Projection of the electrode.

All these factors are dependent on each other and the position of the plug in the engine also has a particular effect.

It has been found that a longer projection of the electrode helps to reduce fouling problems due to low power operation, stop–go driving and high altitude conditions. In order to use greater projection of the electrode, better quality thermal conduction is required to allow suitable heat transfer at higher power outputs. Figure 8.30 shows the heat conducting paths of a spark plug together with changes in design for heat ranges. Also shown is the range of part numbers for NGK plugs.

8.6.4 Electrode materials

The material chosen for the spark plug electrode must exhibit the following properties:

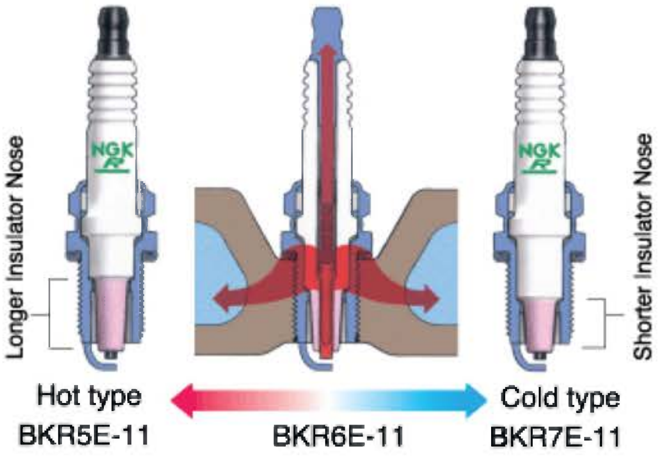


Figure 8.30 Heat conducting paths of a spark plug (Source: NGK)

Key fact

For normal applications, alloys of nickel are used for the electrode material.

- High thermal conductivity.
- High corrosion resistance.
- High resistance to burn-off.

For normal applications, alloys of nickel are used for the electrode material. Chromium, manganese, silicon and magnesium are examples of the alloying constituents. These alloys exhibit excellent properties with respect to corrosion and burn-off resistance. To improve on the thermal conductivity, compound electrodes are used. These allow a greater nose projection for the same temperature range, as discussed in the last section. A common example of this type of plug is the copper-core spark plug.

Silver electrodes are used for specialist applications as silver has very good thermal and electrical properties. Again, with these plugs nose length can be increased within the same temperature range. The thermal conductivity of some electrode materials is listed for comparison.

- Silver 407 W/m K.
- Copper 384 W/m K.
- Platinum 70 W/m K.
- Nickel 59 W/m K.

Compound electrodes have an average thermal conductivity of about 200 W/mΩ K. Platinum tips are used for some spark plug applications due to the very high burn-off resistance of this material. It is also possible because of this to use much smaller diameter electrodes, thus increasing mixture accessibility. Platinum also has a catalytic effect, further accelerating the combustion process (Figure 8.31).

Figure 8.31 The plug has a good anti-fouling characteristics

8.6.5 Electrode gap

Spark plug electrode gaps have, in general, increased as the power of the ignition systems driving the spark has increased. The simple relationship between plug gap and voltage required is that, as the gap increases so must the voltage (leaving aside engine operating conditions). Furthermore, the energy available to form a spark at a fixed engine speed is constant, which means that a larger gap using higher voltage will result in a shorter duration spark. A smaller gap will allow a longer duration spark. For cold starting an engine and for igniting weak mixtures, the duration of the spark is critical. Likewise the plug gap must be as large as possible to allow easy access for the mixture in order to prevent quenching of the flame.

The final choice is therefore a compromise reached through testing and development of a particular application. Plug gaps in the region of 0.6–1.2 mm seem to be the norm at present.

8.6.6 V-grooved spark plug

The V-grooved plug is a development by NGK designed to reduce electrode quenching and allow the flame front to progress more easily from the spark. This is achieved by forming the electrode end into a ‘V’ shape, as shown in Figure 8.32.

This allows the spark to be formed at the side of the electrode, giving better propagation of the flame front and less quenching due to contact with the earth and centre electrodes. Figure 8.33 shows a V-grooved plug firing together with a graphical indication of the potential improvements when compared with the conventional plug.

Figure 8.32 NGK V-power plug (Source: NGK)

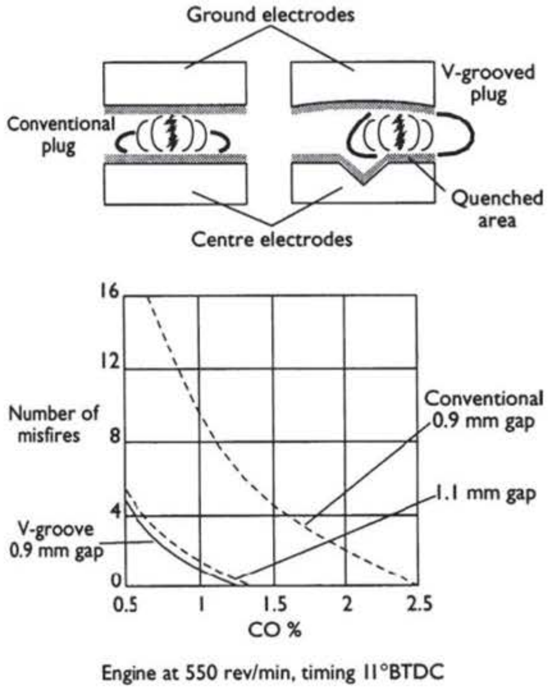


Figure 8.33 V-grooved spark plug firing, together with a graph indicating potential improvements when compared with the conventional plug

8.6.7 Choosing the correct plug

Two methods are often used to determine the best spark plug for a given application. In the main it is the temperature range that is of prime importance. The first method of assessing plug temperature is the thermocouple spark plug, as shown in Figure 8.34. This allows quite accurate measurement of the temperature but does not allow the test to be carried out for all types of plug.

A second method is the technique of ionic current measurement. When combustion has been initiated, the conductivity and pattern of current flow across the plug gap is a very good indication of the thermal load on the plug. This process allows accurate matching of the spark plug heat range to every engine, as well as providing data on the combustion temperature of a test engine. This technique is starting to be used as feedback to engine management systems to assist with accurate control.

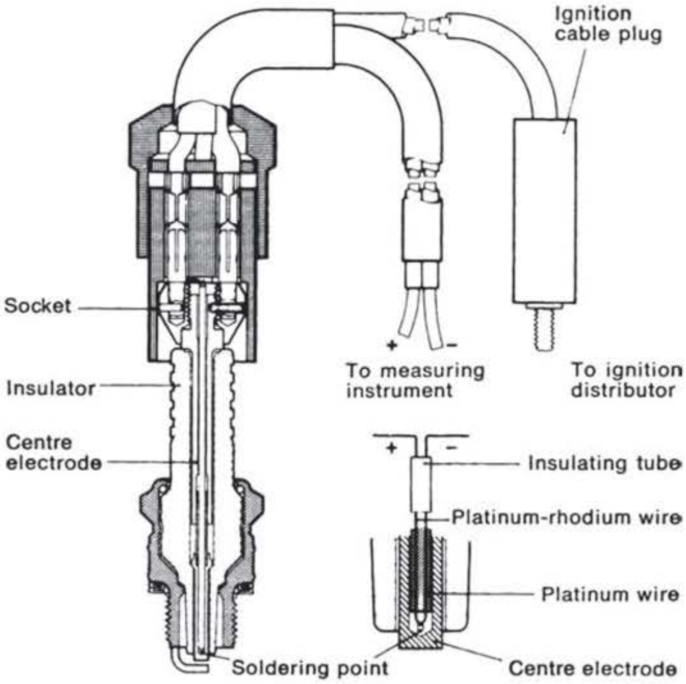


Figure 8.34 Thermocouple spark plug

In the after-market, choosing the correct plug is a matter of using manufacturers' parts catalogues.

8.6.8 Spark plugs development

Most developments in spark plug technology are incremental. Recent trends have been towards the use of platinum plugs and the development of a plug that will stay within acceptable parameters for long periods (i.e. in excess of 50 000 miles/80 000 km).

Multiple electrode plugs are a contribution to long life and reliability. Do note though that these plugs only produce one spark at one of the electrodes each time they fire. The spark will jump across the path of least resistance and this will normally be the path that will produce the best ignition or start to combustion. Equally, the wear rate is spread over two or more electrodes. Figure 8.35 shows a number of platinum spark plugs.

8.7 Summary

8.7.1 Overview

Modern ignition systems are now part of the engine management, which controls fuel delivery, ignition and other vehicle functions. These systems are under continuous development and reference to the manufacturer's workshop manual is essential when working on any vehicle. The main ignition components are the engine speed and load sensors, knock sensor, temperature sensor and the ignition coil. The ECU reads from the sensors, interprets and compares the data, and sends output signals to the actuators. The output component for ignition is the coil. Some form of electronic ignition is now fitted to all spark ignition vehicles.

Key fact

The main ignition components are the engine speed and load sensors, knock sensor, temperature sensor and the ignition coil.

Figure 8.35 Platinum spark plugs (Source: Bosch Media)

In order for a constant energy electronic ignition system to operate, the dwell must increase with engine speed. This will only be of benefit, however, if the ignition coil can be charged up to its full capacity in a very short time. Constant energy means that, within limits, the energy available to make the spark at the plug remains constant under all operating conditions. An energy value of about 0.3 mJ is all that is required to ignite a static stoichiometric (ideal proportion) mixture. However, with lean or rich mixtures, together with high turbulence, energy values in the region of 3 to 4 mJ are necessary. This has made constant energy ignition essential on all of today's vehicles so they can meet emission and performance requirements.

Programmed ignition is the term used by some manufacturers for digitally controlled ignition; others call it electronic spark advance (ESA). Constant energy electronic ignition was a major step forwards and is still used on many vehicles, together with a standard distributor. However, its limitations lie in still having to rely upon mechanical components for speed and load advance characteristics. In many cases these did not match ideally the requirements of the engine. With a digital system, information about the operating requirements of a particular engine is programmed in to memory inside the electronic control unit. This data, stored in read only memory (ROM), is obtained from testing on an engine dynamometer and then under various operating conditions.

Distributorless ignition has all the features of programmed ignition systems but, by using a special type of ignition coil, operates the spark plugs without the need for a distributor. The basic principle is that of the 'lost spark'. On a four-cylinder engine, the distribution of the spark is achieved by using two double-ended coils, which are fired alternately by the ECU. The timing is determined from a crankshaft speed and position sensor as well as load and other corrections. When one of the coils is fired a spark is delivered to two engine cylinders, either 1 and 4, or 2 and 3. The spark delivered to the cylinder on the compression stroke will ignite the mixture as normal. The spark produced in the other cylinder will have no effect, as this cylinder will be just completing its exhaust stroke.

Coil on plug (COP) or direct ignition is similar, but has one ignition coil for each cylinder, which is mounted directly on the spark plug. The use of an individual coil for each plug ensures that the charge time for the low inductance primary winding is very fast. This ensures that a very high voltage, high-energy spark is produced.

Ignition timing and dwell are controlled digitally. On some systems a camshaft sensor is used to provide information about which cylinder is on the compression stroke. An interesting method, which does not require a sensor to determine which cylinder is on compression (engine position is known from a crank sensor), determines the information by initially firing all of the coils. The voltage across the plugs allows measurement of the current for each spark and will indicate which cylinder is on its combustion stroke. This works because a burning mixture has a lower resistance. The cylinder with the highest current at this point will be the cylinder on the combustion stroke.

Modern ignition systems, that are part of an engine management system, usually have a limp-home facility that allows the engine to continue to operate when defects are detected by the ECU. Basic settings are substituted and a warning light is illuminated to alert the driver. Self-test and onboard diagnostic (OBD) links are provided for diagnostic tests to be carried out.

Ignition systems continue to develop and will continue to improve. However, keep in mind that the simple purpose of an ignition system is to ignite the fuel–air mixture every time at the right time. And, no matter how complex the electronics may seem, the high voltage is produced by switching a coil on and off.

Key fact

High voltage sparks are produced by switching a coil on and off.

Safety first

Caution/Achtung/Attention - High voltages can seriously damage your health!

8.7.2 Testing procedure

Table 8.3 lists some common symptoms of an ignition system malfunction together with suggestions for the possible fault.

Figure 8.36 Combustion taking place (Source: Ford Media)

Table 8.3 Ignition diagnostics

Symptom	Possible fault
Engine rotates but does not start	Damp ignition components Spark plugs worn to excess Ignition system open circuit
Difficult to start when cold	Spark plugs worn to excess High resistance in ignition circuit
Engine starts but then stops immediately	Ignition wiring connection intermittent Ballast resistor open circuit (older cars)
Erratic idle	Incorrect plug gaps Incorrect ignition timing
Misfire at idle speed	Ignition coil or distributor cap tracking Spark plugs worn to excess
Misfire through all speeds	Incorrect plugs or plug gaps HT leads breaking down
Lack of power	Ignition timing incorrect HT components tracking
Backfires	Incorrect ignition timing Tracking
Runs on when switched off	Ignition timing incorrect Carbon build up in engine
Pinking or knocking under load	Ignition timing incorrect Ignition system electronic fault Knock sensor not working

The following procedure is generic but with a little adaptation can be applied to any ignition system. Refer to manufacturer's recommendations if in any doubt.

1. Check battery state of charge (at least 70%).
2. Hand and eye checks (all connections secure and clean).
3. Check supply to ignition coil (within 0.5 V of battery).
4. Spark from coil via known good HT lead (jumps about 10 mm, but do not try more).
5. If good spark then check HT system for tracking and open circuits. Check plug condition (leads should be a maximum resistance of about 30 kΩ/m and per lead) – Stop here in this procedure.
6. If no spark or it will only jump a short distance continue with this procedure (colour of spark is not relevant).
7. Check continuity of coil windings (primary 0.5 to 3 Ω, secondary *several* kΩ.
8. Supply and earth to 'module' (12 V minimum supply, earth drop 0.5 V maximum).
9. Supply to pulse generator if appropriate (10 to 12 V).
10. Output of pulse generator (inductive about 1 V AC when cranking, Hall type switches 0 V to 8 V DC).
11. Continuity of LT wires (0 to 0.1 Ω).
12. Replace 'module' or ECU but only if appropriate tests above are satisfactory.

8.8 Advanced ignition technology

8.8.1 Ignition coil performance

The instantaneous value of primary current in the inductive circuit of the ignition coil is determined by a number of factors. The HT produced is mainly dependent on this value of primary current. The rate of increase of primary current is vital because this determines the value of current when the circuit is 'broken' to produce the collapse of the magnetic field.

If the electrical constants of the primary ignition system are known it is possible to calculate the instantaneous primary current. This requires the exponential equation:

$$i = \frac{V}{R}\left(1 - e^{-Rt/L}\right)$$

where: i = instantaneous primary current, R = total primary resistance, L = inductance of primary winding, t = time current has been flowing, e = base of natural logs.

Some typical values for comparison are:

Contact breaker ignition	Electronic ignition
$R = 3$ to $4\ \Omega$	$R = 1\ \Omega$
$V = 14$ V	$V = 14$ V
$L = 10$ mH	$L = 4$ mH

Using as an example a four cylinder engine running at 3000 rpm, 6000 sparks per minute are required (four sparks during the two revolutions to complete the four stroke cycle). This equates to 6000/60 or 100 sparks per second. At this rate each spark must be produced and used in 10 ms.

Taking a typical dwell period of say 60% the time *t*, at 3000 rpm on a four cylinder engine is 6 ms. At 6000 rpm, *t* will be 3 ms. Employing the exponential equation above, the instantaneous current for each system is:

	3000 rpm	6000 rpm
Conventional system =	3.2 A	2.4 A
Electronic system =	10.9 A	7.3 A

This gives a clear indication of how the energy stored in the coil is much increased by the use of low resistance and low inductance ignition coils. It is important to note that the higher current flowing in the electronic system would have been too much for the conventional contact breakers.

The energy stored in the magnetic field of the ignition coil is calculated as shown:

$$E = \tfrac{1}{2}\left(L \times i^2\right)$$

where: E = energy, L = inductance of primary winding, i = instantaneous primary current.

The stored energy of the electronic system at 6000 rpm is 110 mJ; the energy in the conventional system is 30 mJ. This clearly shows the advantage of electronic ignition as the spark energy is directly related to energy stored in the coil.

This page intentionally left blank

CHAPTER 9

Fuel control

9.1 Combustion

9.1.1 Introduction

The process of combustion in spark and compression ignition engines is best considered for petrol and diesel engines in turn. The knowledge of the more practical aspects of combustion has been gained after years of research and is by no means complete even now. For a complete picture of the factors involved, further reference should be made to appropriate sources. However, the combustion section here will give enough details to allow considered opinion about the design and operation of electronic fuel control systems.

9.1.2 Spark ignition engine combustion process

A simplified description of the combustion process within the cylinder of a spark ignition engine is as follows. A single high intensity spark of high temperature passes between the electrodes of the spark plug leaving behind it a thin thread of flame. From this thin thread combustion spreads to the envelope of mixture immediately surrounding it at a rate that depends mainly on the flame front temperature, but also, to a lesser degree, on the temperature and density of the surrounding envelope.

In this way, a bubble of flame is built up that spreads radially outwards until the whole mass of mixture is burning. The bubble contains the highly heated

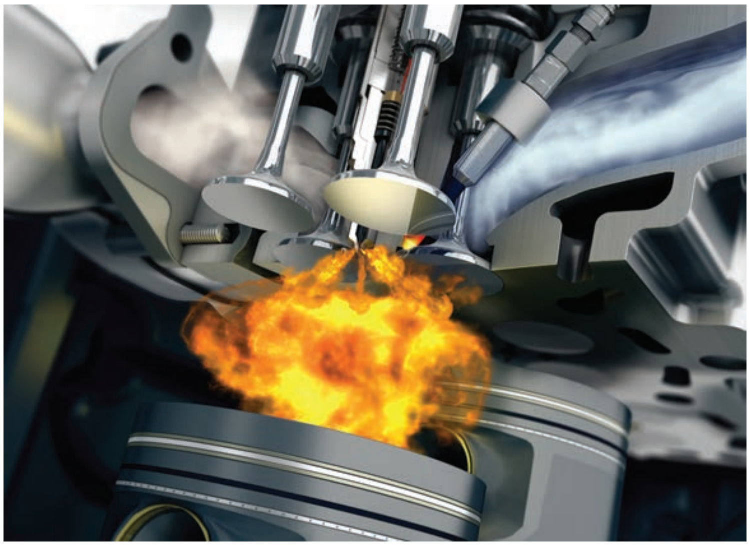


Figure 9.1 Combustion taking place

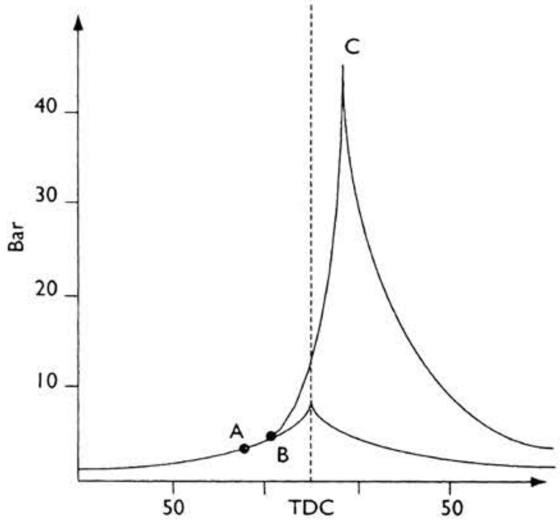


Figure 9.2 The speed at which fuel will oxidize or burn

products of combustion, while ahead of it, and being compressed by it, is the still unburnt mixture.

If the cylinder contents were at rest this bubble would be unbroken, but with the air turbulence normally present within the cylinder, the filament of flame is broken up into a ragged front, which increases its area and greatly increases the speed of advance. While the rate of advance depends on the degree of turbulence, the direction is little affected, unless some definite swirl is imposed on the system. The combustion can be considered in two stages.

1. Growth of a self-propagating flame.
2. Spread through the combustion chamber.

The first process is chemical and depends on the nature of the fuel, the temperature and pressure at the time and the speed at which the fuel will oxidize or burn. Shown in Figure 9.1, it appears as the interval from the spark (A) to the time when an increase in pressure due to combustion can first be detected (B).

Key fact

The time interval from ignition to burning occurs with all fuels but may be reduced with an increase of compression temperature.

This ignition delay period can be clearly demonstrated. If fuel is burned at constant volume, having been compressed to a self-ignition temperature, the pressure–time relationship is as shown in Figure 9.3.

The time interval occurs with all fuels but may be reduced with an increase of compression temperature. A similar result can be demonstrated, enabling the effect of mixture strength on ignition delay to be investigated.

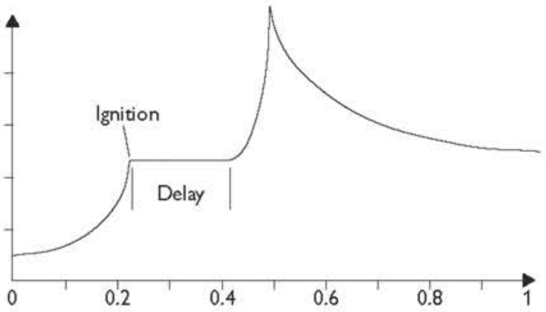


Figure 9.3 Fuel is burned at a constant volume having been compressed to a self-ignition temperature. The pressure–time relationship is shown

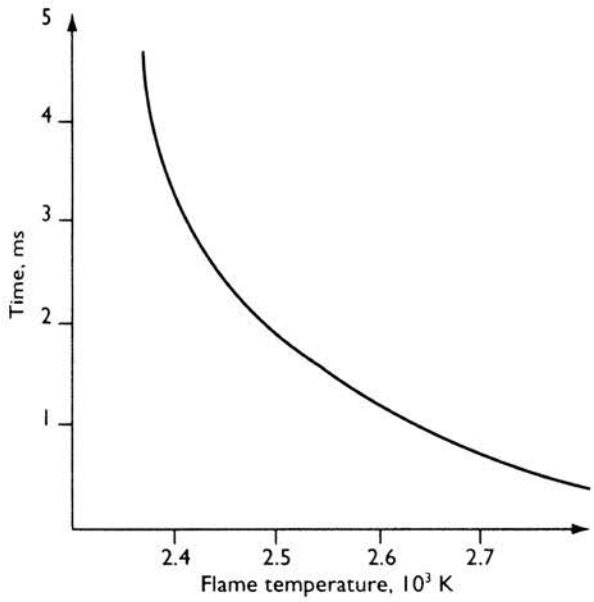


Figure 9.4 Approximated relationship between flame temperature and the time from spark to propagation of flame for a hydrocarbon fuel

Referring to Figure 9.2, with the combustion under way, the pressure rises within the engine cylinder from (B) to (C), very rapidly approaching the 'constant volume' process of the four-stroke cycle. While (C) represents the peak cylinder pressure and the completion of flame travel, all available heat has not been liberated due to re-association, and what can be referred to as after-burning continues throughout the expansion stroke.

9.1.3 Range and rate of burning

The range and rate of burning issue can be summarised by reference to the following graphs:

Figure 9.4 shows the approximate relation between flame temperature and the time from spark to propagation of flame for a hydrocarbon fuel.

Figure 9.5 shows the relation between the flame temperature and the mixture strength.

Figure 9.6 shows the relationship between mixture strength and rate of burning.

These graphs show that the minimum delay time (A to B) is about 0.2 ms seconds with the mixture slightly rich. While the second stage (B to C) is roughly dependent upon the degree of the turbulence (and on the engine speed) the initial delay necessitates ignition advance as the engine speed increases.

Figure 9.7 shows the effects of incorrect ignition timing. As the ignition is advanced there is an increase in firing pressure or maximum cylinder pressure, generally accompanied by a reduction in exhaust temperature. The effect of increasing the range of the mixture strength speeds the whole process up and thus increases the tendency to detonate.

9.1.4 Detonation

The detonation phenomenon is the limiting factor on the output and efficiency of the spark ignition engine. The mechanism of detonation is the setting up within

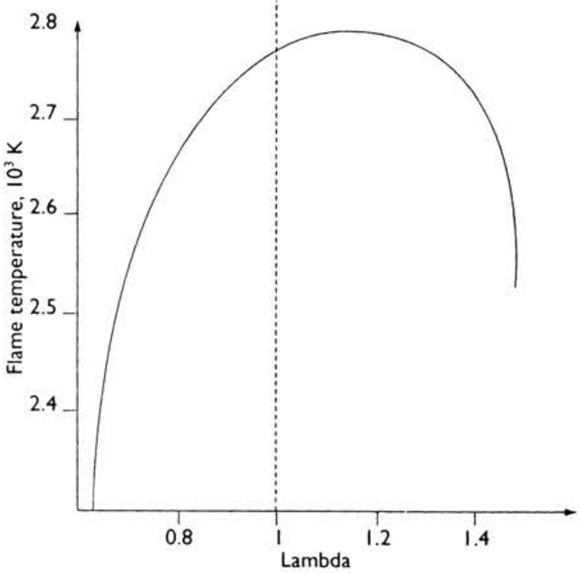


Figure 9.5 Relationship between flame temperature and mixture strength

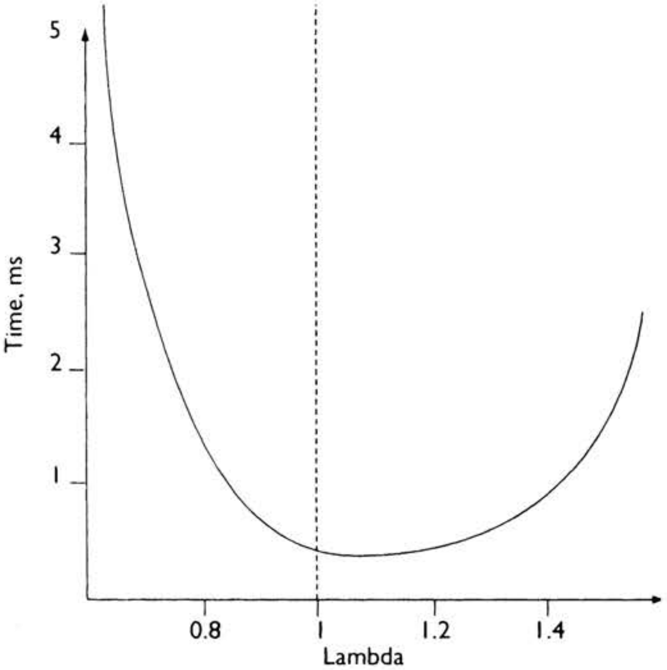


Figure 9.6 Relationship between mixture strength and rate of burning

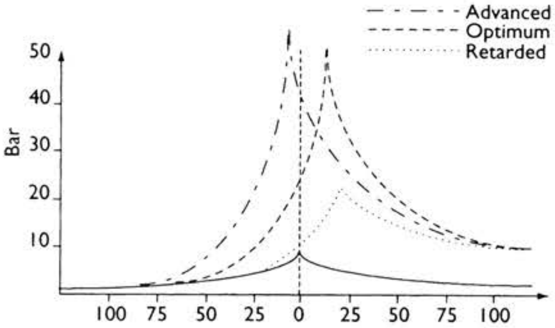


Figure 9.7 Effects of incorrect ignition timing on fuel burn

the engine cylinder of a pressure wave travelling at such velocity as, by its impact against the cylinder walls, to set them in vibration, and thus produce a high pitched ‘ping’. When the spark ignites a combustible mixture of the fuel and air, a small nucleus of flame builds up, slowly at first but accelerating rapidly. As the flame front advances it compresses the remaining unburned mixture ahead of it. The temperature of the unburned mixture is raised by compression and radiation from the advancing flame until the remaining charge ignites spontaneously. The detonation pressure wave passes through the burning mixture at a very high velocity and the cylinder walls emit the ringing knock.

Detonation is not too dangerous in small engines because it is usually avoided at the first warning by easing the load, but at higher speeds, where the noise level is high, the characteristic noise can and often does go undetected. It can be extremely dangerous, prompting pre-ignition and possibly the complete destruction of the engine.

High compression temperature and pressure tend to promote detonation. In addition, the ability of the unburnt mixture to absorb or get rid of the heat radiated to it by the advancing flame front is also important. The latent enthalpy of the mixture and the design of the combustion chamber affect this ability. The latter must be arranged for adequate cooling of the unburnt mixture by placing it near a well-cooled feature such as an inlet valve.

Key fact

High compression temperature and pressure tend to promote detonation.

The length of flame travel should be kept as short as possible by careful positioning of the point of ignition. Other factors include the time (hence the ignition timing), since the reaction in the unburnt mixture must take some time to develop; the degree of turbulence (in general, higher turbulence tends to reduce detonation effects); and, most importantly, the tendency of the fuel itself to detonate.

Some fuels behave better in this respect. Fuel can be treated by additives (e.g. tetra-ethyl lead) to improve performance. However, this aggravates an already difficult pollution problem. A fuel with good anti-knock properties is iso-octane, and a fuel that is susceptible to detonation is normal heptane.

To obtain the octane number or the anti-knock ratings of a particular blend of fuel, a test is carried out on an engine run under carefully monitored conditions, and the onset of detonation is compared with those values obtained from various mixtures of iso-octane and normal heptane. If the performance of the fuel is identical to, for example, a mixture of 90% iso-octane and 10% heptane, then the fuel is said to have an octane rating of 90.

Mixing water, or methanol and water, with the fuel can reduce detonation. A mainly alcohol-based fuel, which enables the water to be held in solution, is also helpful so that better use can be made of the latent enthalpy of the water.

9.1.5 Pre-ignition

Evidence of the presence of pre-ignition is not so apparent at the onset as detonation, but the results are far more serious. There is no characteristic ‘ping’. In fact, if audible at all, it appears as a dull thud. Since it is not immediately noticeable, its effects are often allowed to take a serious toll on the engine. The process of combustion is not affected to any extent, but a serious factor is that control of ignition timing can be lost.

Pre-ignition can occur at the time of the spark with no visible effect. More seriously, the ‘auto-ignition’ may creep earlier in the cycle. The danger of pre-ignition lies not so much in development of high pressures but in the very great increase in heat flow to the piston and cylinder walls. The maximum pressure does not, in fact, increase appreciably although it may occur a little early.

In a single-cylinder engine, the process is not dangerous since the reduction usually causes the engine to stall. In a multiple-cylinder engine the remaining cylinders (if only one is initially affected), will carry on at full power and speed, dragging the pre-igniting cylinder after them. The intense heat flow in the affected cylinder can result in piston seizure followed by the breaking up of the piston with catastrophic results to the whole engine.

Key fact

Pre-ignition is often initiated by some form of hot spot, perhaps red-hot carbon or some poorly cooled feature of combustion space.

Pre-ignition is often initiated by some form of hot spot, perhaps red-hot carbon or some poorly cooled feature of combustion space. In some cases, if the incorrect spark plug is used, over-heated electrodes are responsible, but often detonation is the prime cause. The detonation wave scours the cylinder walls of residual gases present in a film on the surface with the result that the prime source of resistance to heat flow is removed and a great release of heat occurs. Any weaknesses in the cooling system are tested and any hot spots formed quickly give rise to pre-ignition.

9.1.6 Combustion chamber

To avoid the onset of detonation and pre-ignition, a careful layout of the valves and spark plugs is essential. Smaller engines, for automotive use, are firmly tied to the poppet valve. This, together with the restriction of space involved with high compression ratios, presents the designer with interesting problems.

The combustion chamber should be designed bearing in mind the following factors:

- The compression ratio should be 9 : 1 for normal use, 11 or 12 : 1 for higher performance.
- The plug or plugs should be placed to minimize the length of flame travel. They should not be in pockets or otherwise shrouded since this reduces effective cooling and also increases the tendency toward cyclical variations.

Experimental evidence shows a considerable variation in pressure during successive expansion stokes. This variation increases, as the mixture becomes too weak or too rich. Lighter loads and lower compression ratios also aggravate the process. While the size and position of the point of maximum pressure changes, the mean effective pressure and engine output is largely unaffected.

9.1.7 Stratification of cylinder charge

Key fact

One technique to get around the problem of igniting weak mixtures is stratification.

A very weak mixture is difficult to ignite but has great potential for reducing emissions and improving economy. One technique to get around the problem of igniting weak mixtures is stratification.

It is found that if the mixture strength is increased near the plug and weakened in the main combustion chamber an overall reduction in mixture strength results, but with a corresponding increase in thermal efficiency. To achieve this, petrol injection is used – stratification being very difficult with a conventional carburation system. A system that uses this technique is gasoline direct injection, which can allow a petrol engine to run with much weaker air-to-fuel ratios. This is discussed in a later section.

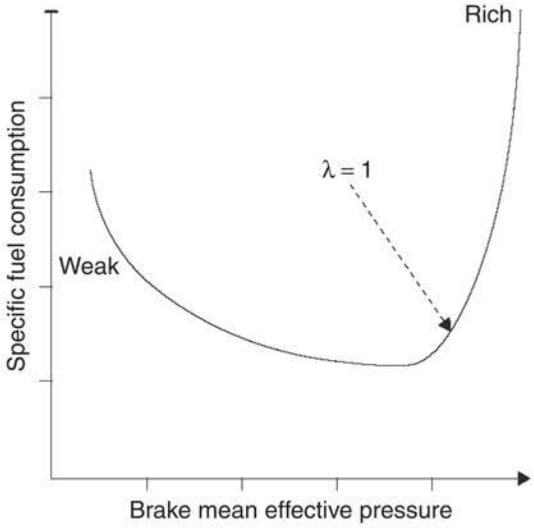


Figure 9.8 Effect of varying mixture strength while keeping throttle position, engine speed and ignition timing constant

9.1.8 Mixture strength and performance

The effect of varying the mixture strength while maintaining the throttle position, engine speed and ignition timing constant is shown in Figure 9.8.

Figure 9.9 shows the effect of operating at part throttle with varying mixture strength. The chemically correct mixture of approximately 14.7 : 1 lies between the ratio that provides maximum power (12 : 1), and minimum consumption (16 : 1). The stoichiometric ratio of 14.7 : 1 is known as a lambda value of one. Figure 9.10 compares engine power output and fuel consumption with changes in air–fuel ratio.

9.1.9 Compression ignition (CI) engines

The process of combustion in the compression ignition engine differs from that in a spark ignition engine. In this case the fuel is injected in a liquid state, into a highly compressed, high-temperature air supply in the engine cylinder. Each

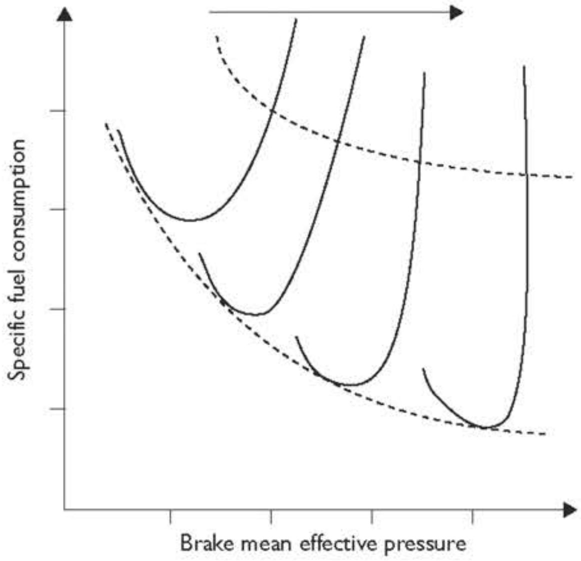


Figure 9.9 Effect operating at part throttle with varying mixture strength

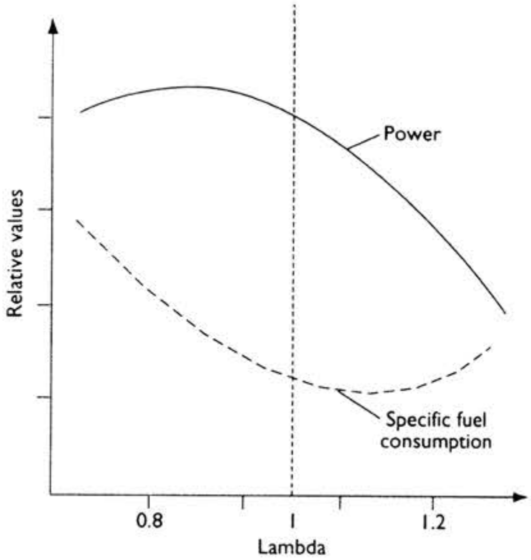


Figure 9.10 Comparison of engine power output and fuel consumption with changes in air-fuel ratio

Key fact

The process oxidization of hydrocarbon fuel is a lengthy process, but one that may be accelerated artificially by providing the most suitable conditions.

minute droplet is quickly surrounded by an envelope of its own vapour as it enters the highly heated air. This vapour, after a certain time, becomes inflamed on the surface. A cross-section of any one droplet would reveal a central core of liquid, a thin surrounding film of vapour, with an outer layer of flame. This sequence of vaporization and burning persists as long as combustion continues.

The process of combustion (oxidization of the hydrocarbon fuel), is in itself a lengthy process, but one that may be accelerated artificially by providing the most suitable conditions. The oxidization of the fuel will proceed in air at normal atmospheric temperatures, but it will be greatly accelerated if the temperature is raised. It will take years at 20 °C, a few days at 200 °C and just a few minutes at 250 °C. In these cases, the rate of temperature rise due to oxidization is less than the rate at which the heat is being lost due to convection and radiation. Ultimately, as the temperature is raised, a critical stage is reached where heat is being generated by oxidization at a greater rate than it is being dissipated.

The temperature then proceeds to rise automatically. This, in turn, speeds up the oxidization process and with it the release of heat. Events now take place very rapidly, a flame is established and ignition takes place. The temperature at which this critical change takes place is usually termed the self-ignition temperature of the fuel. This, however, depends on many factors such as pressure, time and the ability to transmit heat from the initial oxidization.

We will now look at the injection of the fuel as a droplet into the heated combustion chamber. At a temperature well above the ignition point, the extreme outer surface of the droplet immediately starts to evaporate, surrounding the core with a thin film of vapour. This involves a supply of heat from the air surrounding the droplet in order to supply the latent enthalpy of evaporation. This supply is maintained by continuing to draw on the main supply of heat from the mass of hot air.

Ignition can and will occur on the vapour envelope even with the core of the droplet still liquid and relatively cold. Once the flame is established, the combustion proceeds at a more rapid rate. This causes a delay period, after injection commences and before ignition takes place. The delay period therefore depends on:

- Excess of air temperature over and above the self-ignition temperature of the fuel.
- Air pressure, both from the point of view of the supply of oxygen and improved heat transfer between the hot air and cold fuel.

Once the delay period is over, the rate at which each flaming droplet can find fresh oxygen to replenish its consumption controls the rate of further burning. The relative velocity of the droplet to the surrounding air is thus of considerable importance. In the compression ignition engine, the fuel is injected over a period of perhaps 40–50° of crank angle. This means that the oxygen supply is absorbed by the fuel first injected, with a possible starvation of the last fuel injected.

This necessitates a degree of turbulence of the air so that the burnt gases are scavenged from the injector zone and fresh air is brought into contact with the fuel. It is clear that the turbulence should be orderly and not disorganized, as in a spark ignition engine, where it is only necessary in order to break up the flame front.

Key fact

In a compression ignition engine the combustion can be regarded as occurring in three distinct phases: ignition delay, rapid pressure rise (flame spread), afterburn (controlled combustion).

In a compression ignition engine the combustion can be regarded as occurring in three distinct phases as shown in Figure 9.11.

- Delay period.
- Rapid pressure rise.
- After-burning, i.e. the fuel is burning as it leaves the injector.

The longer the delay, the greater and more rapid the pressure rise since more fuel will be present in the cylinder before the rate of burning comes under direct control of the rate of injection. The aim should be to reduce the delay as much as possible, both for the sake of smooth running, the avoidance of knock and also to maintain control over the pressure change. There is, however, a lower limit to the delay since, without delay, all the droplets would burn as they leave the nozzle. This would make it almost impossible to provide enough combustion air within the concentrated spray and the delay period also has its use in providing time for the proper distribution of the fuel. The delay period therefore depends on:

- The pressure and temperature of the air.

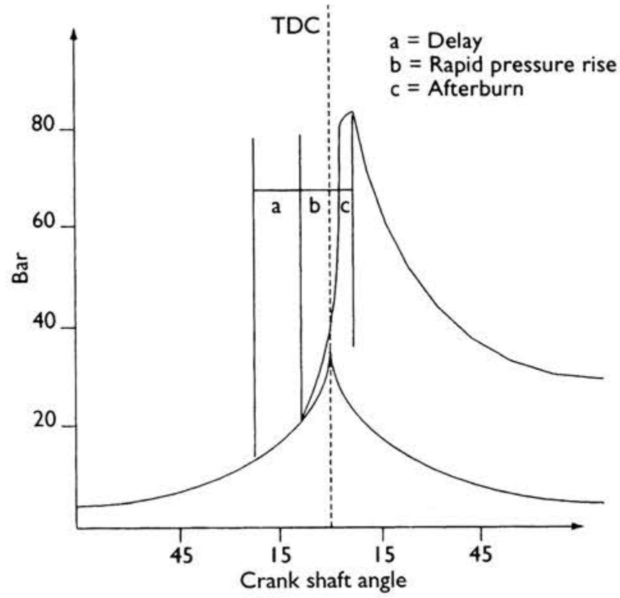


Figure 9.11 Phases of diesel combustion

- The cetane rating of the fuel.
- The volatility and latent enthalpy of the fuel.
- The droplet size.
- Controlled turbulence.

The effect of droplet size is important, as the rate of droplet burning depends primarily on the rate at which oxygen becomes available. It is, however, vital for the droplet to penetrate some distance from the nozzle around which burning will later become concentrated. To do this, the size of the droplets must be large enough to obtain sufficient momentum at injection. On the other hand, the smaller the droplet the greater the relative surface area exposed and the shorter the delay period. A compromise between these two effects is clearly necessary.

With high compression ratios (15 : 1 and above) the temperature and pressure are raised so that the delay is reduced, which is an advantage. However, high compression ratios are a disadvantage mechanically and also inhibit the design of the combustion chamber, particularly in small engines where the bumping clearance consumes a large proportion of the clearance volume.

9.1.10 Combustion chamber design – diesel engine

The combustion chamber must be designed to:

- Give the necessary compression ratio.
- Provide the necessary turbulence.
- Position for correct and optimum operation of the valves and injector.

These criteria have effects that are interrelated. Turbulence is normally obtained at the expense of volumetric efficiency. Masked inlet valves (which are mechanically undesirable) or ‘tangent’ directional ports restrict the air flow and therefore are restrictive to high-speed engines.

To assist in breathing, four or even six valves per cylinder can be used. This arrangement has the advantage of keeping the injector central, a desirable aim for direct injection engines. Large valves and their associated high lift, in addition to providing mechanical problems often require heavy piston recesses, which disturb squish and orderly movement of the air.

A hemispherical combustion chamber assists with the area available for valves, at the expense of using an offset injector. Pre-combustion chambers, whether of the air cell or ‘combustion swirl’ type have the general disadvantage of being prone to metallurgical failure or at least are under some stress since, as they are required to produce a ‘hot spot’ to assist combustion, the temperature stresses in this region are extremely high. There is no unique solution and the resulting combustion chamber is always a compromise.

9.1.11 Summary of combustion

This section has looked at some of the issues of combustion, and is intended to provide a background to some of the other sections in this book. The subject is very dynamic and improvements are constantly being made. Some of the key issues this chapter has raised so far include points such as the time to burn a fuel–air mixture, the effects of changes in mixture strength and ignition timing, the consequences of detonation and other design problems.

Accurate control of engine operating variables is one of the keys to controlling the combustion process.

9.2 Engine fuelling and exhaust emissions

9.2.1 Operating conditions

The ideal air-fuel ratio is about 14.7 : 1. This is the theoretical amount of air required to burn the fuel completely. It is given a 'lambda (λ)' value of 1.

λ = actual air quantity ÷ theoretical air quantity

Key fact

The ideal air-fuel ratio is about 14.7 : 1.

The air–fuel ratio is altered during the following operating conditions of an engine to improve its performance, drivability, consumption and emissions.

- Cold starting – a richer mixture is needed to compensate for fuel condensation and improves drivability.
- Load or acceleration – a richer mixture to improve performance.
- Cruise or light loads – a weaker mixture for economy.
- Overrun – very weak mixture (if any) to improve emissions and economy.

The more accurately the air-fuel ratio is controlled to cater for external conditions, then the better the overall operation of the engine.

9.2.2 Exhaust emissions

Figure 9.12 shows, first, the theoretical results of burning a hydrocarbon fuel and, second, the actual combustion results. The top part of the figure is ideal but the lower part is the realistic result under normal conditions. Note that this result is prior to any further treatment, for example by a catalytic converter.

Figure 9.13 shows the approximate percentages of the various exhaust gas emissions. The volume of pollutants is small but, because they are so poisonous, they are undesirable and strong legislation now exists to encourage their

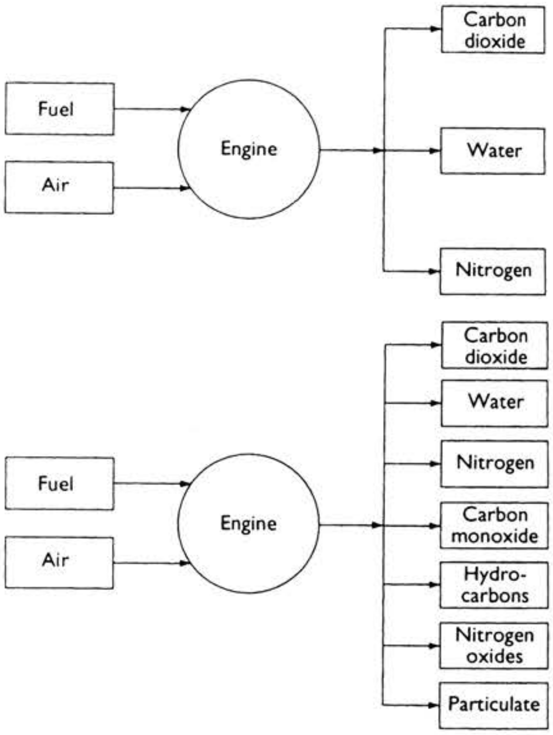


Figure 9.12 Theoretical results of burning a hydrocarbon fuel and actual combustion results

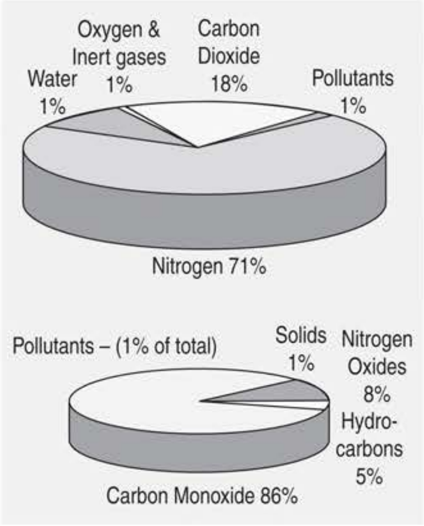


Figure 9.13 Composition of exhaust

Table 9.1 Emissions hazardous to health

Substance	Description
Carbon monoxide (CO)	This gas is very dangerous even in low concentrations. It has no smell or taste and is colourless. When inhaled it combines in the body with the red blood cells preventing them from carrying oxygen. If absorbed by the body it can be fatal in a very short time.
Nitrogen oxides (NOx)	Oxides of nitrogen are colourless and odourless when they leave the engine but as soon as they reach the atmosphere and mix with more oxygen, nitrogen oxides are formed. They are reddish brown and have an acrid and pungent smell. These gasses damage the body's respiratory system when inhaled. When combined with water vapour nitric acid can be formed which is very damaging to the windpipe and lungs. Nitrogen oxides are also a contributing factor to acid rain.
Hydrocarbons (HC)	A number of different hydrocarbons are emitted from an engine and are partly or unburnt fuel. When they mix with the atmosphere they can help to form smog. It is also believed that hydrocarbons may be carcinogenic.
Particulate matter (PM)	This heading in the main covers lead and carbon. Lead was traditionally added to petrol to slow its burning rate to reduce detonation. It is detrimental to health and is thought to cause brain damage especially in children. Lead will eventually be phased out as all new engines now run on unleaded fuel. Particles of soot or carbon are more of a problem on diesel-fuelled vehicles and these now have limits set by legislation.

reduction. The actual values of these emissions varies depending on engine design, operating conditions, temperature and smooth running, to name just a few variables.

Table 9.1 lists the four main emissions that are hazardous to health, together with a short description of each.

9.2.3 Other sources of emissions

The main source of vehicle emissions is the exhaust, but other areas of the vehicle must also come under scrutiny.

As well as sulphur in fuel, another area of contention between car manufacturers and oil companies is the question of who should bear the cost of collecting fuel vapour at filling stations. The issue of evaporative fuel emissions (EFEs) has become a serious target for environmentalists. Approximately 10% of EFEs escape during refuelling.

The vehicle emission control system that captures fuel vapours from the vehicle gas tank during refuelling is known as onboard refuelling vapour recovery (ORVR). The fuel tank and fill pipe are designed so that when refuelling the vehicle, fuel vapours in the tank travel to an activated carbon packed canister, which adsorbs the vapour. When the engine is in operation, it draws the vapours into the engine intake manifold to be used as fuel. ORVR has been mandated on all passenger cars in the United States since 2000 by the United States Environmental Protection Agency. The use of onboard vapour recovery is intended to make vapour recovery at fuel stations obsolete.

Table 9.2 Crankcase and evaporative emissions

Source	Comments
Fuel evaporation from the tank and system	Fuel evaporation causes hydrocarbons to be produced. The effect is greater as temperature increases. A charcoal canister is the preferred method for reducing this problem. The fuel tank is usually run at a pressure just under atmospheric by a connection to the intake manifold drawing the vapour through the charcoal canister. This must be controlled by the management system however, as even a 1% concentration of fuel vapour would shift the lambda value by 20%. This is done by using a 'purge valve', which under some conditions is closed (full-load and idle for example) and can be progressively opened under other conditions. The system monitors the effect by use of the lambda sensor signal.
Crankcase fumes (blow by)	Hydrocarbons become concentrated in the crankcase mostly due to pressure blowing past the piston rings. These gases must be conducted back into the combustion process. This is usually via the air intake system. This is described as positive crankcase ventilation.

In the EU the capture of these emissions has been made the responsibility of the fuelling stations.

This still leaves the matter of preventing evaporation from the fuel line itself, another key problem for car manufacturers. Technological advances in design actually increase fuel evaporation from within the fuelling system. This is because of the increasing use of plastics, rather than metal, for manufacturing fuel lines. Plastics allow petrol vapour to permeate through into the atmosphere. The proximity of catalytic converters, which generate tremendous heat to the fuel tank and the under-body shielding, contributes to making the fuel hotter and therefore more liable to evaporate.

Table 9.2 describes this issue further and also looks at crankcase emissions.

Evaporative emissions are measured in a 'shed'! This Sealed Housing for Evaporative Determination (SHED) is used in two ways:

- The vehicle with 40% of its maximum fuel is warmed up (from about 14–28 °C) in the shed and the increased concentration of the hydrocarbons measured.
- The vehicle is first warmed up over the normal test cycle and then placed in the shed. The increase in HC concentration is measured over one hour.

9.2.4 Leaded and unleaded fuel

Tetra-ethyl lead was first added to petrol in the 1920s to slow down the rate of burning, improve combustion and increase the octane rating of the fuel. All this was achieved using the lead additive - at less cost than further refining by the petrol companies.

The first real push for unleaded fuel was from Los Angeles in California. To reduce this city's severe smog problem, the answer at the time seemed to be to employ catalytic converters. However, if leaded fuel is used, the 'cat' can be rendered inoperative. A further study showing that lead causes brain damage in children sounded the death knell for leaded fuel. This momentum spread worldwide and still exists.

Key fact

Tetra-ethyl lead was first added to petrol in the 1920s to slow down the rate of burning, improve combustion and increase the octane rating of the fuel.

New evidence is now coming to light showing that the additives used instead of lead were ending up in the environment. The two main culprits are benzene, which is strongly linked to leukaemia, and methyl tertiary-butyl ether (MTBE), which poisons water and is very toxic to almost all living things. This is potentially a far worse problem than lead, which is now not thought to be as bad as the initial reaction suggested.

MTBE as a petrol/gasoline additive, is used as an oxygenate and to raise the octane number. Its use has declined in response to environmental and health concerns. It has been found to easily pollute large quantities of groundwater when fuel with MTBE is spilled or leaked at refuelling stations. MTBE spreads more easily underground than other gasoline components due to its higher solubility in water

It is important, however, to note that this is still in the ‘debate’ stage; further research is necessary for a fully reasoned conclusion. Note though, how any technological issue usually has far more to it than first meets the eye!

Modern engines are now designed to run on unleaded fuel, with one particular modification being hardened valve seats. In Europe and other places, leaded fuel has now been phased out completely. This is a problem for owners of classic vehicles. Many additives are available but these are not as good as lead. Here is a list of comments I have collated from a number of sources.

- All engines with cast iron heads and no special hardening of the exhaust valve seats will suffer some damage running on unleaded. The extent of the damage depends on the engine and on the engine revs.
- No petrol additives prevent valve seat recession completely. Some are better than others but none replace the action of lead.
- The minimum critical level of lead in the fuel is about 0.07 g Pb/l. Current levels in some leaded fuel are 0.15 g Pb/l and so mixing alternate tanks of leaded and unleaded is likely to be successful.
- It is impossible to predict wear rates accurately and often wear shows up predominantly in only one cylinder.
- Fitting hardened valve seats or performing induction hardening on the valve seats is effective in engines where either of these processes can be done.
- Tests done by Rover appear to back up the theory that, although unleaded petrol does damage all iron heads, the less spirited driver will not.

9.3 Emissions and driving cycles

9.3.1 Exhaust emission regulations

Limits and regulations relating to exhaust emissions vary in different countries and in different situations. For example, in the UK certain limits have to be met during the annual test. The current test default limits (for vehicles since September 2002 fitted with a catalytic converter) are:

Minimum oil temperature 60 °C

Fast idle (2500 to 3000 rpm)

- CO <= 0.2%
- HC <= 200 ppm
- Lambda 0.97 to 1.03

Idle (450 to 1500 rpm)

- CO <= 0.3%

Table 9.3 European past and future emission limits

Emissions Standard	Particulate matters (PM)/(mg/km)		Oxides of nitrogen (NOx) (mg/km)		Hydrocarbons (HC) (mg/km)	
	Diesel	Petrol	Diesel	Petrol	Diesel	Petrol
Euro 2 (1996)	80–100	–	–	–	–	–
Euro 3 (2000)	50	–	500	150	–	200
Euro 4 (2005)	25	–	250	80	–	100
Euro 5 (2009)	5	5	180	70	–	100
Euro 6 (2014)	5	5	80	70	–	100

Manufacturers however, have to meet stringent regulations when producing new vehicles. In Europe the emission standards are defined in a series of EU directives staging the progressive introduction of increasingly stringent standards (see Table 9.3).

Key fact

Manufacturers have to meet stringent regulations when producing new vehicles.

In the USA, what are known as Tier II standards are divided into several numbered 'bins' (see Table 9.4). Eleven bins were initially defined, with bin 1 being the cleanest (Zero Emission Vehicle) and 11 the dirtiest. However, bins 9, 10 and 11 are temporary. Only the first ten bins were used for light-duty vehicles below 8500 pounds GVWR, but medium-duty passenger vehicles up to 10 000 pounds (4536 kg) GVWR and to all 11 bins. Manufacturers can make vehicles which fit into any of the available bins, but still must meet average targets for their entire fleets.

The two least-restrictive bins for passenger cars, 9 and 10, were phased out at the end of 2006. However, bins 9 and 10 were available for classifying a restricted number of light-duty trucks until the end of 2008, when they were removed along with bin 11 for medium-duty vehicles. As of 2009, light-duty trucks must meet the same emissions standards as passenger cars.

Phase 2 was 2004 to 2009 and now even more stringent standards are coming into use. Also, the California Air Resources Board (CARB) may also adopt and enforce its own emissions standards. However, regardless of whether a manufacturer receives CARB approval, all new motor vehicles and engines must still receive certification from the environmental protection agency (EPA) before a vehicle is introduced.

In Europe the EU has recently set new CO_2 targets for vans. These are 175 g/km by 2017 and 147 g/km by 2020 (to be reviewed in 2013). The 2017 target will be phased in from 2014 after the introduction of the Euro 6 emissions limits.

9.3.2 Test cycles

All new vehicles must pass stringent emissions tests before being 'approved'. This is done by running the vehicle through test cycles and collecting the exhaust for analysis.

9.3.2.1 Europe

The New European Driving Cycle (NEDC) is a driving cycle consisting of four repeated ECE-15 driving cycles and an Extra-Urban driving cycle, or EUDC. The NEDC is meant to represent the typical usage of a car in Europe, and is

Table 9.4 Tier 2 exhaust emission standards (USA)

Standard		**Emission limits at 50 000 miles**					**Emission limits at full useful life (120 000 miles)[a]**				
		NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)	NOx (g/mi)	NMOG (g/mi)	CO (g/mi)	PM (g/mi)	HCHO (g/mi)
Federal	Bin 1	–	–	–	–	–	0	0	0	0	0
	Bin 2	–	–	–	–	–	0.02	0.01	2.1	0.01	0.004
	Bin 3	–	–	–	–	–	0.03	0.055	2.1	0.01	0.011
	Bin 4	–	–	–	–	–	0.04	0.07	2.1	0.01	0.011
	Bin 5	0.05	0.075	3.4	–	0.015	0.07	0.09	4.2	0.01	0.018
	Bin 6	0.08	0.075	3.4	–	0.015	0.1	0.09	4.2	0.01	0.018
	Bin 7	0.11	0.075	3.4	–	0.015	0.15	0.09	4.2	0.02	0.018
	Bin 8	0.14	0.100/0.125[c]	3.4	–	0.015	0.2	0.125/ 0.156	4.2	0.02	0.018
	Bin 9[b]	0.2	0.075/0.140	3.4	–	0.015	0.3	0.090/ 0.180	4.2	0.06	0.018
	Bin 10[b]	0.4	0.125/0.160	3.4/ 4.4	–	0.015/ 0.018	0.6	0.156/ 0.230c	4.2/ 6.4	0.08	0.018/ 0.027
	Bin 11[b]	0.6	0.195	5	–	0.022	0.9	0.28	7.3	0.12	0.032

[a] In lieu of intermediate useful life standards (50,000 miles) or to gain additional nitrogen oxides credit, manufacturers may optionally certify to the Tier 2 exhaust emission standards with a useful life of 150,000 miles.

[b] Bins 9–11 expire in 2006 for light-duty vehicles and light light-duty trucks and 2008 for heavy light-duty trucks and medium-duty passenger vehicles.

[c] Pollutants with two numbers have a separate certification standard (1st number) and in-use standard (2nd number).

Key fact

Driving cycles are meant to represent the typical usage of a car.

used, among other things, to measure emissions. It is sometimes referred to as MVEG cycle (Motor Vehicle Emissions Group).

The *old* European ECE-15 driving cycle lies between 0s and 800s and represented an urban drive cycle (Figure 9.14). The section from 800s represents a suburban drive cycle, and is now called the New European Driving Cycle (NEDC).

The cycle must be performed on a cold vehicle at 20 °C (68 °F). The cycles may be performed on a normal flat road, in the absence of wind. However, to improve repeatability, they are generally performed on a rolling road.

Several measurements are usually performed during the cycle. The figures made available to the general public are:

- Urban fuel economy (first 800 seconds).
- Extra-urban fuel economy (800 to 1200s).
- Overall fuel economy (complete cycle).
- CO_2 emission (complete cycle).

The following parameters are also generally measured to validate the compliance to European emission standards:

- Carbon monoxide (CO).
- Unburnt hydrocarbons (HC).
- Nitrogen oxides (NOx).
- Particulate matter (PM).

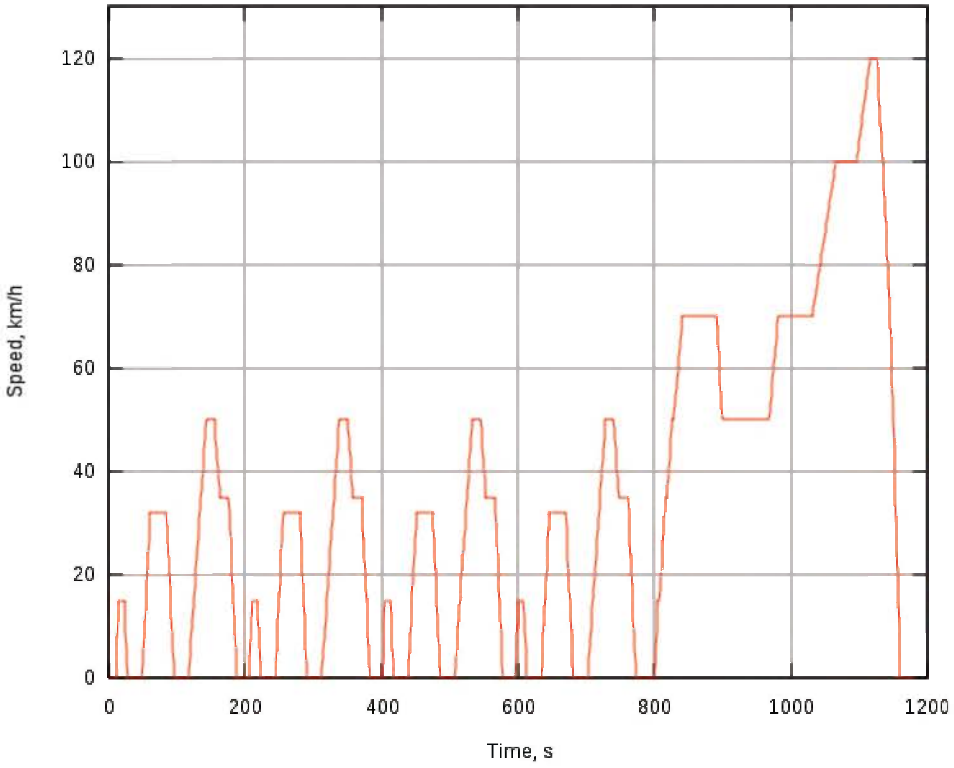


Figure 9.14 New European Driving Cycle (NEDC)

A further tightening of the driving cycle is the Modified New European Driving Cycle (MNEDC), which is very similar to the NEDC except that there is no warm up time at the start (Figure 9.15).

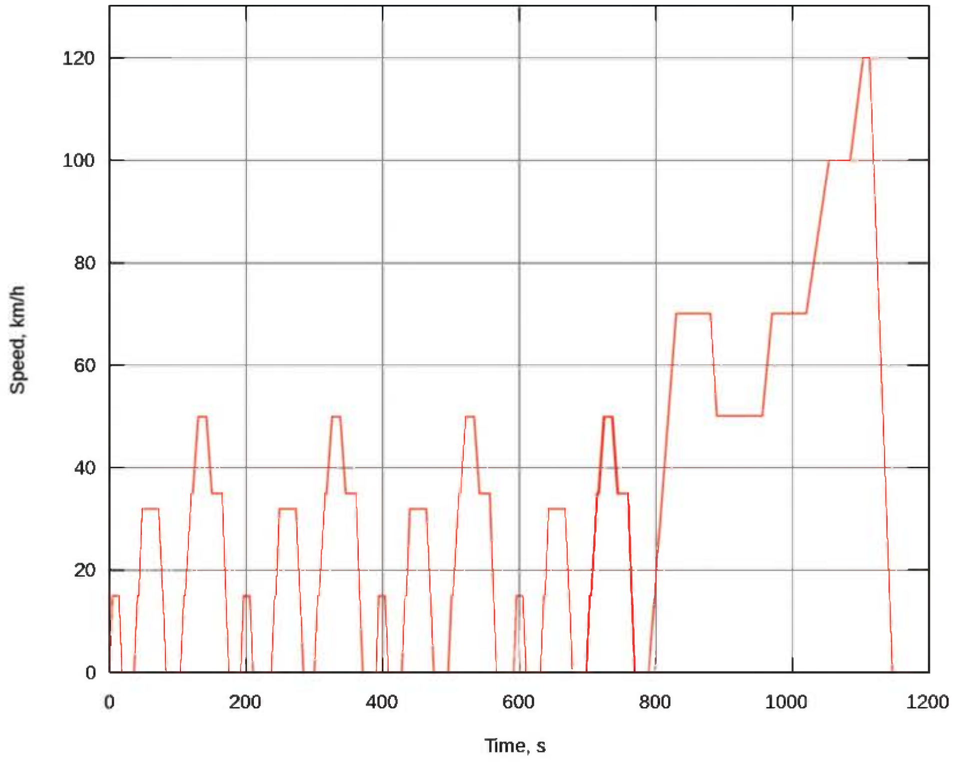


Figure 9.15 Modified New European Driving Cycle (MNEDC)

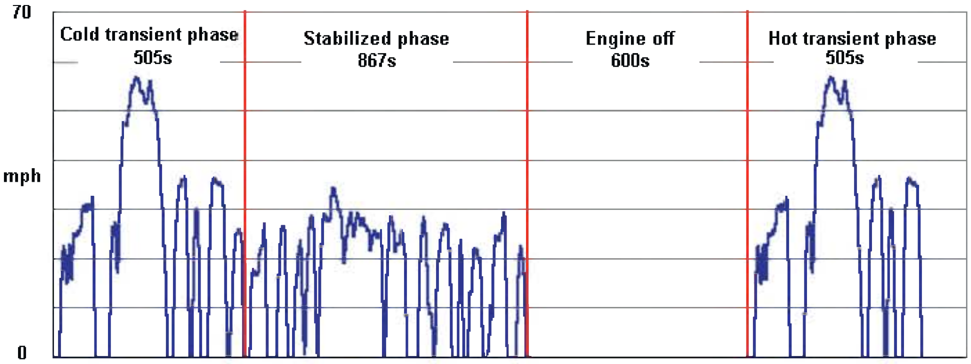


Figure 9.16 USA Federal Test Procedure

9.3.2.2 USA

In the USA a cycle known as the Federal Test Procedure FTP-75 is used (Figure 9.16). This has been added to and became known as the Supplementary Federal Test Procedure (SFTP).

9.4 Electronic control of carburation

9.4.1 Basic carburation

Figure 9.17 shows a simple fixed choke carburettor, which shows the principles and operation of this device. The float and needle valve assembly ensure a constant level of petrol in the float chamber. The Venturi causes an increase in air speed and hence a drop in pressure in the area of the outlet. The main jet regulates how much fuel can be forced into this intake air stream by the higher pressure now apparent in the float chamber. The basic principle is that as more air is forced into the engine then more fuel will be mixed into the air stream.

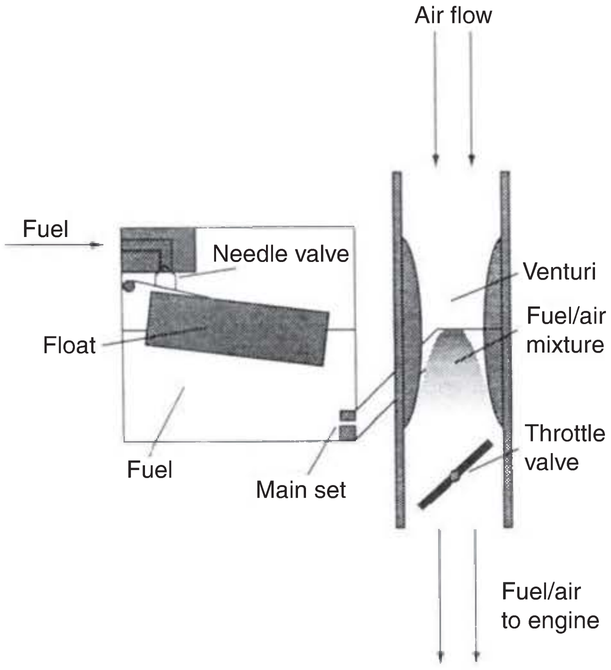


Figure 9.17 Simple fixed choice carburettor

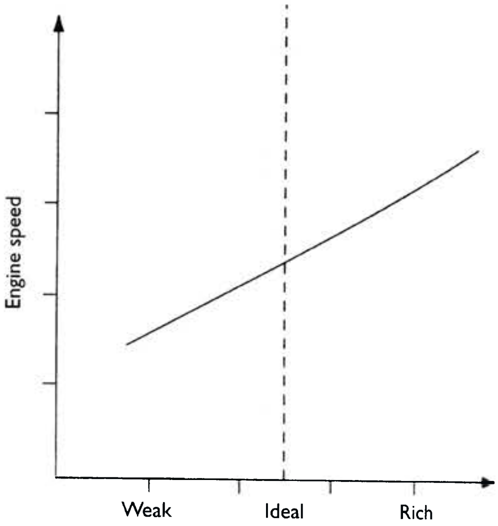


Figure 9.18 Fuel forced into the air stream does not linearly follow the increase in air quantity with a simple fixed choke carburettor

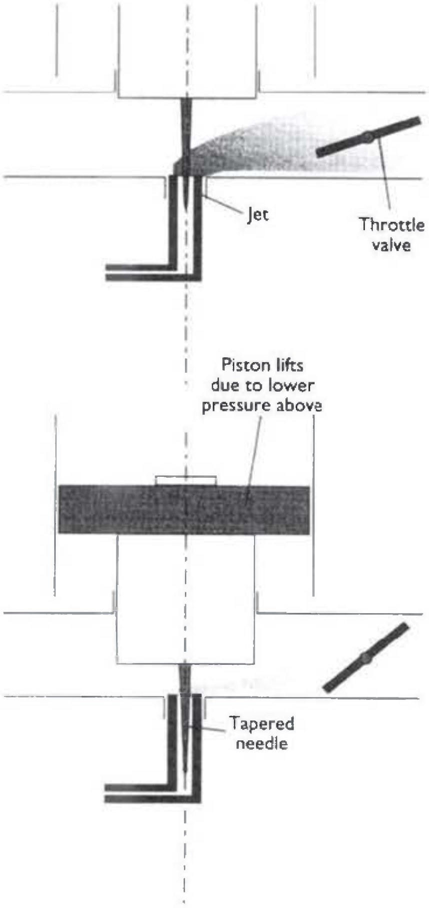


Figure 9.19 Variable Venturi carburettor

Figure 9.18 shows the problem with this very simple system; the amount of fuel forced into the air stream does not linearly follow the increase in air quantity. This means further compensation fuel and air jets are required to meet all operating requirements.

Figure 9.19 shows a variable Venturi carburettor, which keeps the air pressure in the Venturi constant, and uses a tapered needle to control the amount of fuel.

9.4.2 Areas of control

One version of the variable Venturi carburettor (Figure 9.20) has been used with electronic control. In general, electronic control of a carburettor is used in the following areas.

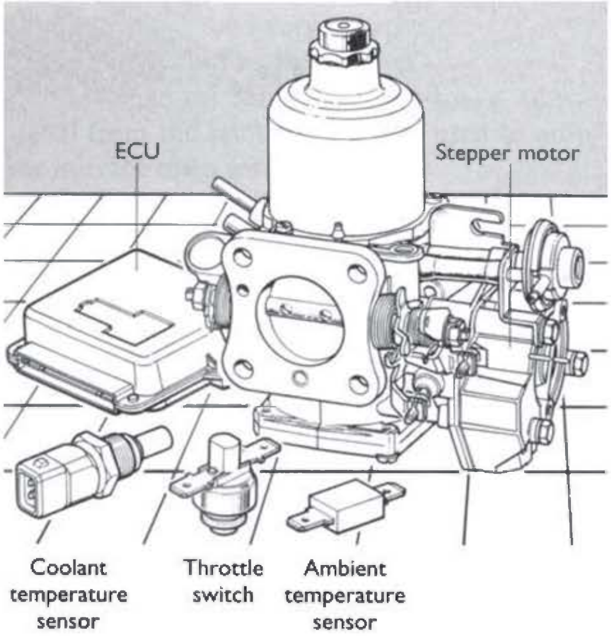


Figure 9.20 HIF variable Venturi carburettor with electronic control components

Idle speed

Controlled by a stepper motor to prevent stalling but still allow a very low idle speed to improve economy and reduce emissions. Idle speed may also be changed in response to a signal from an automatic gearbox to prevent either the engine from stalling or the car from trying to creep.

Fast idle

The same stepper motor as above controls fast idle in response to a signal from the engine temperature sensor during the warm up period.

Choke (warm up enrichment)

A rotary choke or some other form of valve or flap operates the choke mechanism depending on engine and ambient temperature conditions.

Overrun fuel cut off

A small solenoid operated valve or similar cuts off the fuel under particular conditions. These are often that the engine temperature is above a set level, the engine speed is above a set level and that the accelerator pedal is in the off position.

The main control of the air–fuel ratio is a function of the mechanical design and is very difficult to control by electrical means. Some systems have used electronic control of a needle and jet but this did not prove to be very popular.

9.5 Fuel injection

9.5.1 Advantages of fuel injection

The major advantage of any type of fuel injection system is accurate control of the fuel quantity injected into the engine. The basic principle of fuel injection is that if petrol is supplied to an injector (electrically controlled valve), at a constant differential pressure, then the amount of fuel injected will be directly proportional to the injector open time.

Key fact

The major advantage of any type of fuel injection system is accurate control of the fuel quantity injected into the engine.

Most systems are now electronically controlled even if containing some mechanical metering components. This allows the operation of the injection system to be very closely matched to the requirements of the engine. This matching process is carried out during development on test beds and dynamometers, as well as development in the car. The ideal operating data for a large number of engine operating conditions are stored in a read only memory in the ECU. Close control of the fuel quantity injected allows the optimum setting for mixture strength when all operating factors are taken into account (see the air–fuel ratio section).

Further advantages of electronic fuel injection control are that overrun cut off can easily be implemented, fuel can be cut at the engine's rpm limit and information on fuel used can be supplied to a trip computer.

Fuel injection systems can be classified into two main categories:

- Single-point injection – see Figure 9.21.
- Multipoint injection – see Figure 9.22.

Both of these systems are discussed in more detail in later sections of this chapter.

9.5.2 System overview

Figure 9.23 shows a typical control layout for a fuel injection system. Depending on the sophistication of the system, idle speed and idle mixture adjustment can be either mechanically or electronically controlled.

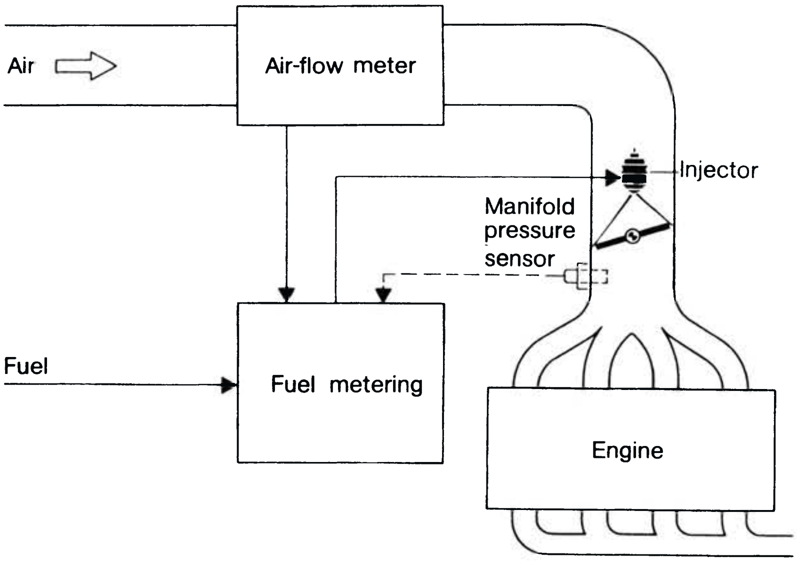


Figure 9.21 Fuel Injection, single-point

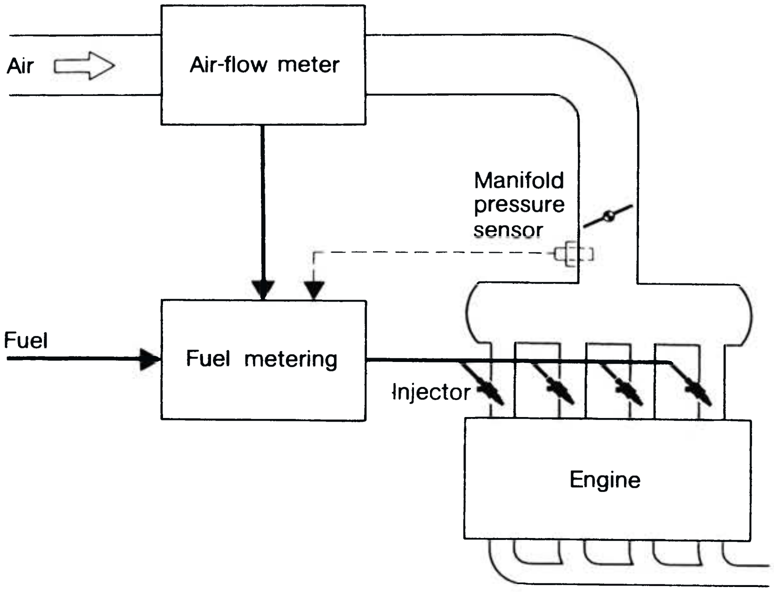


Figure 9.22 Fuel injection, multipoint

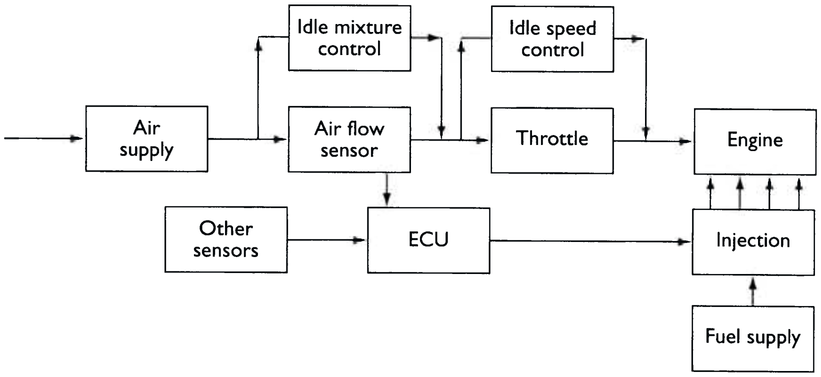


Figure 9.23 Typical control layout for a fuel injection system

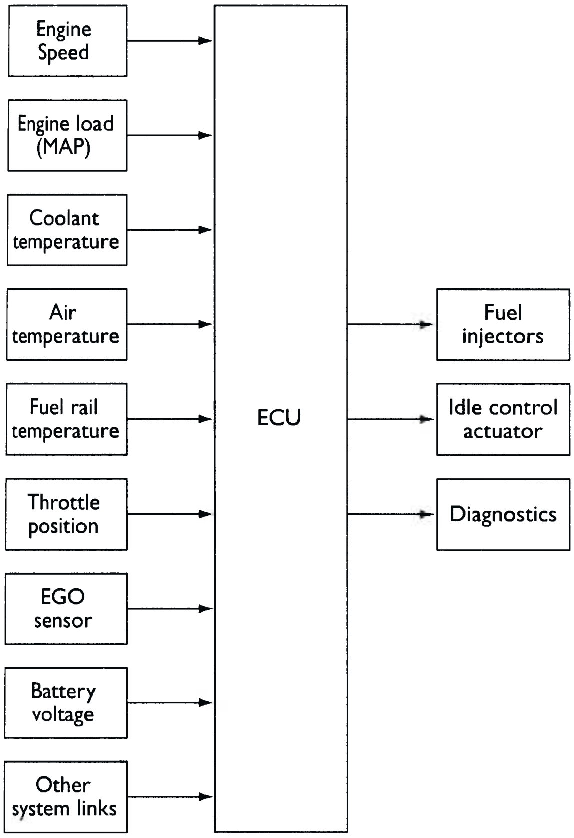


Figure 9.24 Block diagram of inputs and outputs common to most fuel injection systems

Figure 9.24 shows a block diagram of inputs and outputs common to most fuel injection systems. Note that the two most important input sensors to the system are speed and load. The basic fuelling requirement is determined from these inputs in a similar way to the determination of ignition timing, as described in a previous section.

A three-dimensional cartographic map, shown in Figure 9.25, is used to represent how the information on an engine's fuelling requirements is stored.

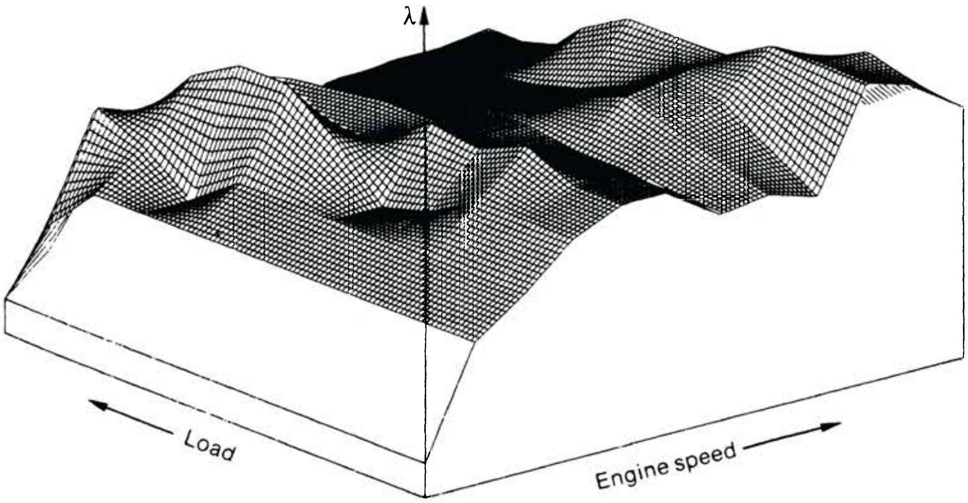


Figure 9.25 Cartographic map used to represent how the information on an engine's fuelling requirements are stored

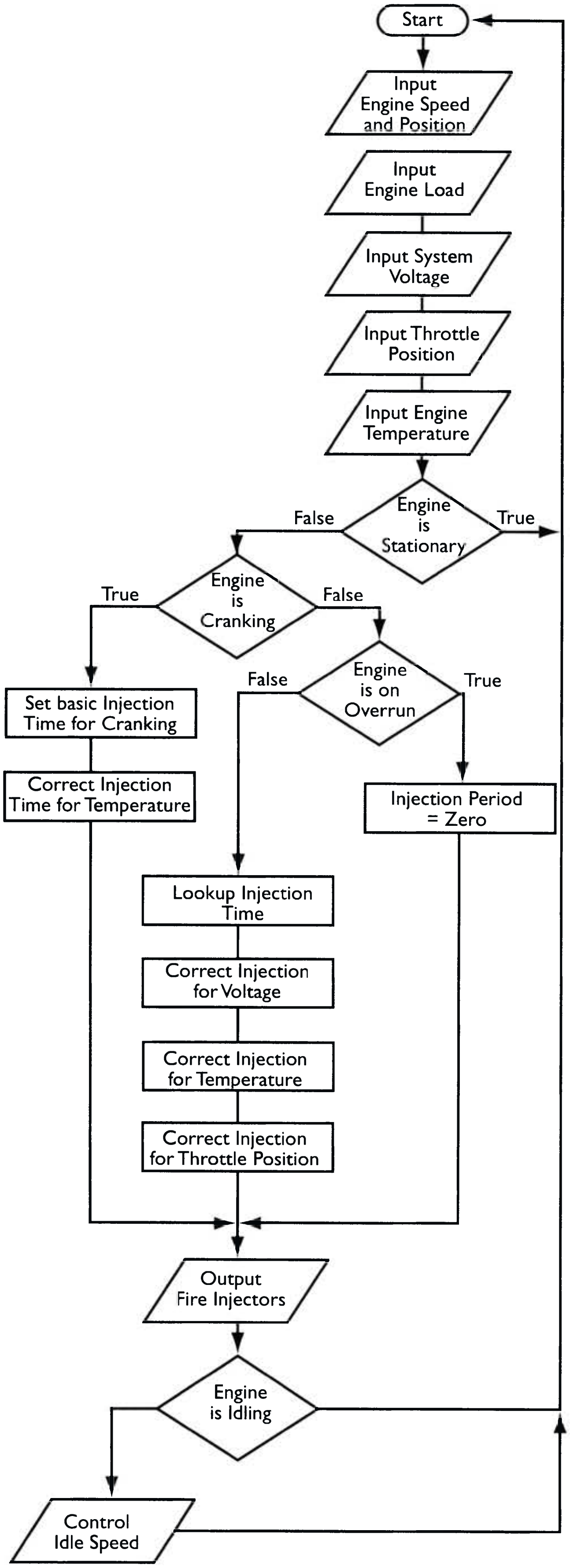


Figure 9.26 Fuel and idle speed flow diagram

This information forms part of a read only memory (ROM) chip in the ECU. When the ECU has determined the look-up value of the fuel required (injector open time), corrections to this figure can be added for battery voltage, temperature, throttle change or position and fuel cut off.

Idle speed, and fast idle, are also controlled by the ECU and a suitable actuator. It is also possible to have a form of closed loop control with electronic fuel injection. This involves a lambda sensor to monitor exhaust gas oxygen content. This allows very accurate control of the mixture strength, as the oxygen content of the exhaust is proportional to the air–fuel ratio. The signal from the lambda sensor is used to adjust the injector open time.

Key fact

A lambda sensor is used to monitor exhaust gas oxygen content.

Figure 9.26 is a flow chart showing one way in which the information from the sensors could be processed to determine the best injector open duration as well as control of engine idle speed.

9.5.3 Components of a fuel injection system

The following components, with some additions, are typical of the Bosch ‘L’ Jetronic systems. This system is not now used as almost all cars have full engine management (i.e. combined ignition and fuel control systems). However, it is a good example to use as an introduction to the subject. The various components are only discussed briefly, as most are included in other sections in more detail.

Flap type air flow sensor (Figure 9.27)

A Bosch vane-type sensor is shown which moves due to the air being forced into the engine. The information provided to the ECU is air quantity and engine load.

Engine speed sensor (Figure 9.28)

Most injection systems, which are not combined directly with the ignition, take a signal from the coil negative terminal. This provides speed data but also engine position to some extent. A resistor in series is often used to prevent high voltage surges reaching the ECU.

Temperature sensor (Figure 9.29)

A simple thermistor provides engine coolant temperature information.

Throttle position sensor (Figure 9.30)

Various sensors are shown consisting of the two-switch types, which only provide information that the throttle is at idle, full load or anywhere else in between; and potentiometer types, which give more detailed information.

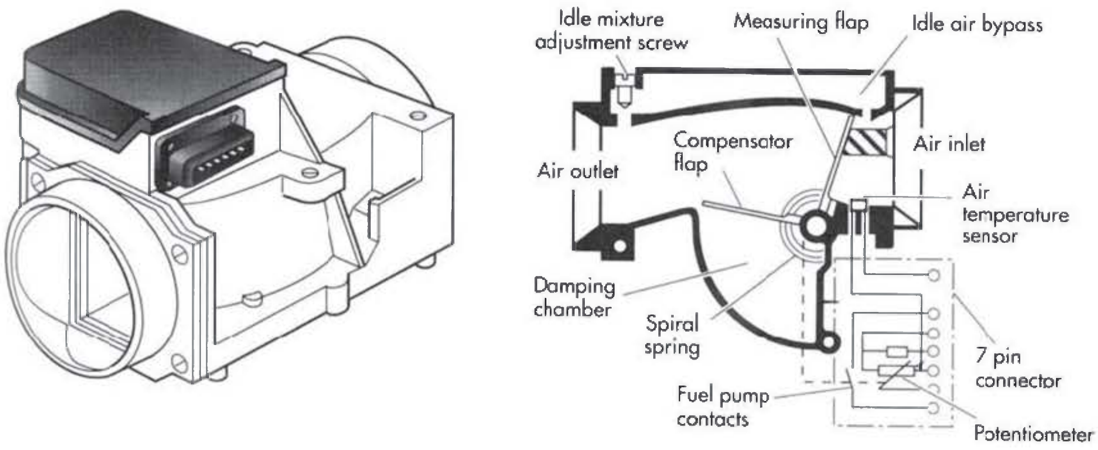


Figure 9.27 Air flow meter

Figure 9.28 Crank sensor in position

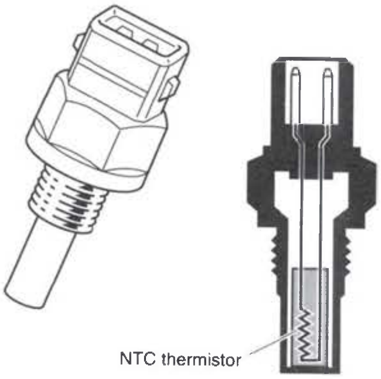


Figure 9.29 Coolant temperature sensor

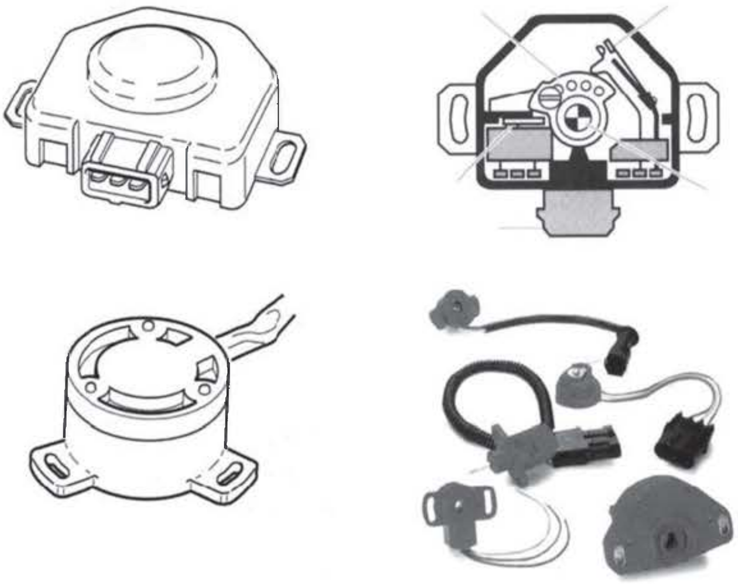


Figure 9.30 Throttle position sensors

Lambda sensor (Figure 9.31)

This device provides information to the ECU on exhaust gas oxygen content. From this information, corrections can be applied to ensure the engine is kept at or very near to stoichiometry. Also shown in this figure is a combustion chamber pressure sensor.

Idle or fast idle control actuator (Figure 9.32)

Bimetal or stepper motor actuators are used but the one shown is a pulsed actuator. The air that it allows through is set by its open/close ratio.

Fuel injector(s) (Figure 9.33)

Two types are shown – the pintle and disc injectors. They are simple solenoid-operated valves designed to operate very quickly and produce a finely atomized spray pattern.

Injector resistors

These resistors were used on some systems when the injector coil resistance was very low. A lower inductive reactance in the circuit allows faster operation of the injectors. Most systems now limit injector maximum current in the ECU in much the same way as for low resistance ignition on coils.

Figure 9.31 Lambda sensor in the exhaust downpipe

Figure 9.32 Rotary idle actuator

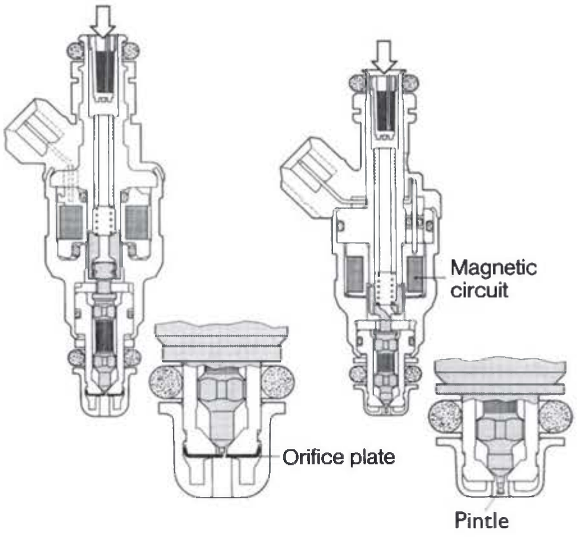


Figure 9.33 Fuel Injector

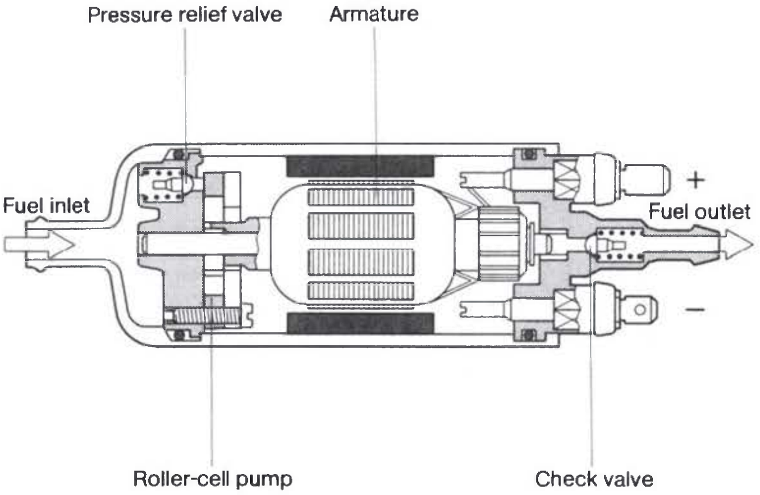


Figure 9.34 Fuel pump (high pressure)

Key fact

The pump ensures a constant supply of fuel to the fuel rail.

Fuel pump (Figure 9.34)

The pump ensures a constant supply of fuel to the fuel rail. The volume in the rail acts as a swamp to prevent pressure fluctuations as the injectors operate. The pump must be able to maintain a pressure of about 3 bar.

Fuel pressure regulator (Figure 9.35)

This device ensures a constant differential pressure across the injectors. It is a mechanical device and has a connection to the inlet manifold.

Cold start injector and thermo-time switch (Figure 9.36)

An extra injector was used on earlier systems as a form of choke. This worked in conjunction with the thermo-time switch to control the amount of cold enrichment. Both engine temperature and a heating winding heat it. This technique has been replaced on newer systems, which enrich the mixture by increasing the number of injector pulses or the pulse length.

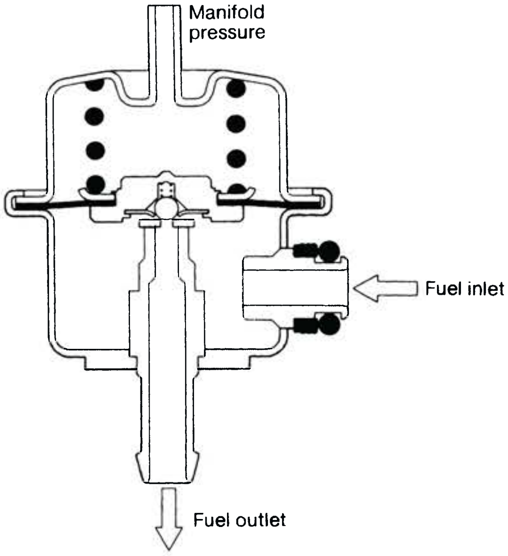


Figure 9.35 Pressure regulator

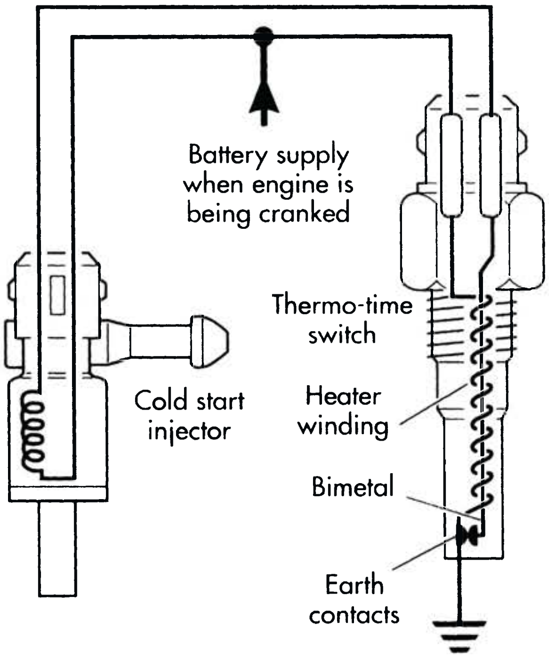


Figure 9.36 Typical cold start arrangement

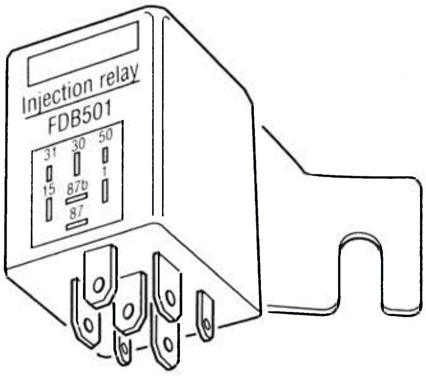


Figure 9.37 Combination relay

Combination relay (Figure 9.37)

This takes many forms on different systems but is basically two relays, one to control the fuel pump and one to power the rest of the injection system. The relay is often controlled by the ECU or will only operate when ignition pulses are sensed as a safety feature. This will only allow the fuel pump to operate when the engine is being cranked or is running.

Electronic control unit (Figure 9.38)

Earlier ECUs were analogue in operation. All ECUs now in use employ digital processing.

Figure 9.38 Electronic control unit

9.5.4 Bosch 'L' Jetronic – Variations

Owing to continued demands for improvements, the 'L' Jetronic system, along with all other injection systems, developed and changed over the years. This section will highlight the main changes that have taken place.

L2-Jetronic

This system is changed little except for the removal of the injector series resistors as the ECU now limits the output current to the injectors. The injector resistance is 16 Ω.

LE1-Jetronic

No current resistors are used and the throttle switch is adjustable. The fuel pump does not have safety contacts in the air flow sensor. The safety circuit is incorporated in the electronic relay. This will only allow the fuel pump to operate when an ignition signal is present; that is, when the engine is running or being cranked.

LE2-Jetronic

This is very similar to the LE1 systems except the thermo-time switch and cold start injector are not used. The ECU determines cold starting enrichment and adjusts the injector open period accordingly.

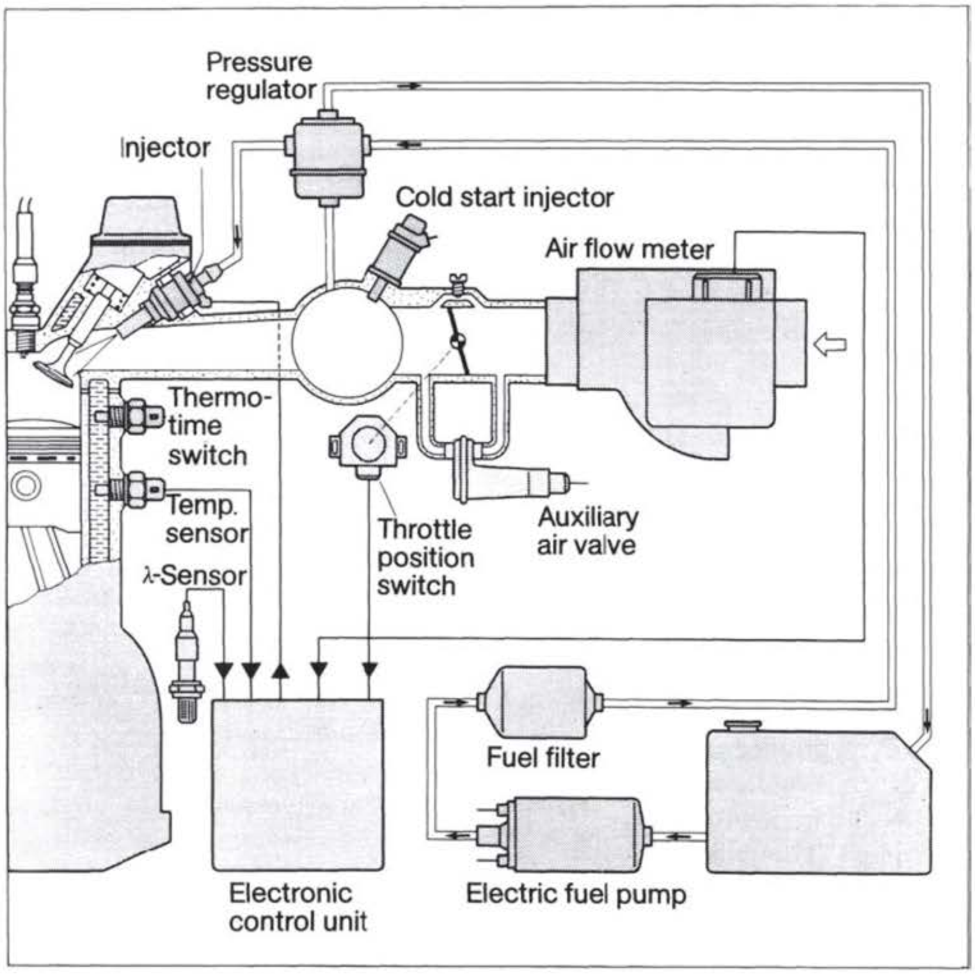


Figure 9.39 L-Jetronic

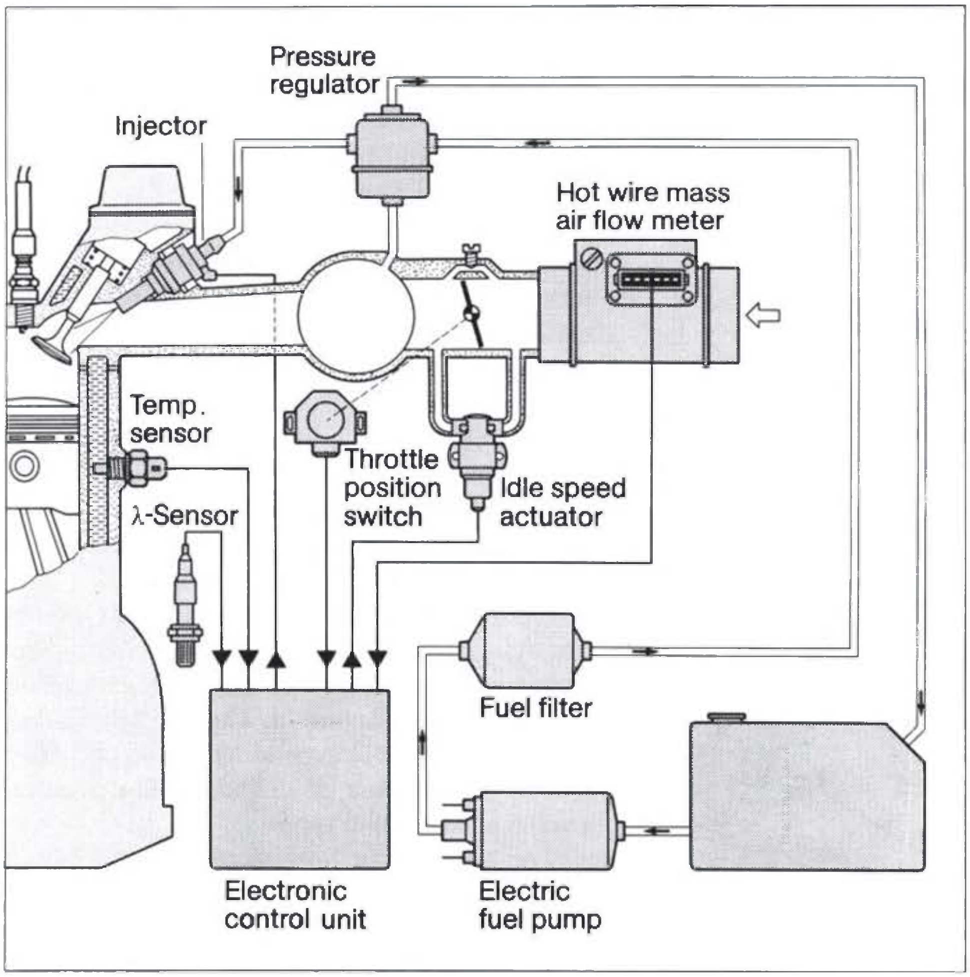


Figure 9.40 LH-Jetronic

LU-Jetronic

This system is a further refinement of the LE systems but also utilizes closed loop lambda control.

L3-Jetronic

The ECU for the L3-Jetronic forms part of the air flow meter installation. The ECU now includes a 'limp home' facility. The system can be operated with or without lambda closed loop control. The air–fuel ratio can be adjusted by a screw-operated potentiometer on the side of the ECU.

LH-Jetronic

The LH system incorporates most of the improvements noted above. The main difference is that a hot-wire type of air flow meter is used. The component layout is shown in Figure 9.40. Further developments are continuing but, in general, most systems have now developed into combined fuel and ignition control systems as discussed in the next chapter.

9.5.5 Bosch Mono Jetronic – single point injection

The Mono Jetronic is an electronically controlled system utilizing just one injector positioned above the throttle butterfly valve. The throttle body assembly is similar in appearance to a carburettor and the system is often referred to as throttle body injection. A low pressure (1 bar) fuel supply pump, as shown in Figure 9.41, is used to supply the injector, which injects the fuel intermittently into the inlet manifold. In common with most systems, sensors measuring engine variables supply the operating data. The ECU computes the ideal fuel requirements and outputs to the injector. The width of the injector pulses determines the quantity of fuel introduced.

The injector for the system is a very fast-acting valve. Figure 9.42 shows the injector in section. A pintle on the needle valve is used and a conical spray pattern is produced. This ensures excellent fuel atomization and hence a better 'burn' in the cylinder. In order to ensure accurate metering of small fuel quantities the valve needle and armature have a very small mass. This permits opening and closing times of less than 1 ms. The fuel supply to the injector is continuous, this prevents air locks and a constant supply of cool fuel. This also provides for good hot starting performance, which can be inhibited by evaporation if the fuel is hot.

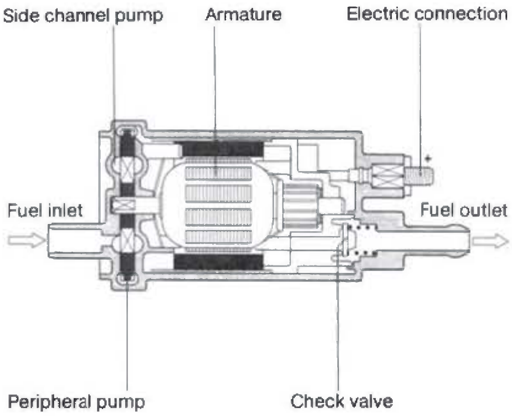


Figure 9.41 Electric fuel pump (low pressure)

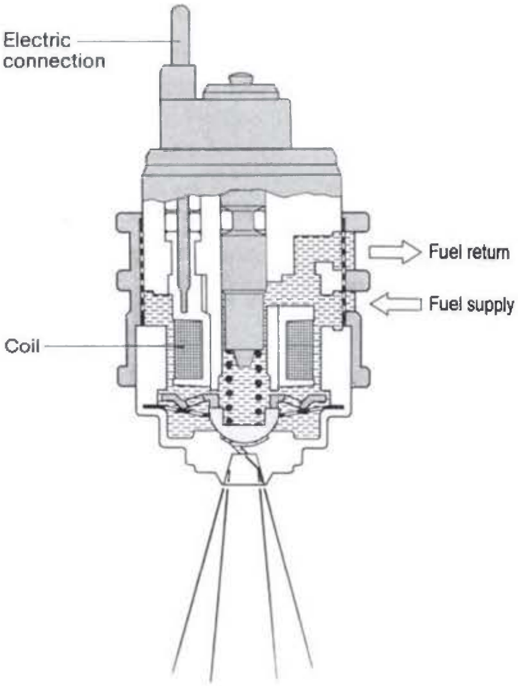


Figure 9.42 Low pressure injector (mini-Injector)

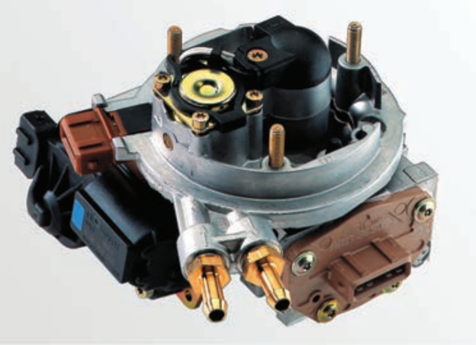


Figure 9.43 Central injection unit of the mono Jetronic

Figure 9.43 shows the main components of the Mono Jetronic system. The component most noticeable by its absence is an air flow sensor which is not used by this system. Air mass and load are calculated from the throttle position sensor, engine speed and air intake temperature. This is sometimes known as the speed density method. At a known engine speed with a known throttle opening, the engine will 'consume' a known volume of air. If the air temperature is known then the air mass can be calculated.

The basic injection quantity is generated in the ECU as a function of engine speed and throttle position. A ROM chip, represented by a cartographic map, stores data at 16 speed and 16 throttle angle positions, giving 256 references altogether. If the ECU detects deviations from the ideal air–fuel ratio by signals from the lambda sensor then corrections are made. If these corrections are required over an extended period then the new corrected values are stored in memory. These are continuously updated over the life of the system. Further corrections are added to this look-up value for temperature, full load and idle conditions. Over-run fuel cut-off and high engine speed cut-off are also implemented when required.

The Bosch Mono Jetronic system also offers adaptive idle control. This is to allow the lowest possible smoothed idle speed to reduce fuel consumption and exhaust emissions. A throttle valve actuator changes the position of the valve in response to a set speed calculated in the ECU, which takes into account the engine temperature and electrical loads on the alternator. The required throttle angle is computed and placed in memory. The adaptation capability of this system allows for engine drift during its life and also makes corrections for altitude.

The electronic control unit checks all signals for plausibility during normal operation. If a signal deviates from the normal, this fault condition is memorized and can be output to a diagnostic tester or read as a blink code from a fault lamp.

Key fact

Air mass and load can be calculated from the throttle position sensor, engine speed and air intake temperature. This is sometimes known as the speed density method.

Key fact

A single-point system injects the fuel in continuous pulses; whereas most multipoint injectors fire at the same time, injecting half of the required fuel.

9.5.6 Sequential multipoint injection

A single-point system injects the fuel in continuous pulses; whereas most multipoint injectors fire at the same time, injecting half of the required fuel. A sequential injection system injects fuel on the induction stroke of each cylinder in the engine firing order. This system, while more complicated, allows stratification of the cylinder charge to be controlled to some extent, resulting in an overall weaker charge. Sequential injection is normally incorporated with full engine management, which is discussed further in Chapter 10. Figure 9.44 shows a comparison between normal and sequential injection.

Key fact

A sequential injection system injects fuel on the induction stroke of each cylinder in the engine firing order.

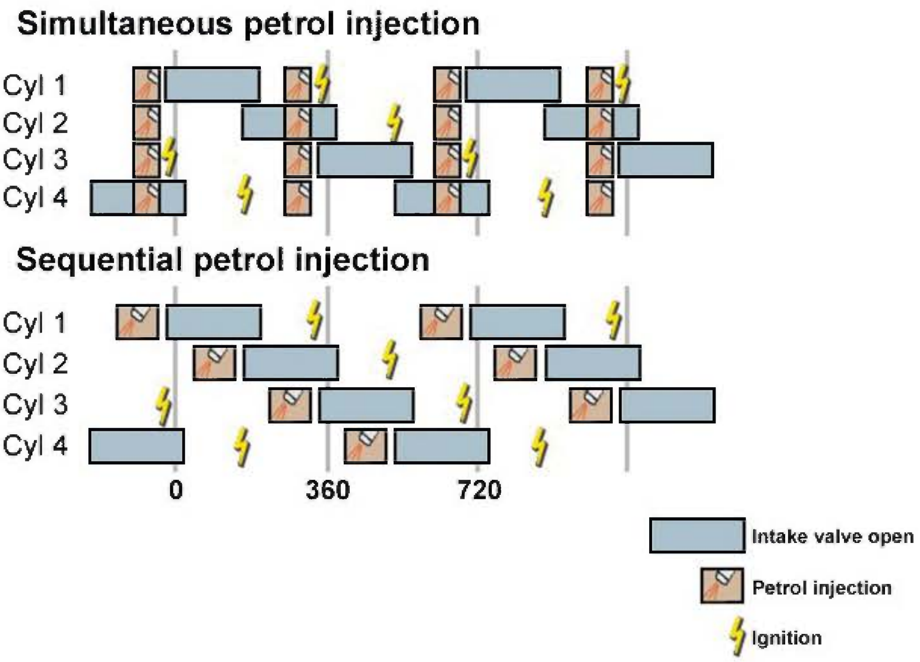


Figure 9.44 Simultaneous and sequential petrol injection

9.5.7 Lean burn technology

The optimum air–fuel ratio is 14.7 : 1 to ensure complete combustion. Increasing this ratio (introducing more air) results in what is known as lean burn. Fuel economy is maximized when the ratio is in the 20 to 22 : 1 range. Running leaner mixtures also reduces NOx emissions. However, the potential for unstable combustion increases. Reducing NOx emissions under lean burn conditions is difficult because the normal catalytic converter needs certain conditions to work properly. Mazda produced a ‘Z-lean engine’ that offers both a wide lean burn range and good power output at normal rpm. Figure 9.45 shows a cutaway view of this engine.

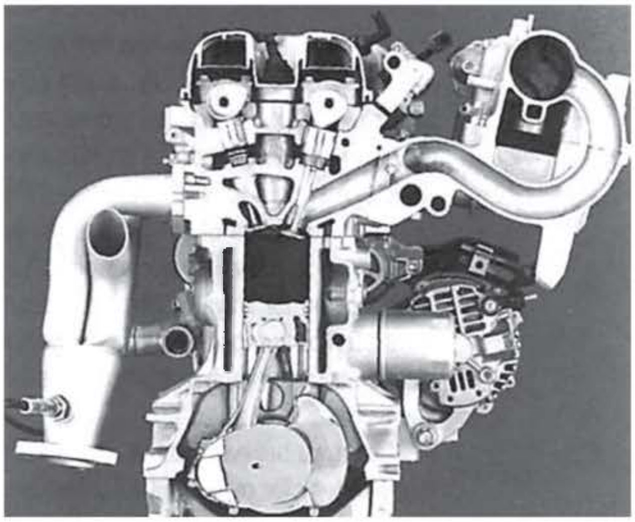


Figure 9.45 Cylinder-head and inlet path of a lean burn engine

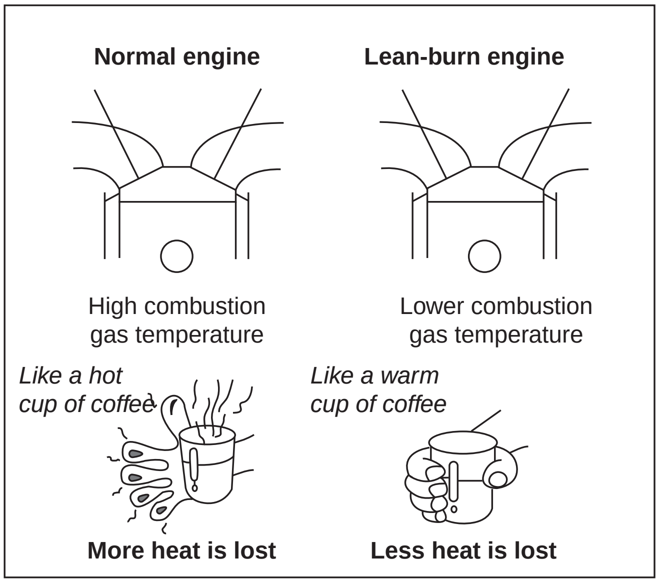
Introducing more air into the cylinder necessarily results in a lower fuel density in the mixture and thus a lower combustion temperature. This in turn means that less heat energy is lost from the combustion chamber to the surrounding parts of the engine. In addition to reduced heat loss, pumping loss is also lower because one can open the throttle wider when adjusting air input. These two effects contribute to the higher fuel economy of lean burn engines. Figure 9.46 shows these features.

The tumble swirl control (TSC) valve and its effects are shown as Figure 9.47. The Z-lean engine uses a feature known as a TSX (tumble swirl multiplex) port to control the vortex inside the cylinder. Combining this with an air mixture type injector which turns the fuel into a very fine spray and a high-energy ignition system ensures that it can operate on very lean mixtures up to 25 : 1. A special catalytic converter combines the NOx and HC into H_2O, CO_2 and N.

Key fact

Introducing more air into the cylinder results in a lower fuel density in the mixture and thus a lower combustion temperature.

■ **Reduced heat loss**



■ **Reduced pumping loss**

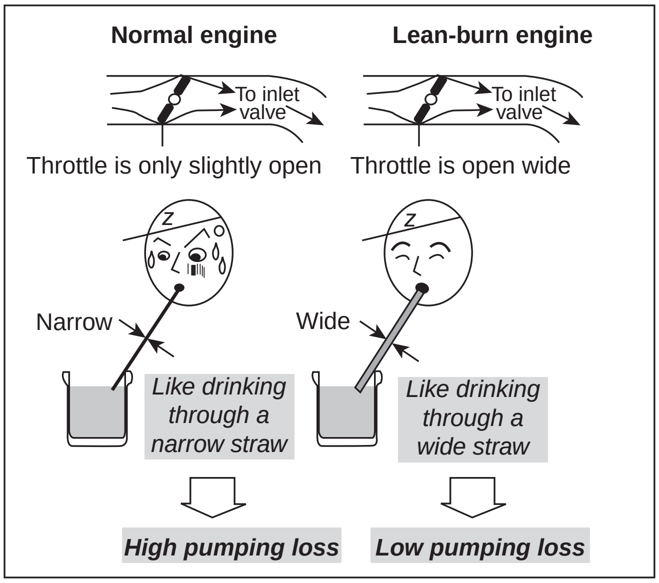
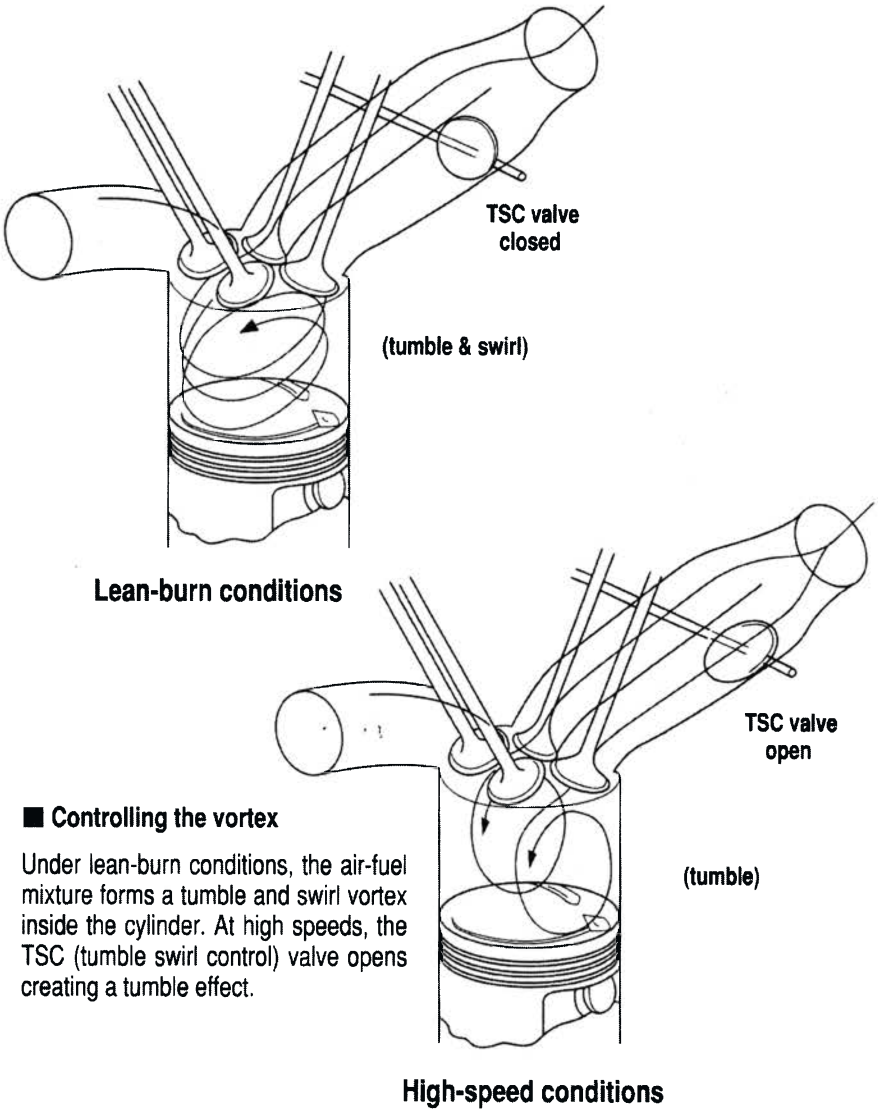


Figure 9.46 Features of the lean burn system



■ Controlling the vortex

Under lean-burn conditions, the air-fuel mixture forms a tumble and swirl vortex inside the cylinder. At high speeds, the TSC (tumble swirl control) valve opens creating a tumble effect.

Figure 9.47 Tumble swirl control

9.5.8 Double fuel injectors

Nissan have developed a fuel injection system that uses two injectors; one in each inlet port of a four-valve, four-cylinder engine. This has allowed 30 micron injector holes instead of the usual 70 microns. This use of two injectors and a finer spray has resulted in improved combustion efficiency.

Continuously variable valve timing on both inlet and exhaust valves makes it possible to close the exhaust valves very late, ‘losing’ some of the intake air at idle. This is known as the Atkinson high ratio expansion cycle (also used in the Toyota Prius). With this system the compression stroke is effectively shorter than the power stroke, compression is reduced and little of no combustion energy is lost through the exhaust.

This method is used to improve idle efficiency. It would normally result in an unstable idle but the finer fuel atomization makes mitigates this. With valve timing adjusted to achieve internal EGR to reduce pumping losses at part throttle and reduce NOx, the system can achieve a 4% reduction in CO_2 emission.

Definition

Micron: A unit of length equal to one millionth of a meter or 1 μm. The micron is not an official unit but is in common use.

9.6 Diesel fuel injection

9.6.1 Introduction

Diesel engines have the fuel injected into the combustion chamber where it is ignited by heat in the air charge. This is known as compression ignition (CI) because no spark is required. The high temperature needed to ignite the fuel is obtained by a high compression of the air charge. Diesel fuel is injected under high pressure from an injector nozzle, into the combustion chambers. The fuel is pressurised in a diesel injection pump. It is supplied and distributed to the injectors through high pressure fuel pipes or directly from a rail and/ or an injector. The high pressure generation is from a direct acting cam or a separate pump.

The air flow into a diesel engine is usually unobstructed by a throttle plate so a large air charge is always provided. Throttle plates may be used to provide control for emission devices. Engine speed is controlled by the amount of fuel injected. The engine is stopped by cutting off the fuel delivery. For all engine operating conditions a surplus amount of air is needed for complete combustion of the fuel.

Earlier diesel engine tended to be considered as indirect and direct injection. Nowadays, almost all are direct and there are a number of methods used as shown in Figures 9.48 and 9.49. The rotary pumped direct injection and common rail systems will be discussed further in this section.

Small high speed diesel engine compression ratios are about 20:1 for direct injection systems. This compression ratio is capable of raising the air charge to temperatures of between 500 °C and 800 °C. Very rapid combustion of the fuel therefore occurs when it is injected into the hot air charge.

The combustion process in a diesel engine follows three phases (Figure 9.51). These are ignition delay, flame spread and controlled combustion. In addition, an

Key fact

The high temperature needed to ignite fuel in a diesel engine is obtained by high compression of the air charge.

Figure 9.48 Diesel fuel injection components (Source: Bosch Media)

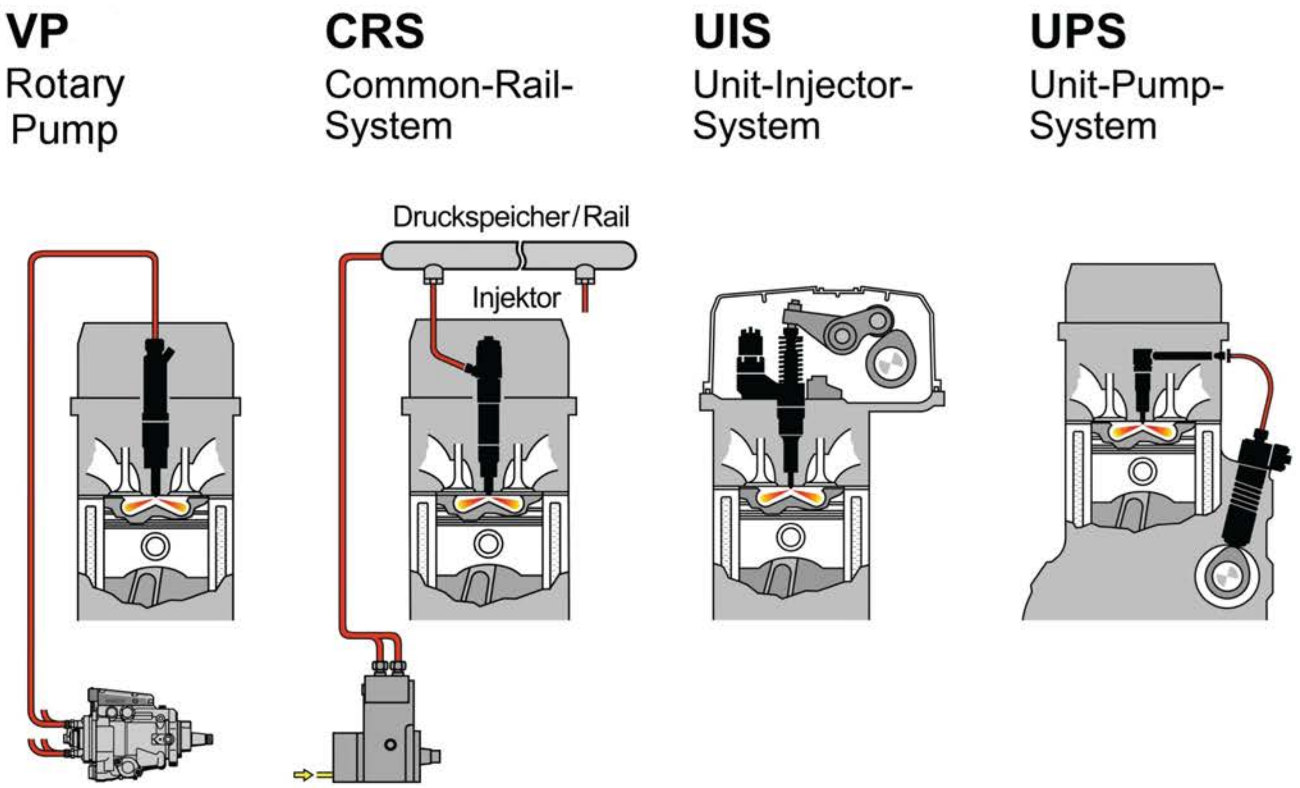


Figure 9.49 Types of diesel injection: Direct using a rotary pump, common rail, unit injection and pumped unit injection (Source: Bosch Media)

Key fact

The combustion process in a diesel engine follows three phases. These are ignition delay, flame spread and controlled combustion.

injection lag occurs in the high pressure pipes, of earlier systems, as the pressure builds up just before injection.

The most important phase of controlled combustion is when fuel is being injected into a burning mixture. This must be at a rate that maintains an even combustion pressure onto the piston throughout the critical crankshaft rotational angles. This

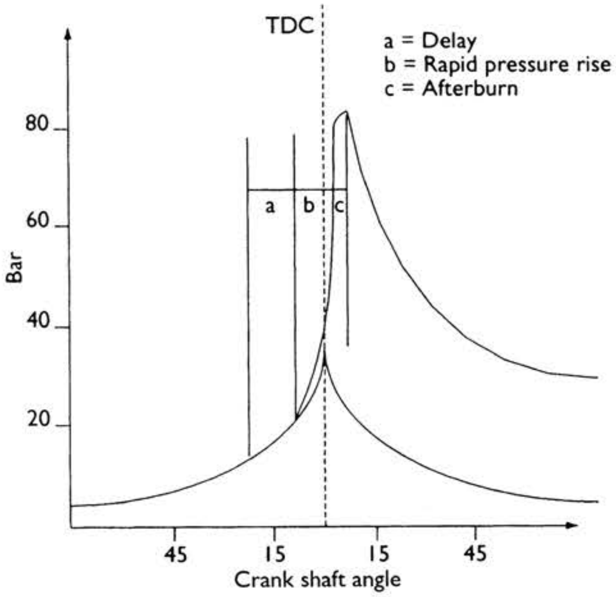


Figure 9.50 Combustion phases: a-delay, b- flame spread (rapid pressure rise), c-controlled combustion (afterburn)

gives maximum torque and efficient fuel usage, because temperatures remain controlled and the heat lost to the exhaust is minimised. The low temperatures also help to keep nitrogen oxide emissions (NOx) to a minimum.

The speed of flame spread in a diesel engine is affected by the air charge temperature and the atomisation of the fuel. These characteristics are shared with the delay period. A sufficiently high air charge temperature, of at least 450 °C, is a minimum requirement for optimum ignition and combustion.

The delay phase or ignition lag for diesel fuel combustion lasts a few milliseconds. It occurs immediately on injection as the fuel is heated up to the self-ignition temperature. The length of the delay is dependent on the compressed air charge temperature and the grade of fuel. The air charge temperature is also affected by the intake air temperature and the engine temperature.

A long delay period allows a high volume of fuel to be injected before ignition and flame spread occurs. In this situation diesel knock is at its most severe. When a diesel engine is cold, there may be insufficient heat in the air charge to bring the fuel up to the self-ignition temperature. When ignition is slow, heavy knocking occurs.

To aid starting and to reduce diesel knock, cold start devices may be used. For indirect injection engines, starting at lower than normal operating temperatures requires additional combustion chamber heating. For direct injection engines, cold start devices are only required in frosty weather.

An initial delay, known as injection lag, occurs in the high pressure fuel lines of rotary pumped systems. This occurs between the start of the pressure rise and the point when pressure is sufficient to overcome the compression spring force in the injectors.

Ignition of the fuel occurs in the combustion chamber at the time of injection into the heated air charge. The injection point and the ignition timing are therefore

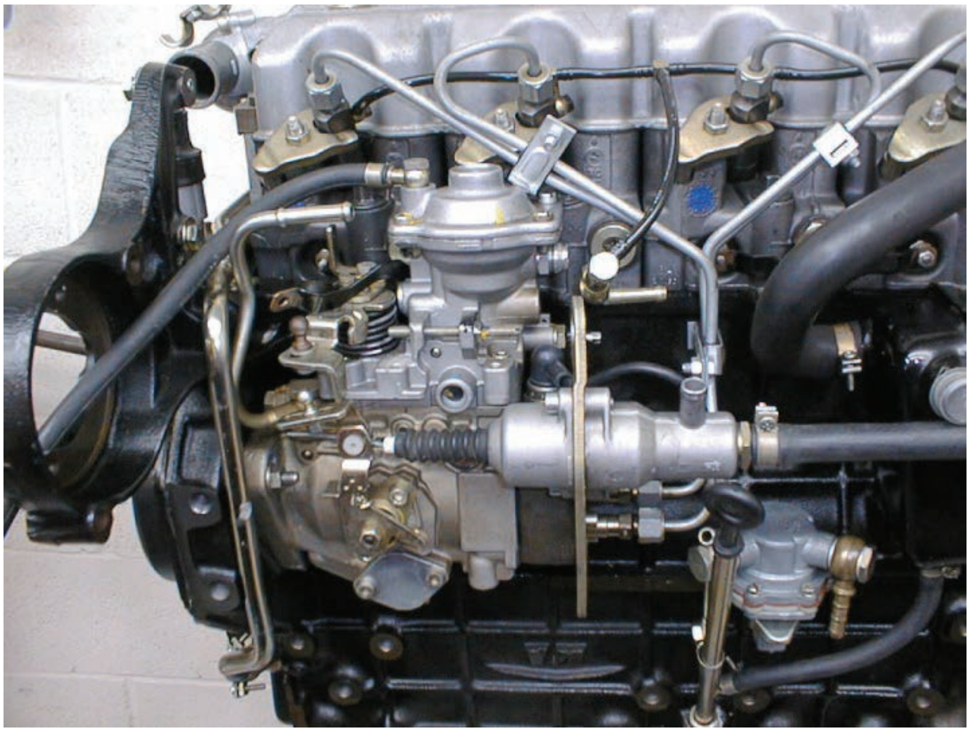


Figure 9.52 Pipes and injectors

Figure 9.51 Glow plugs (Source: Bosch Media)

Key fact

The delay phase or ignition lag for diesel fuel combustion lasts a few milliseconds.

Key fact

An initial delay, known as injection lag, occurs in the high pressure fuel lines of rotary pumped systems.

effectively the same thing. Diesel engine injection timing is equivalent to the ignition timing for petrol engines. Injection timing must fall within a narrow angle of crankshaft rotation. It is advanced and retarded for engine speed and load conditions. Injection timing is set by accurate positioning of the fuel injection pump. Incorrect timing leads to power loss. An increase in the production of nitrogen oxides (NOx) when too far advanced or an increase in the hydrocarbon (HC) emissions, when too far retarded also occurs.

Particulate emissions result from incomplete combustion of fuel. Particulates are seen as black carbon smoke in the exhaust under heavy load or when fuel delivery and/or timing is incorrect. White smoke may also be visible at other times, such as when the injection pump timing is incorrect. It also occurs when compression pressures are low or when coolant has leaked into the combustion chambers.

Key fact

Diesel engines are particularly suitable for turbocharging.

Recent developments in electronic diesel fuel injection control have made it possible to produce small direct injection engines. Diesel engines are built to withstand the internal stresses, which are greater than other engines. Diesel engines are particularly suitable for turbocharging, which improves power and torque outputs.

EGR is used to reduce nitrogen oxide (NOx) emissions on petrol engines; this is also true for diesel engines (Figure 9.54). Additionally, a small quantity of hot exhaust gas in the air charge of a cold engine helps to reduce the delay period and the incidence of cold engine diesel knock.

Many diesel engined vehicles are now fitted with oxidation catalytic converters that work in conjunction with other emission components to reduce hydrocarbon and particulate emissions.

The fuel systems for traditional direct and indirect injection are similar and vary only in injection pressures and injector types. Until more recently, all light high speed diesel engines used rotary diesel fuel injection pumps. These pumps producing injection pressures of over 100 bar for indirect engines. However,

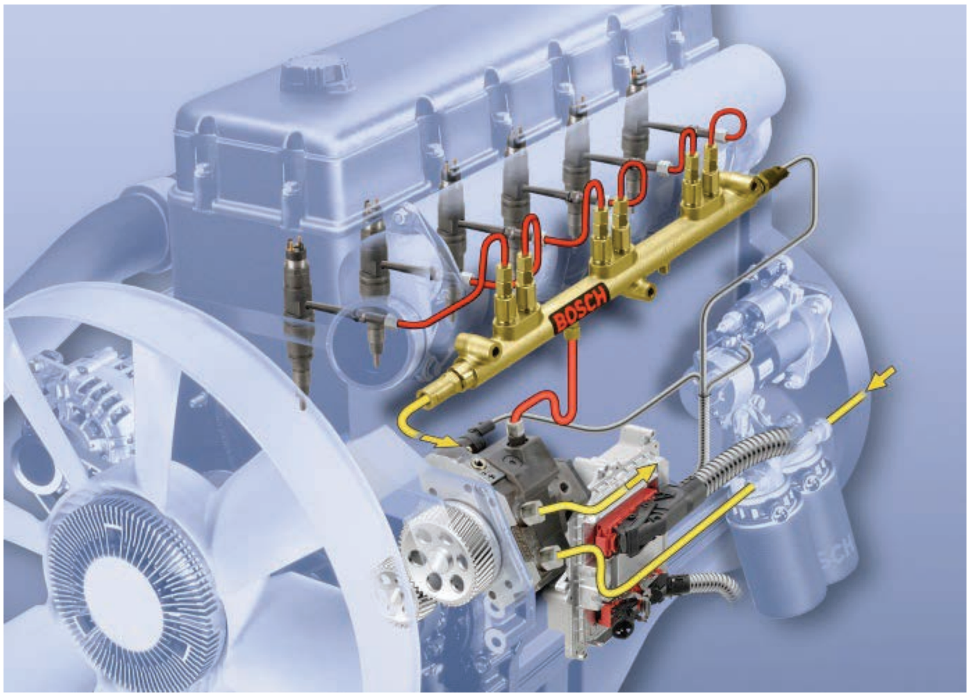


Figure 9.53 Common rail injection (Source: Bosch Media)

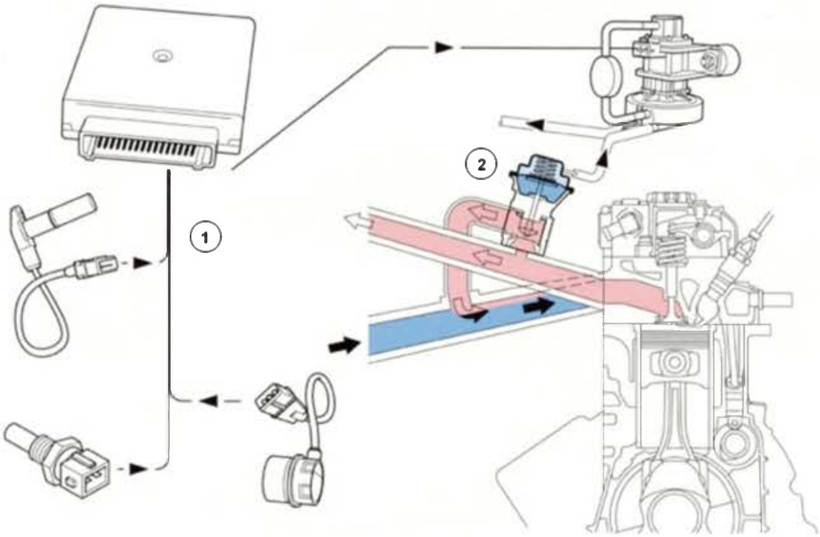


Figure 9.54 Diesel exhaust gas recirculation (EGR): 1-ECU and sensors, 2-EGR actuator and control valve

these can rise up to 1000 bar at the pump outlet, for turbocharged direct injection engines.

Injectors operate with a pulsing action at high pressure to break the fuel down into finely atomised parts. Atomisation is critical to good fuel distribution in the compressed air charge. The air charge pressure may be in excess of 60 bar. The pressure differential, between the fuel injection pressure and air charge pressure, must be sufficient to overcome the resistance during injection. This will also give good fuel atomisation and a shorter injection time.

The main components of a diesel fuel system provide for either the low pressure or the high pressure functions. The low pressure components are the fuel tank, the fuel feed and return pipes and hoses, a renewable fuel filter with a water trap and drain tap, and a priming or lift pump.

Safety first

Caution: pressure can be up to 1000 bar at the outlet of a rotary pump.

Key fact

Injectors operate with a pulsing action at high pressure to break the fuel down into finely atomised parts.

Figure 9.55 Bosch rotary injection pump (Source: Bosch Media)

Figure 9.56 Fuel filter

The high pressure components are the fuel injector pump, the high pressure pipes and the injectors (Figure 9.55). Other components provide for cold engine starting. Electronically controlled systems include sensors, an electronic diesel control (EDC) module and actuators in the injection pump.

Key fact

The internal components of the pump and injectors are manufactured to very fine tolerances.

All diesel fuel entering the injection pump and injectors must be fully filtered. The internal components of the pump and injectors are manufactured to very fine tolerances. Even very small particles of dirt could be damaging to these components.

The most common rotary injection pumps are axial-piston designs having a roller ring and cam plate attached to an axial piston or plunger in the distributor head to generate the high pressure. The latest versions have full electronic control. The details are examined further in the next section.

The high pressure pipes are of double thickness steel construction and are all of the same length. This is so that the internal pressure rise characteristics are identical for all cylinders. The high pressure connections are made by rolled flanges on the pipe ends and threaded unions securing the rolled flanges to convex, or occasionally concave, seats in the delivery valves and injectors.

Key fact

The fuel pressure needed is in excess of 100 bar (1500 psi).

The fuel injectors are fitted into the cylinder head with the nozzle tip projecting into the pre-combustion (IDI) or combustion chamber (DI). The injectors for indirect combustion are of a pintle or ‘pintaux’ design (similar to petrol injectors in many ways) and produce a conical spray pattern on injection. The injectors for direct injection (DI) are of a pencil type multi-hole design that produces a broad distribution of fuel on injection.

Fuel injectors are held closed by a compression spring. They are opened by hydraulic pressure when it is sufficient to overcome the spring force on the injector needle. The hydraulic pressure is applied to a face on the needle where it sits in a pressure chamber. The fuel pressure needed is in excess of 100 bar (1500 psi). This pressure lifts the needle and opens the nozzle, so that fuel is injected in a fine spray pattern into the combustion chamber (Figure 9.57).

Figure 9.57 DI injector

9.6.2 Injection overview

The basic principle of the four-stroke diesel engine is very similar to the petrol system. The main difference is that the mixture formation takes place in the cylinder combustion chamber as the fuel is injected under very high pressure.

The timing and quantity of the fuel injected is important from the usual viewpoints of performance, economy and emissions.

Fuel is metered into the combustion chamber by way of a high pressure pump connected to injectors via heavy duty pipes. When the fuel is injected it mixes with the air in the cylinder and will self-ignite at about 800 °C. See the section on diesel combustion for further details. The mixture formation in the cylinder is influenced by the following factors.

Start of delivery and start of injection (timing)

The timing of a diesel fuel injection pump to an engine is usually done using start of delivery as the reference mark. The actual start of injection, in other words when fuel starts to leave the injector, is slightly later than start of delivery, as this is influenced by the compression ratio of the engine, the compressibility of the fuel and the length of the delivery pipes. This timing increases the production of carbon particles (soot) if too early, and increases the hydrocarbon emissions if too late.

Spray duration and rate of discharge (fuel quantity)

The duration of the injection is expressed in degrees of crankshaft rotation in milliseconds. This clearly influences fuel quantity but the rate of discharge is also important. This rate is not constant due to the mechanical characteristics of the injection pump.

Injection pressure

Pressure of injection will affect the quantity of fuel, but the most important issue here is the effect on atomization. At higher pressures, the fuel will atomize into smaller droplets with a corresponding improvement in the burn quality. Indirect injection systems use pressures up to about 350 bar, while direct injection systems can be up to about 1000 bar.

Emissions of soot are greatly reduced by higher pressure injection.

Injection direction and number of jets

The direction of injection must match very closely the swirl and combustion chamber design. Deviations of only 2° from the ideal can greatly increase particulate emissions.

Excess air factor (air–fuel ratio)

Diesel engines do not, in general, use a throttle butterfly as the throttle acts directly on the injection pump to control fuel quantity. At low speeds in particular, the very high excess air factor ensures complete burning and very low emissions. Diesel engines operate where possible with an excess air factor even at high speeds.

9.6.3 Diesel exhaust emissions

Overall, the emissions from diesel combustion are far lower than emissions from petrol combustion. Figure 9.58 shows a general comparison between petrol and diesel emissions. The CO, HC and NOx emissions are lower, mainly due to the higher compression ratio and excess air factor. The higher compression ratio improves the thermal efficiency and thus lowers the fuel consumption. The excess air factor ensures more complete burning of the fuel.

The main problem area is that of particulate emissions (PMs). These particle chains of carbon molecules can also contain hydrocarbons, mostly aldehydes. The effect of this emission is a pollution problem but the possible carcinogenic effect of this

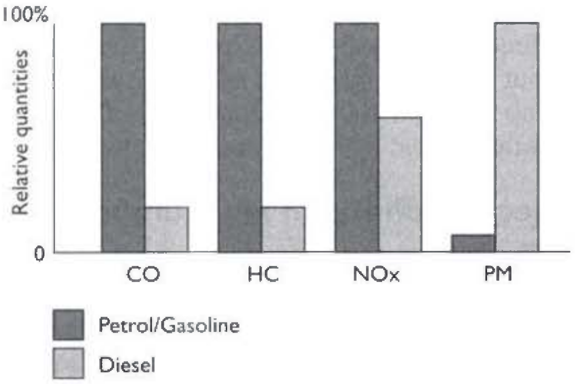


Figure 9.58 Comparison between petrol and diesel emissions

soot also gives a cause for concern. The diameter of these particles is only a few ten thousandths of a millimetre – consequently they float in the air and can be inhaled. However, many systems are now fitted with PM traps and filters.

9.6.4 Electronic control of diesel injection

The advent of electronic control over the diesel injection pump has allowed many advances over the purely mechanical system. The production of high pressure and injection is, however, still mechanical with all current systems. The following advantages are apparent over the non-electronic control system.

- More precise control of fuel quantity injected.
- Better control of start of injection.
- Idle speed control.
- Control of exhaust gas recirculation.
- Drive by wire system (potentiometer on throttle pedal).
- An anti-surge function.
- Output to data acquisition systems etc.
- Temperature compensation.
- Cruise control.

A distributor-type injection pump can be used with electronic control. Because fuel must be injected at high pressure, the hydraulic head, pressure pump and drive elements are still used. An electromagnetic moving iron actuator adjusts the position of the control collar, which in turn controls the delivery stroke and therefore the injected quantity of fuel. Fuel pressure is applied to a roller ring and this controls the start of injection. A solenoid-operated valve controls the supply to the roller ring. These actuators together allow control of the start of injection and injection quantity.

Figure 9.59 shows a block diagram of a typical electronic diesel control system. Ideal values for fuel quantity and timing are stored in memory maps in the electronic control unit. The injected fuel quantity is calculated from the accelerator position and the engine speed. The start of injection is determined from the following:

- Fuel quantity.
- Engine speed.
- Engine temperature.
- Air pressure.

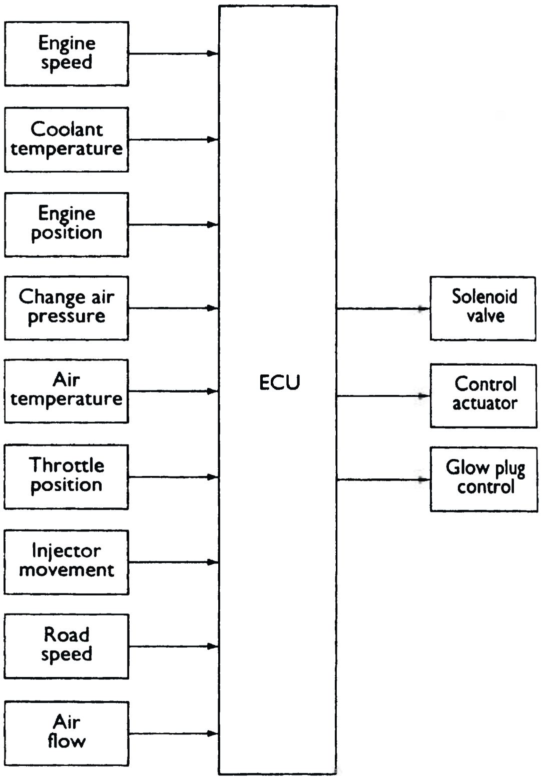


Figure 9.59 Block diagram of an electronic diesel control (EDC) system

The ECU is able to compare start of injection with actual delivery from a signal produced by the needle motion sensor in the injector.

Control of exhaust gas recirculation is by a simple solenoid valve. This is controlled as a function of engine speed, temperature and injected quantity. The ECU is also in control of the stop solenoid and glow plugs via a suitable relay. Figure 9.60 is the complete layout of an electronic diesel control system.

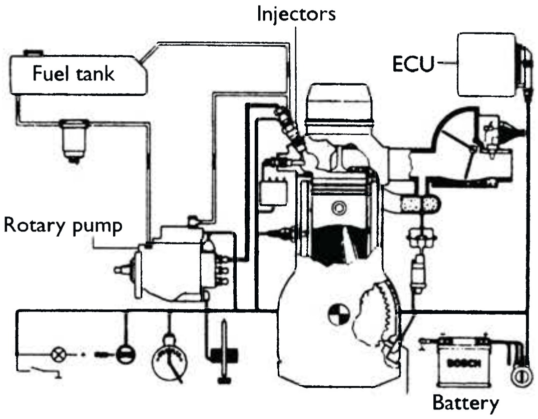


Figure 9.60 Layout of an electrical diesel control system

9.6.5 Rotary Pump System

The Bosch rotary VR pumps are used on high-speed direct injection diesel engines for cars and light commercial vehicles. They are radial-piston distributor injection pumps having opposing plungers that are forced inwards by cam lobes

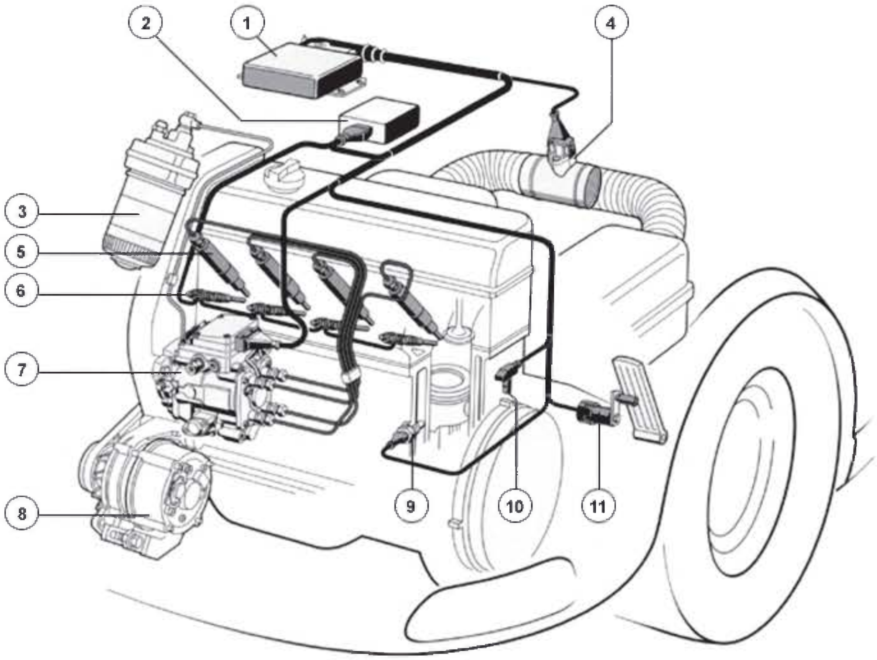


Figure 9.61 Electronic diesel control (EDC) rotary pump system 1, Engine ECU; 2, Glow control unit; 3. Filter; 4, Air-mass meter; 5, Injector; 6, Glowplugs; 7, Rotary distributor pump with ECU; 8, Alternator; 9, Coolant temperature sensor; 10, Crenkshaft sensor; 11, Throttle pedal sensor

on the inside of a cam ring, in order to produce high pressure, which can be up to 1400 bar in some applications. The cam is located in the pump body and the plungers are in the rotor driven by the pump spindle. Four cylinder engines have two plungers and four cam lobes. Six cylinder engines have three plungers and six cam lobes. The pump is driven from the engine at half crankshaft speed.

A low pressure feed to the injection pump is provided by a submerged electrical pump in the fuel tank (Figure 9.62). This provides for priming and positive pressure in the injection pump. In common with all diesel fuel systems, a fuel filter and water trap is used to ensure that only very clean fuel is delivered to the

Safety first

Injection pressures even on rotary systems can be up to 1400 bar.

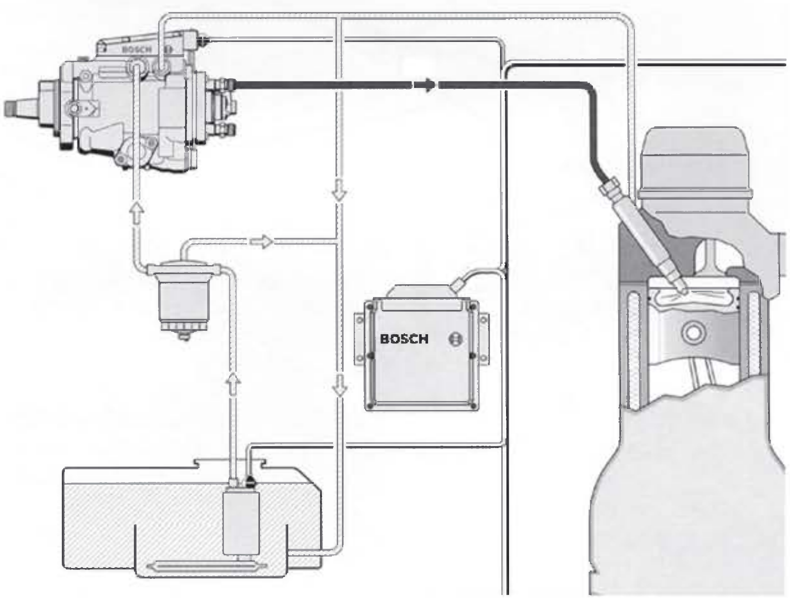


Figure 9.62 Low (in white) and high (in black) pressure fuel system

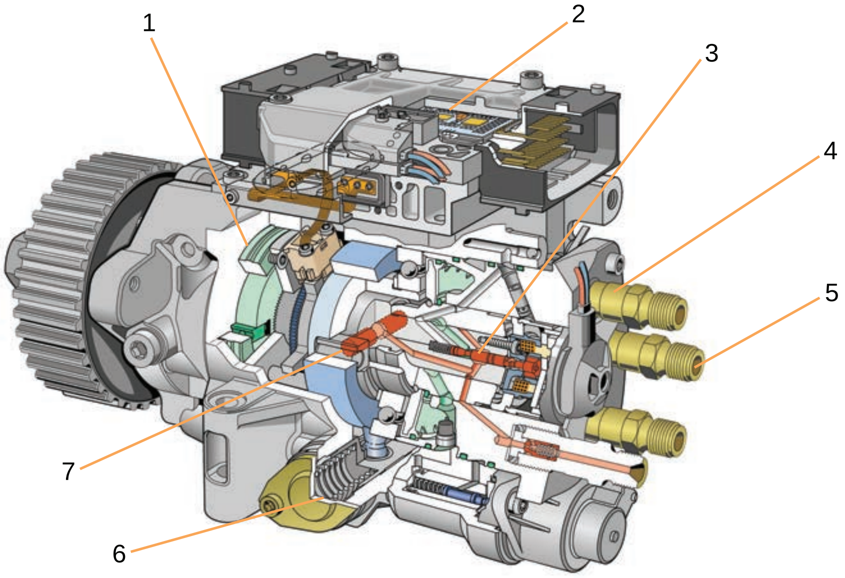


Figure 9.63 Solenoid-valve controlled radial-piston distributor pump: 1-Sensor (position/timing), 2-ECU, 3-High pressure solenoid valve needle, 4-Solenoid, 5-Outlets to injectors, 6-Timing device (ignition advance mechanism), 7-Radial-piston high pressure pump (Source: Bosch Media)

pump. Return pipes are used for excess fuel leakage, for purging the pump and for lubrication of the injectors.

Inside the distributor pump is a vane type pump, which is used to produce the pump body pressure. Pump body pressure is used for charging the high-pressure chamber between the plungers and for injection advance. A pressure control valve is used to prevent excessive pressure. It is a spring loaded plunger that is lifted by hydraulic pressure to expose ports in the valve bore. This will then allow fuel to flow back to the inlet side of the vane type pump.

An overflow throttle valve, in the pump housing, is used to allow a defined quantity of fuel to flow back to the fuel tank at all times. This provides some cooling in the pump and venting of air during pump priming. A second larger overflow bore in the valve opens at a given pressure to allow a flow of fuel from the distributor head.

The Bosch pump shown as Figure 9.63, has full electronic control for fuel metering and for injection advance. The electronic diesel control unit consists of two electronic control modules to perform the control functions. These two modules are the engine control ECU (Figure 9.64) and the injection-pump ECU. The pump ECU is fitted on top of the pump.

Fuel metering is controlled by the high-pressure solenoid valve. This is an electrically actuated valve set centrally inside the distributor rotor. There are connecting bores in the distributor rotor for filling of the high-pressure circuit, through the inlet port at pump body pressure, and for delivery at high pressure to the fuel injectors. These are either connected or separated by the position of the valve.

A high-pressure solenoid valve is closed by an electrical signal from the pump electronic control unit. When the valve is closed, fuel under high-pressure passes from the high-pressure pump chamber, through the bores in the rotor and distributor head, the return-flow throw throttle valve (delivery valve) and out to the injectors. It is then injected into the engine combustion chambers. The few microseconds of time, during which the valve remains closed, is referred to as the delivery or injection period.

Figure 9.64 EDC ECU

Key fact

Fuel metering is controlled by the ECU which operates a high-pressure solenoid valve.

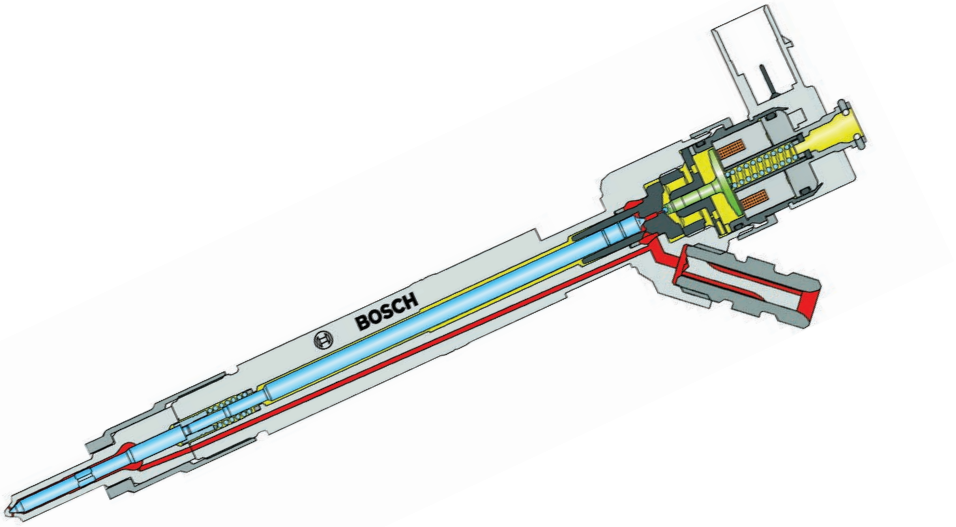


Figure 9.65 Injector (Source: Bosch Media)

The quantity of fuel that is metered for injection at any time is computed by the engine ECU, which sends signals to the injection-pump ECU for control of the high-pressure solenoid valve. The electrical current for operating this valve is high and the two electronic control units are separated, in order to avoid high current interference, in the more electronically vulnerable engine ECU.

The electronic diesel control units are provided with data signals from sensors and switches attached to the engine, the pump and other vehicle systems. The sensors are used for comparisons to programmed operating parameters and for calculations for metering the amount of fuel delivered and for controlling the injection advance.

Injection advance is obtained by rotation of the cam ring by pump body pressure in the injection advance mechanism. The injection advance mechanism consists of a transverse timing device piston and control components and an electrical solenoid valve. Maximum advance is 40° of crankshaft rotation.

A needle motion sensor in the injector sends a signal to the engine ECU at the instant of opening of the injector (Figure 9.65). This point, relative to the crankshaft rotational angle before top dead centre, is used for load and speed injection timing calculations and for control of the exhaust gas recirculation valve.

The Bosch, VR electronic diesel control system, uses a number of sensors and control actuators. This allows it to achieve optimum performance. However, even this sophisticated system has virtually been superseded by the common rail injection.

9.6.6 Common rail system

The development of diesel fuel systems is continuing, with many new electronic changes to the control and injection processes. One of the most significant is the CR 'common rail' system, which operates at very high injection pressures. It also has piloted and phased injection to reduce noise and vibration.

The common rail system has made it easier for small high speed diesel engines to have all the advantages of direct injection. These developments have resulted in significant improvements in fuel consumption and performance.

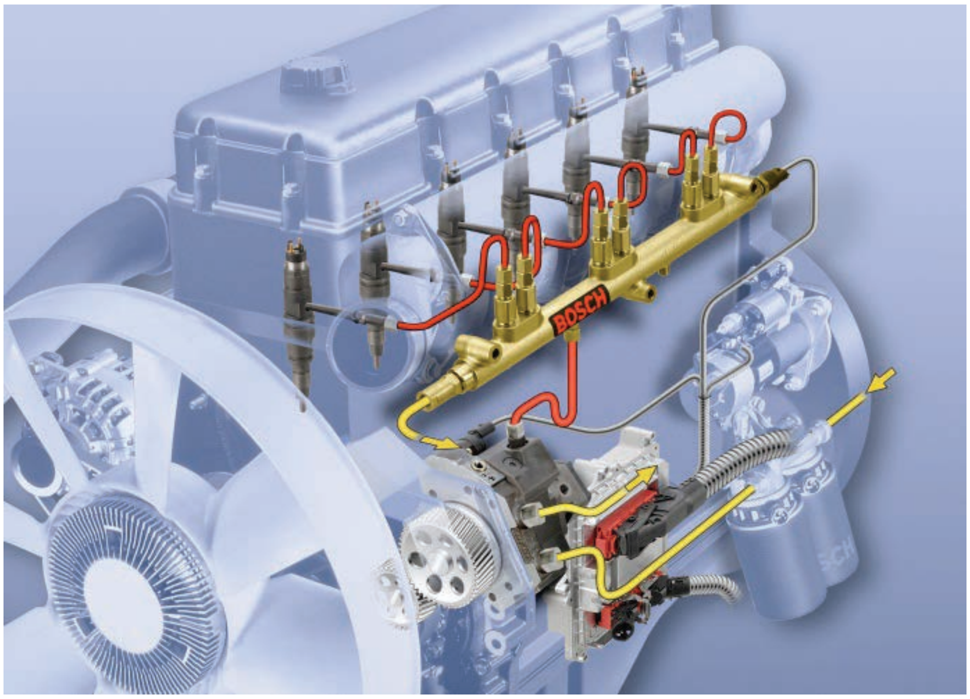


Figure 9.66 Common rail system (Source: Bosch Media)

The combustion process, with common rail injection, is improved by a pilot injection of a very small quantity of fuel, at between 40° and 90 ° BTDC (Figure 9.68). This pilot fuel ignites in the compressing air charge so that the cylinder temperature and pressure are higher than in a conventional diesel injection engine at the start of injection. The higher temperature and pressure reduces ignition lag to a minimum, so that the controlled combustion phase during the main injection period, is softer and more efficient.

Key fact

The combustion process, with common rail injection, is improved by a pilot injection of a very small quantity of fuel, at between 40° and 90 ° BTDC.

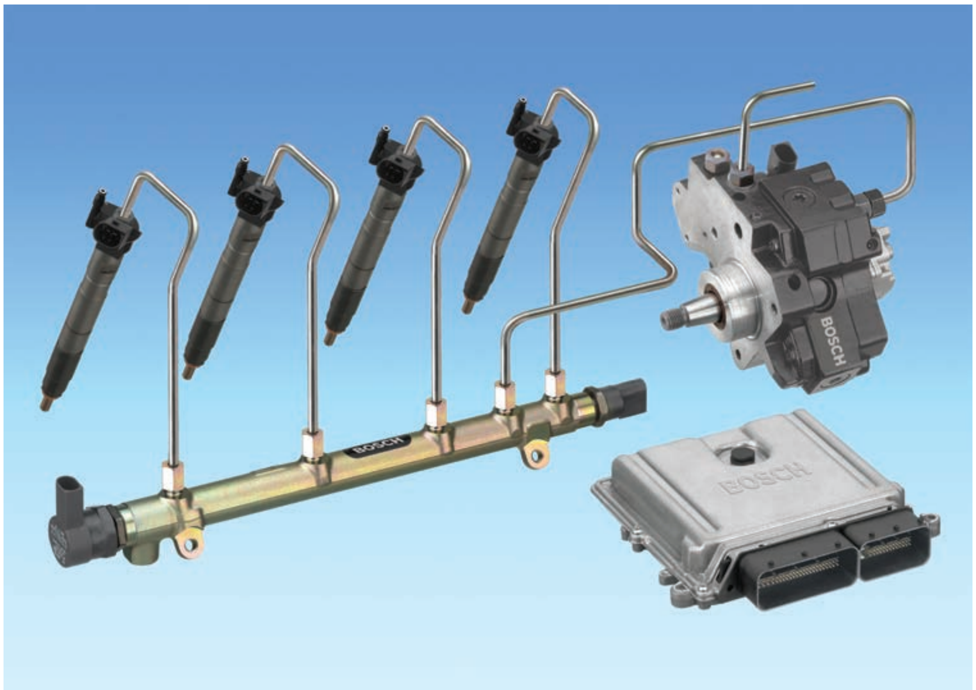
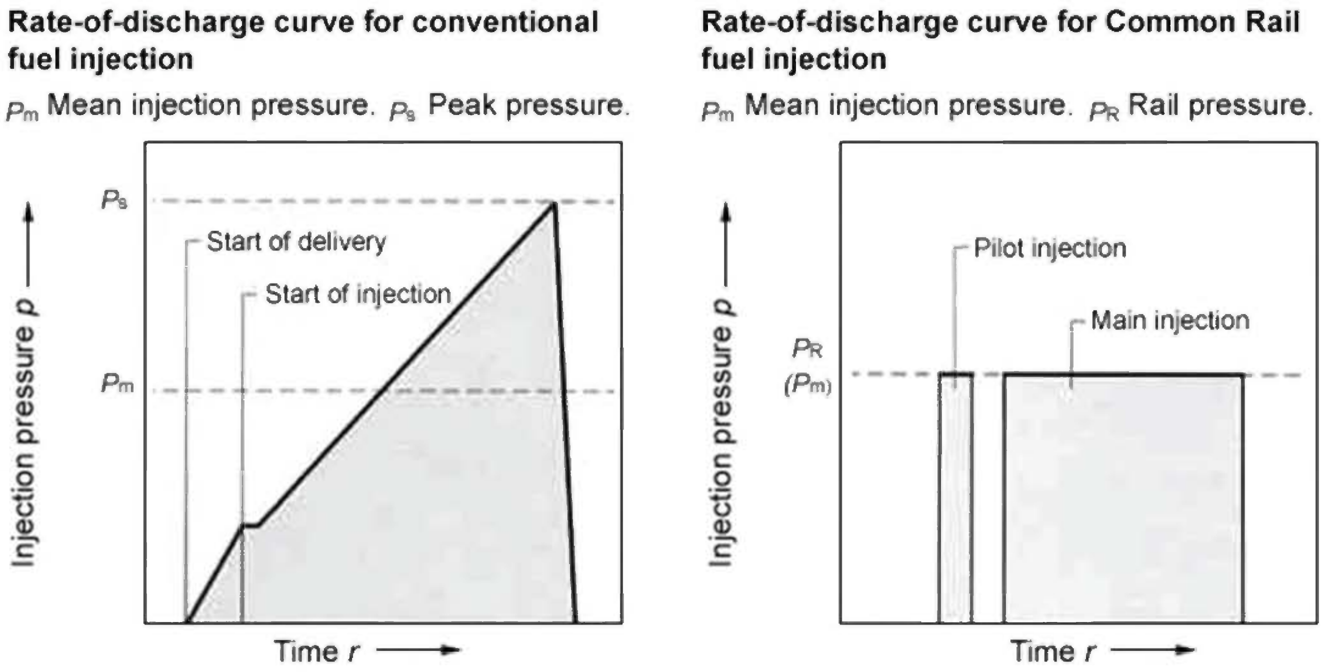


Figure 9.67 Four cylinder system (Source: Bosch Media)



Rate-of-discharge curve for conventional fuel injection

P_m Mean injection pressure. P_s Peak pressure.

Rate-of-discharge curve for Common Rail fuel injection

P_m Mean injection pressure. P_R Rail pressure.

Figure 9.68 Conventional system and common rail system

Key fact

CR fuel injection pressures can be varied, throughout the engine speed and load range.

Fuel injection pressures are varied, throughout the engine speed and load range, to suit the instantaneous conditions of driver demand and engine speed and load conditions. Data input, from other vehicle system ECUs, is used to further adapt the engine output, to suit changing conditions elsewhere on the vehicle. Examples are traction control, cruise control and automatic transmission gearshifts. The electronic diesel control (EDC) module carries out calculations to determine the quantity of fuel delivered. It also determines the injection timing based on engine speed and load conditions. The actuation of the injectors, at a specific crankshaft angle (injection advance), and for a specific duration (fuel quantity), is made by signal currents from the EDC module. A further function of the EDC module is to control the accumulator (rail) pressure.

Definition

Accumulator: an apparatus by means of which energy can be stored (the rail on a CR system).

The Bosch CR common rail diesel fuel injection system, for light vehicles, consists of four main component areas. These are the low pressure delivery, high pressure delivery with a high pressure pump and accumulator (rail), the electronically controlled injectors and electronic control unit and associated sensors and switches.

The low pressure delivery components are the fuel tank, a pre-filter, pre-supply (low pressure) pump, a fuel filter and the low pressure delivery pipes to the high pressure pump and for excess fuel return. The low pressure pump, depending on application, can be of the roller cell type and be fitted in either the fuel tank or in-line where it is mounted to the vehicle body close to the fuel tank. Where the pump is fitted in the fuel tank, it includes a pre-filter and has the fuel gauge sender unit attached to the same attachment flange on the side or top of the fuel tank.

Key fact

An inertia switch is generally used to cut the electrical current to the pump motor in an accident.

The electrical supply to the fuel pump is made, either directly, or through a relay from the electronic diesel control module. An inertia switch is generally used to cut the electrical current to the pump motor in an accident. On some vehicles, a gear type pump may be incorporated into the high pressure pump and be driven

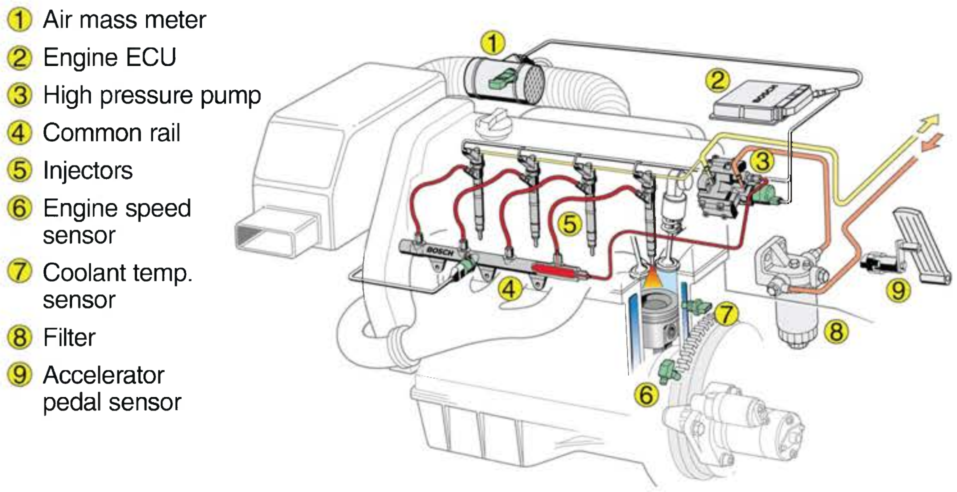


Figure 9.69 Main components

from a common drive shaft. It can be a separate pump attached to the engine with a geared drive from the camshaft or crankshaft. The low pressure delivery pipes connect to a fuel filter and water trap. A continuous flow of fuel runs through the filter and primes the high pressure pump or returns to the fuel tank.

The high pressure pump is driven from the engine crankshaft through a geared drive at half engine speed and can be fitted where a conventional distributor pump would be. It can also be fitted on the end of the camshaft housing and be driven by the camshaft. It is lubricated by the diesel fuel that flows through it.

The pump has to produce all of the high pressure for fuel injection. It is a triple piston radial pump, with a central cam for operation of the pressure direction of the pistons and return springs to maintain the piston rubbing shoes in contact with the cam. The pump has a positive displacement with inlet and outlet valves controlling the direction of flow through the pump (Figure 9.70).

Figure 9.70 High pressure pump (Source: Bosch Media)

Safety first

A CR pump can produce pressures in the region of 1400 bar.

The pump delivery rate is proportional to the speed of rotation of the engine so that it meets most engine speed requirements. To meet the engine load requirements the pump has a high volume. To meet the high pressure requirements, for fine atomisation of the fuel on injection, the pump can produce pressures in the region of 1400 bar (scary!). A pressure control valve returns excess fuel to the fuel tank.

The pressure control valve is a mechanical and electrical unit. It is fitted on the pump or the high pressure accumulator (rail). The mechanical part of the valve consists of a compression spring that acts on a plunger and ball valve. The electrical component is a solenoid that puts additional and variable force to the ball valve. The solenoid is actuated on signal currents from the EDC module. When the solenoid is not actuated, the ball valve opens at 100 bar against the resistance of the compression spring. This spring valve damps some of the high frequency pressure fluctuations produced by the pump.

Key fact

The solenoid in the pressure control valve is used for setting pressure in the rail.

The solenoid in the pressure control valve is used for setting a variable mean pressure in the high pressure accumulator (rail). The pressure in the rail is measured by a sensor and compared with a stored map in the EDC module for the current engine operating conditions. In order to increase the fuel rail pressure an electrical alternating current is applied to the solenoid. The energising current is varied by the EDC module, so that the additional force on the ball valve produces the required fuel rail pressure.

The high pressure accumulator (rail) is common to all cylinders and derives its name ‘common rail’ from this. This term is used in preference to fuel rail, which is used for petrol engines. The rail is an accumulator because it holds a large volume of fuel under pressure. The volume of fuel is sufficient to dampen the pressure pulses from the high pressure pump.

The injectors on the common rail system have nozzles that are similar to all other diesel injectors for direct injection engines. The nozzle needle seats in the nozzle

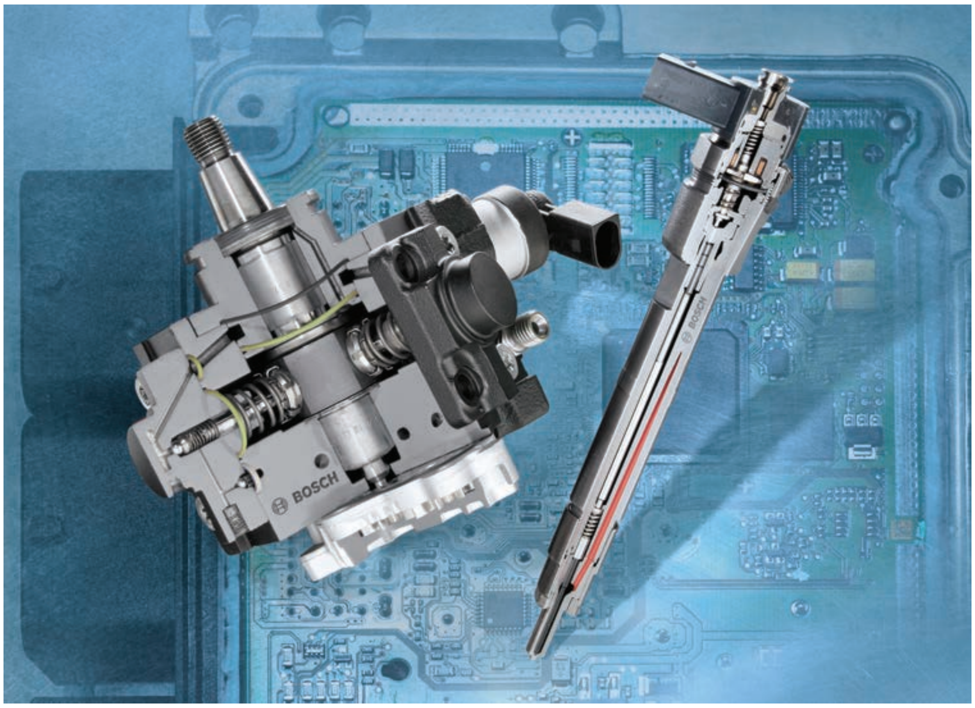


Figure 9.71 Sectioned CR pump and injector (Source: Bosch Media)

Figure 9.72 Piezo CR injector

to obstruct the holes in the tip where the fuel is injected into the combustion chamber. The nozzle needle is held closed by a compression spring and opened by hydraulic pressure (Figures 9.71 and 9.72).

Opening and closing of the injector is controlled, not by high pressure fuel pulse from an injector pump, as in a conventional rotary distributor pump, but by actuation of an electrical solenoid in the injector body. This is controlled by the electronic diesel control module. A permanent high pressure is maintained in the injector at the same pressure as the rail. Operation of the injector is controllable for very small intervals of time.

The electronic control of the common rail diesel injection system allows for precise control of fuelling. This results in excellent economy and very low emissions.

Key fact

The piezo injector reacts very quickly therefore improving fuel control.

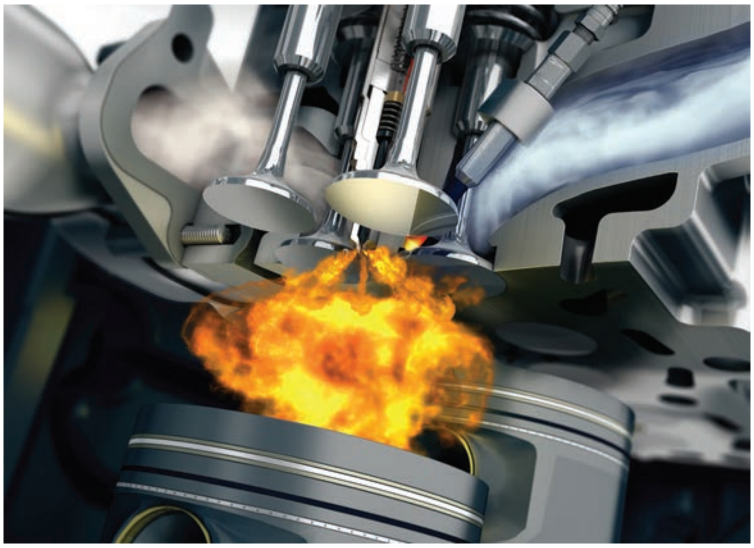


Figure 9.73 Common rail injection combustion

9.6.7 Electronic unit injection (EUI) – diesel fuel

The advantages of electronic unit injection are as follows.

Lower emissions

Through the use of higher injection pressures (up to 2000 bar), lower emissions of particulates and NOx are achieved, together with a reduction in the levels of noise traditionally associated with diesel engines.

Electronic fuel quantity and timing control

Precise electronic control also assists in the reduction of emissions.

Shot to shot fuel adjustment

This feature also provides a very quick transient response, improving vehicle drivability.

Control of all engine functions

Through a series of sensors connected to the electronic control unit (ECU), the EUI system ensures that all the engine functions consistently operate at optimum performance.

Electronically controlled pilot injection

A new feature developed to meet tighter NOx emissions standards, without loss of fuel consumption. Pilot injection also reduces combustion noise.

Communication with other systems

Linked to the ECU, the EUI system can communicate with other vehicle systems such as ABS, transmission and steering, making further systems development possible.

Cylinder cut-out

This is used as a diagnostic aid and offers potential for fuel economy at idling and low loads.

Reliability and durability

The EUI's reliability is proven under field conditions. Experience in the truck market indicates a service life of at least 800 000 km.

Full diagnostics capability

Fault codes can be stored and diagnostic equipment connected.

Further development potential

EUI technology is currently only at the beginning of its life cycle; it has significant further development potential which will enable the system to meet future tough emissions legislation.

In the EUI system, the fuel injection pump, the injector and a solenoid valve are combined in one, single unit; these unit injectors are located in the cylinder-head, above the combustion chamber. The EUI is driven by a rocker arm, which is in turn driven by the engine camshaft. This is the most efficient hydraulic and mechanical layout, giving the lowest parasitic losses. The fuel feed and spill pass through passages integrated in the cylinder-head.

The EUI uses sensors and an electronic control unit (ECU) to achieve precise injection timing and fuel quantities. Sensors located on the engine pass information to the ECU on all the relevant engine functions. This evaluates the information and compares it with optimum values stored in the ECU to decide on the exact injection timing and fuel quantity required to realize optimum performance. Signals are then sent to the unit injector's solenoid-actuated spill valve system to deliver fuel at the timing required to achieve this performance.

Injection is actuated by switching the integrated solenoid valve. The closing point of the valve marks the beginning of fuel delivery, and the duration of closing determines the fuel quantity. The operating principle is as follows.

Each plunger moves through a fixed stroke, actuated by the engine camshaft. On the upward (filling) stroke, fuel passes from the cylinder-head through a series of integrated passages and the open spill valve into a chamber below the plunger.

The ECU then sends a signal to the solenoid stator, which results in the closure of the spill control valve. The plunger continues its downward stroke causing pressure to build in the high pressure passages. At a pre-set pressure the nozzle opens and fuel injection begins. When the solenoid stator is de-energized the spill control valve opens, causing the pressure to collapse, which allows the nozzle to close, resulting in a very rapid termination of injection.

Electronic unit injectors (Figure 9.74) have been developed in a range of sizes to suit all engines, and can be fitted to light- and heavy-duty engines suitable for small cars and the largest premium trucks.

Figure 9.74 Unit injector (Source: Bosch Media)

9.6.8 Diesel lambda sensor

Lambda sensing is now also applicable to diesel engines. This new technology makes cars cleaner and more economical (Figure 9.75). Bosch is now also applying the lambda sensor in the closed loop control concept for diesel engines. The new system allows for a previously unreached fine tuning of injection and engine. This reduces fuel consumption and pollutant emission from diesel engines.

Different from the previous concept, the lambda-based control now optimizes the exhaust gas quality via exhaust gas recirculation, charge-air pressure and start of injection. These parameters decisively influence the emissions from diesel engines. A broad-band lambda sensor, with a wide working range, measures the oxygen content in the exhaust gas and renders important information on the combustion processes in the engine, which can be utilized for the engine management.

Compared to the standard diesel engine management, the new Bosch system permits a stricter adherence to low emission values. Engines are better protected against defects. For example, the harmful combustion in cars running in overrun may be detected and corrected. In engines running under full load, the system offers more effective smoke suppression.

Key fact

Lambda sensing is applicable to diesel engines.

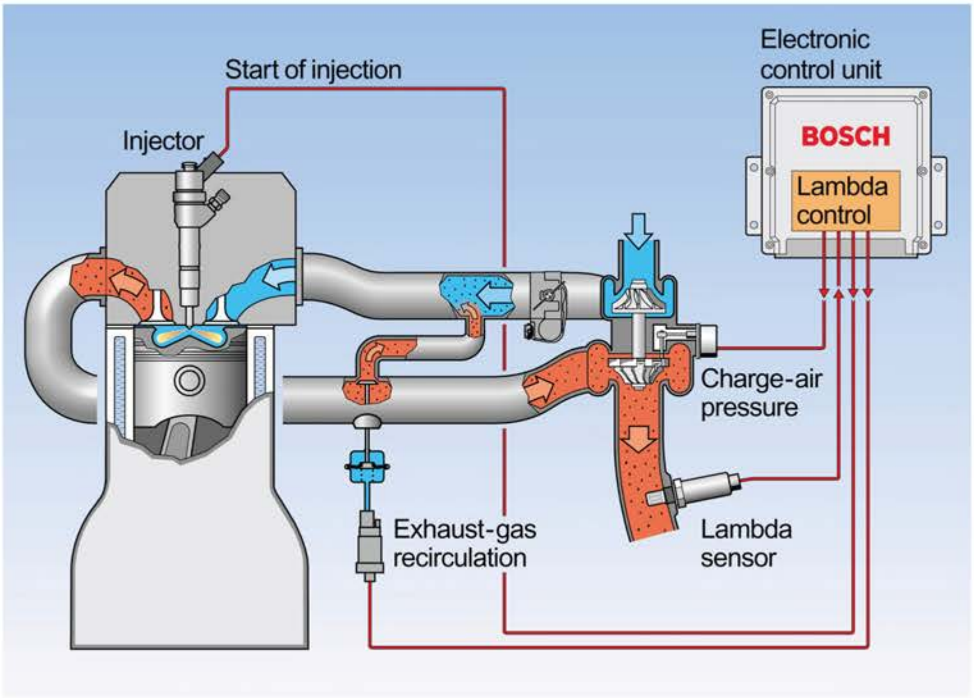


Figure 9.75 Lambda sensing on a diesel system (Source: Bosch Media)

The lambda sensor will also monitor the NOx accumulator catalytic converters (of future emission purification systems). The sensor supplies data for the management of the catalytic converter, which has to be cleaned at regular intervals in order to preserve its storage capability.

9.6.9 Exhaust emission treatments

9.6.9.1 Introduction

The main approach to the further lowering of diesel engine emissions, the focus is primarily on internal engine improvements; improved fuel combustion prevents as far as possible the formation of pollutants and also reduces fuel consumption. In this respect, automobile manufacturers and their component suppliers have already achieved a great deal.

However, heavy passenger cars will not meet the latest Euro standards without treatment systems. EDC (Electronic Diesel Control) handles the management of particulate filters and nitrogen oxide storage catalytic converters. It matches injection flexibly to the requirements of the exhaust emission treatment systems, for example by altering injection timing, quantity and process. EDC also matches the amount of combustion air fed to the engine to the respective demand. This is done by controlling the exhaust gas recirculation, determining the setting of the throttle valve and the operating pressure of the exhaust gas turbocharger. Sensors convey information to the EDC about the exhaust gas temperature, backpressure and composition. Engine management can, therefore, not only determine the condition of the particulate filter and the nitrogen oxide storage catalytic converter, but also improve the quality of combustion.

9.6.9.2 Diesel particulate filters

If the injection system and the particulate filter are working optimally together, exhaust emission values can be further improved. Bosch therefore began mass production of diesel particulate filters in 2005/6. The particulate filter from Bosch is made of sintered metal and lasts considerably longer than current ceramic models, since its special structure offers a high storage capacity for oil and additive combustion residues. The filter is designed in such a way that the filtered particulates are very evenly deposited, allowing the condition of the filter to be identified more reliably and its regeneration controlled far better than

Figure 9.76 Diesel exhaust particulate filter (Source: Bosch Media)

with other solutions. The diesel particulate filter is designed to last as long as the vehicle itself.

Once the storage capacity of the particulate filter has been exhausted, the filter has to be regenerated by passing hot exhaust gases through it which burn up the deposited particulates. In order to produce the necessary high exhaust gas temperatures, the EDC alters the amount of air fed to the engine as well as the amount of fuel injected and the timing of injection. In addition, some unburnt fuel can be fed to the oxidizing catalytic converter by arranging for extra fuel to be injected during the expansion stroke. The fuel combusts in the oxidizing catalytic converter and raises the exhaust temperature even further. Engineers are currently developing a system for injecting fuel directly into the exhaust duct, supplementing the injection into the combustion chamber just referred to.

People often express the hope that particulate filters could be fitted retrospectively to diesel-powered vehicles. Such retro-fitting would require an enormous technical input, since not only would the engine have to be adjusted to the modified exhaust system, but the control unit and the control unit software would also have to be extensively modified.

Key fact

Once the storage capacity of a particulate filter has been exhausted, the filter has to be regenerated by passing hot exhaust gases through it which burn up the deposited particulates.

9.6.9.3 Exhaust gas treatment for commercial vehicles

To meet the ever more stringent Euro parameters, two options are possible for commercial vehicles:

- Exhaust gas recycling, if necessary in combination with the use of a particulate filter.
- Selective Catalytic Reduction (SCR). SCR, in combination with a particulate filter, tended to be the favoured solution for Euro 5 (introduced in 2008).

Definition

SCR: Selective Catalytic Reduction: nitrogen oxide in the exhaust gases reacts with ammonia to produce water and nitrogen.

Bosch has developed the Denoxtronic dosage system for delivering the reducing agents for the SCR system. In the SCR process, the nitrogen oxide in the exhaust gases reacts with ammonia to produce water and nitrogen. The required ammonia is generated directly in the exhaust duct by hydrolysis from the added reducing agent AdBlue – a solution of water and urea. Bosch's Denoxtronic delivers to the catalytic converter the required amount of AdBlue dependent on the actual operating circumstances. It came into use for the first time in a series production vehicle in 2004.

Engine design using an SCR catalytic converter reduces the nitrogen oxide emissions of commercial vehicles by around 85%. This allows injection timing to be advanced, leading to a reduction in fuel consumption of up to 5%. If an oxidizing catalytic converter is used, particulate emissions can also be reduced by up to 30% - the use of an SCR catalytic converter means that it pays to protect the environment, since the extra cost of the exhaust gas treatment system is soon outweighed by the savings in fuel consumption (Dohle 2003).

9.7 Summary

9.7.1 Overview

The developments of fuel injection in general, and the reduced complexity of single-point systems in particular, have made the carburettor obsolete. As emission regulations continue to become more stringent, manufacturers are being forced into using fuel injection, even on lower priced models. This larger market will, in turn, pull the price of the systems down, making them comparable to carburation techniques on price but superior in performance.

9.7.2 Diagnosing fuel control systems

Safety first

Caution/Achtung/Attention – Burning fuel can seriously damage your health!

Table 9.5 lists some common symptoms of a fuel system malfunction together with suggestions for the possible fault. Note that when diagnosing engine fuel system faults, the same symptoms may indicate an ignition problem.

The following procedure is generic and, with a little adaptation, can be applied to any fuel injection system. Refer to manufacturer's recommendations if in any doubt. It is assumed the ignition system is operating correctly. Most tests are carried out while cranking the engine.

1. Check battery state of charge (at least 70%).
2. Hand and eye checks (all fuel and electrical connections secure and clean).
3. Check fuel pressure supplied to rail (in multipoint systems it will be about 2.5 bar but check specifications).

Table 9.5 Common fuel system symptoms and faults

Symptom	Possible fault
Engine rotates but does not start	No fuel in the tank! Air filter dirty or blocked Fuel pump not running No fuel being injected
Difficult to start when cold	Air filter dirty or blocked Fuel system wiring fault Enrichment device not working (choke or injection circuit)
Difficult to start when hot	Air filter dirty or blocked Fuel system wiring fault
Engine starts but then stops immediately	Fuel system contamination Fuel pump or circuit fault (relay) Intake system air leak
Erratic idle	Air filter blocked Inlet system air leak Incorrect CO setting Fuel injectors not spraying correctly
Misfire through all speeds	Fuel filter blocked Fuel pump delivery low Fuel tank ventilation system blocked
Engine stalls	Idle speed incorrect CO setting incorrect Fuel filter blocked Air filter blocked Intake air leak Idle control system not working
Lack of power	Fuel filter blocked Air filter blocked Low fuel pump delivery Fuel injectors blocked
Backfires	Fuel system fault (air flow sensor on some cars)

4. If the pressure is not correct jump to stage 10.
5. Is injector operation OK? – continue if not (suitable spray pattern or dwell reading across injector supply).
6. Check supply circuits from main relay (battery volts minimum).
7. Continuity of injector wiring (0–0.2 and note that many injectors are connected in parallel).

9.8 Advanced fuel control technology

9.8.1 Air–fuel ratio calculations

The ideal ratio by mass of air to fuel for complete combustion is 14.7:1. This is given the lambda value 1, which is known as stoichiometry. This figure can be calculated by working out the exact number of oxygen atoms, which are required to completely oxidise the particular number of hydrogen and carbon atoms in the hydrocarbon fuel, then multiplying by the atomic mass of the respective elements.

Petrol/gasoline consists of a number of ingredients, these are known as fractions and fall into three chemical series:

Paraffins	e.g. Octane	C_8H_{18}
Napthenes	e.g. Cyclohexane	C_6H_{12}
Aromatics	e.g. Benzene	C_6H_6

The ideal air fuel ratio for each of these can be calculated from the balanced chemical equation and the atomic mass of each atom. The atomic masses of interest are:

- Carbon (C) = 12
- Hydrogen (H) = 1
- Oxygen (O) = 16.

The balanced chemical equation for complete combustion of octane is as follows:

$$2C_8H_{18} + 25O_2 \rightarrow 16CO_2 + 18H_2O$$

The molecular mass of $2C_8H_{18}$ is:

$$(2 \times 12 \times 8) + (2 \times 1 \times 18) = 228$$

The molecular mass of $25 \times O_2$ is:

$$(25 \times 16 \times 2) = 800$$

Therefore the oxygen to octane ratio is 800:228 or 3.5:1; in other words 1 kg of fuel uses 3.5 kg of oxygen. Air contains 23% of oxygen by mass (21% by volume), which means 1 kg of air contains 0.23 kg of oxygen. Further, there is 1 kg of oxygen in 4.35 kg of air.

The ideal air to fuel ratio for complete combustion of octane is 3.5 x 4.35 = 15.2:1.

Octane:	$2C_8H_{18} + 25O_2 \rightarrow 16CO_2 + 18H_2O$	A/F ratio = 15.2:1

If a similar calculation is carried out for cyclohexane and benzene the results are as follows:

Cyclohexane:	$C_6H_{12} + 9O_2 \rightarrow 6CO_2 + 6H_2O$	A/F ratio = 14.7:1
Benzene:	$C_6H_6 + 15O_2 \rightarrow 6CO_2 + 3H_2O$	A/F ratio = 13.2:1

The above examples serve to explain how air to fuel ratio is calculated and how petrol/gasoline being a mixture of a number of fractions, has an ideal air fuel ratio

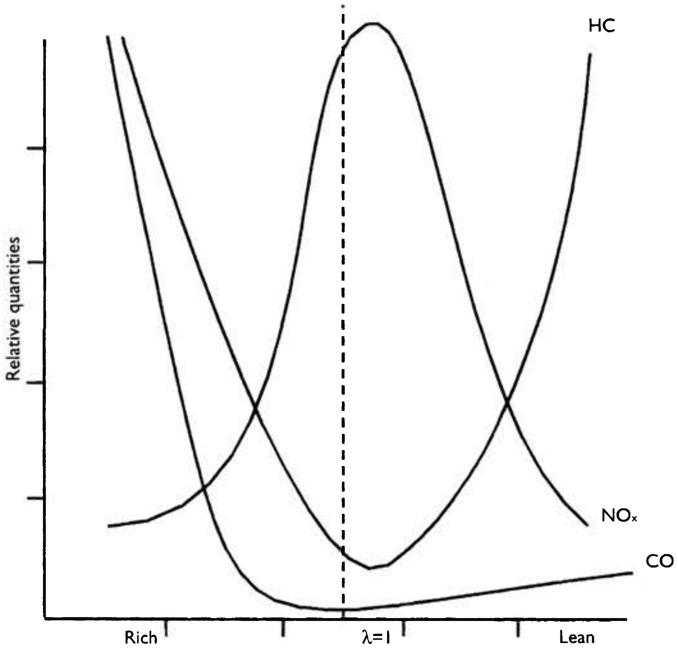


Figure 9.77 Influence of air-fuel ratio on the three main pollutants created from a spark ignition engine (No catalyst in use)

of 14.7:1.

This figure is, however, only the theoretical ideal and takes no account of pollutants produced and the effect air fuel ratio has on engine performance. With modern engine fuel control systems it is possible to set the air fuel ratio exactly at this stoichiometric ratio if desired. As usual though, a compromise must be sought as to the ideal setting.

Figure 9.77 shows the influence of air fuel ratio on the three main pollutants created from a spark ignition, internal combustion engine. A ratio slightly weaker than lambda value of 1 (or about 15.5 : 1 ratio) is often an appropriate compromise.

CHAPTER 10

Engine management

10.1 Combined ignition and fuel introduction

10.1.1 Introduction

As the requirements for lower and lower emissions continue, together with the need for better performance, other areas of engine control are constantly being investigated. This control is becoming even more important as the possibility of carbon dioxide emissions being included in future regulations increases. Some of the current and potential areas for further control of engine operation are included in this section. Although some of the common areas of 'control' have been covered in the previous two chapters, this chapter will cover some aspects in more detail and introduce further areas of engine control. Some of the main issues are:

- Ignition timing.
- Dwell angle.
- Fuel quantity.
- Exhaust gas recirculation (EGR).
- Canister purge.
- Idle speed.

An engine management system can be represented by the standard three-stage model as shown in Figure 10.1. This representation shows closed loop feedback, which is a common feature, particularly related to:

- lambda control,
- knock,
- idle speed.

The block diagram shown as Figure 10.2 can further represent an engine management system. This series of 'inputs' and 'outputs' is a good way of representing a complex system. This section continues with a look at some of the less common 'inputs and outputs'.

Key fact

As the requirements for lower and lower emissions continue, together with the need for better performance, all possible areas of engine control are being investigated.

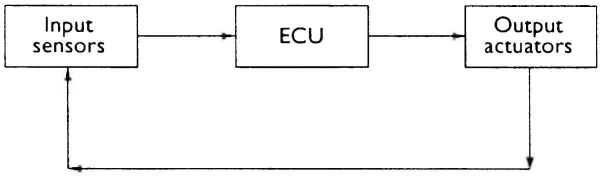


Figure 10.1 Representation of complete engine control as the standard

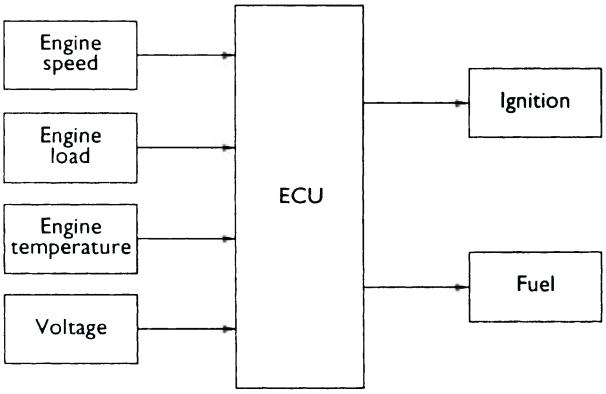


Figure 10.2 General block diagram of an ignition and fuel control system

10.1.2 Variable inlet tract

For an engine to operate at its best, volumetric efficiency is not possible with fixed manifolds. This is because the length of the inlet tract determines the velocity of the intake air and, in particular, the propagation of the pressure waves set up by the pumping action of the cylinders. These standing waves can be used to improve the ram effect of the charge as it enters the cylinder but only if they coincide with the opening of the inlet valves. The length of the inlet tract has an effect on the frequency of these waves. One method of changing the length of the inlet tract is shown in Figure 10.3. The control valves move, which changes the effective length of the inlet.

10.1.3 Combustion flame and pressure sensing

Research is on-going in the development of cost effective sensors for determining combustion pressure and combustion flame quality. These sensors are used during development but currently are prohibitively expensive for use in production. When available, these sensors will provide instantaneous closed loop feedback about the combustion process. This will be particularly important with lean burning engines.

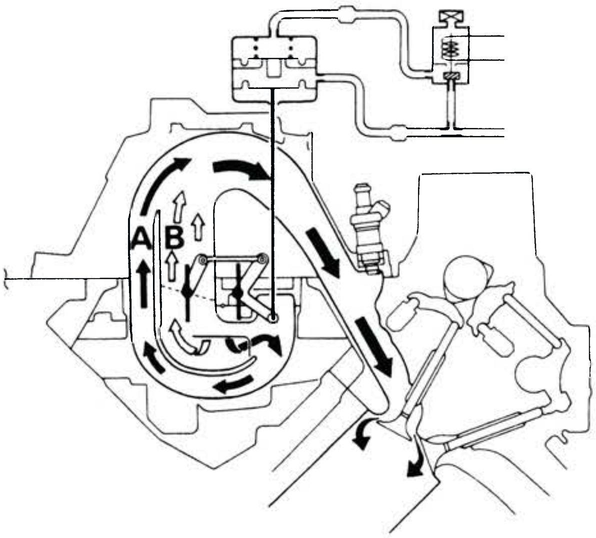


Figure 10.3 Variable length inlet manifold. A = long tract; B = short tract

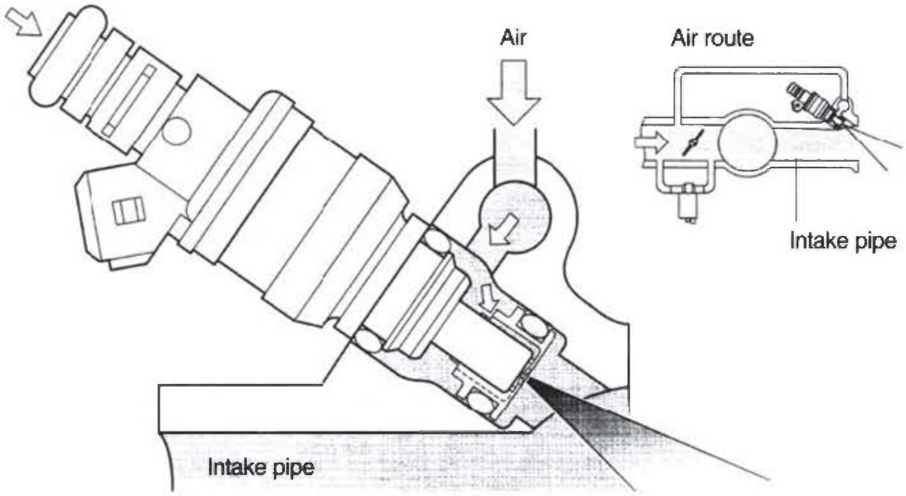


Figure 10.4 Injection valve with air shrouding

Figure 10.5 Improved injection pattern (also split in two in this case)

10.1.4 Wide range lambda sensors

Most lambda sensors provide excellent closed-control of the air–fuel ratio at or very near to stoichiometry (14.7 : 1). A sensor is now available that is able to provide a linear output between air–fuel ratios of 12 : 1 and about 24 : 1. This allows closed loop feedback over a much wider range of operating conditions.

10.1.5 Injectors with air shrouding

If high-speed air is introduced at the tip of an injector, the dispersal of the fuel is considerably improved. Droplet size can be reduced to below 50 μm during idle conditions. Figure 10.4 shows an injector with air shrouding. one without. The improved dispersal and droplet size is clear in Figure 10.5.

10.2 Exhaust emission control

10.2.1 Engine design

Many design details of an engine have a marked effect on the production of pollutant emissions. With this in mind, it will be clear that the final design of an engine is a compromise between conflicting interests. The major areas of interest are as discussed in the following sections.

10.2.2 Combustion chamber design

The main source of hydrocarbon emissions is unburnt fuel that is in contact with the combustion chamber walls. For this reason the surface area of the walls should be kept as small as possible and with the least complicated shape. A theoretical ideal is a sphere but this is far from practical. Good swirl of the cylinder charge is important, as this facilitates better and more rapid burning. Perhaps more important is to ensure a good swirl in the area of the spark plug. This ensures a mixture quality that is easier to ignite. The spark plug is best positioned in the centre of the combustion chamber as this reduces the likelihood of combustion knock by reducing the distance the flame front has to travel.

Key fact

The spark plug is best positioned in the centre of the combustion chamber as this reduces the likelihood of combustion knock.

10.2.3 Compression ratio

Key fact

The higher the compression ratio, the higher, in general, the thermal efficiency of the engine and therefore the better the performance and fuel consumption.

The higher the compression ratio, the higher, in general, the thermal efficiency of the engine and therefore the better the performance and fuel consumption. The two main drawbacks to higher compression ratios are the increased emissions and the increased tendency to knock. The problem with emissions is due to the high temperature, which in turn causes greater production of NOx. The increase in temperature makes the fuel and air mixture more likely to self-ignite, causing a higher risk of combustion knock. Countries which have had stringent emission regulations for some time, such as the USA and Japan, have tended to develop lower compression engines. However, with the changes in combustion chamber design and the more widespread introduction of four valves per cylinder, together with greater electronic control and other methods of dealing with emissions, compression ratios have increased over the years.

10.2.4 Valve timing

The effect of valve timing on exhaust emissions can be quite considerable. One of the main factors is the amount of valve overlap. This is the time during which the inlet valve has opened but the exhaust valve has not yet closed. The duration of this phase determines the amount of exhaust gas left in the cylinder when the exhaust valve finally closes. This has a significant effect on the reaction temperature (the more exhaust gas the lower the temperature), and hence has an effect on the emissions of NOx. The main conflict is that, at higher speeds, a longer inlet open period increases the power developed. The down-side is that this causes a greater valve overlap and, at idle, this can greatly increase emissions of hydrocarbons. This has led to the successful introduction of electronically controlled valve timing.

10.2.5 Manifold designs

Gas flow within the inlet and exhaust manifolds is a very complex subject. The main cause of this complexity is the transient changes in flow that are due not only to changes in engine speed but also to the pumping action of the cylinders. This pumping action causes pressure fluctuations in the manifolds. If the manifolds and both induction and exhaust systems are designed to reflect the pressure wave back at just the right time, great improvements in volumetric efficiency can be attained. Many vehicles are now fitted with adjustable length induction tracts. Longer tracts are used at lower engine speeds and shorter tracts at higher speed.

10.2.6 Charge stratification

Key fact

If the charge mixture can be inducted into the cylinder in such a way that a richer mixture is in the proximity of the spark plug, then overall the cylinder charge can be much weaker.

If the charge mixture can be inducted into the cylinder in such a way that a richer mixture is in the proximity of the spark plug, then overall the cylinder charge can be much weaker. This can bring great advantages in fuel consumption, but the production of NOx can still be a problem. The later section on direct mixture injection development is a good example of the use of this technique. Many lean-burn engines use a form of stratification to reduce the chances of misfire and rough running.

10.2.7 Warm up time

A significant quantity of emissions produced by an average vehicle is created during the warm-up phase. Suitable materials and care in the design of the cooling system can reduce this problem. Some engine management systems even run the ignition timing slightly retarded during the warm-up phase to heat the engine more quickly.

10.2.8 Exhaust gas recirculation

This technique is used primarily to reduce peak combustion temperatures and hence the production of nitrogen oxides (NOx). Exhaust gas recirculation (EGR) can be either internal as mentioned above, due to valve overlap, or external via a simple arrangement of pipes and a valve (Figure 10.6). A proportion of exhaust gas is simply returned to the inlet side of the engine.

This EGR is controlled electronically as determined by a ROM in the ECU. This ensures that drivability is not affected and also that the rate of EGR is controlled. If the rate is too high, then the production of hydrocarbons increases. Figure 10.7 shows the effect of various rates of EGR.

One drawback of EGR systems is that they can become restricted by exhaust residue over a period of time, thus changing the actual percentage

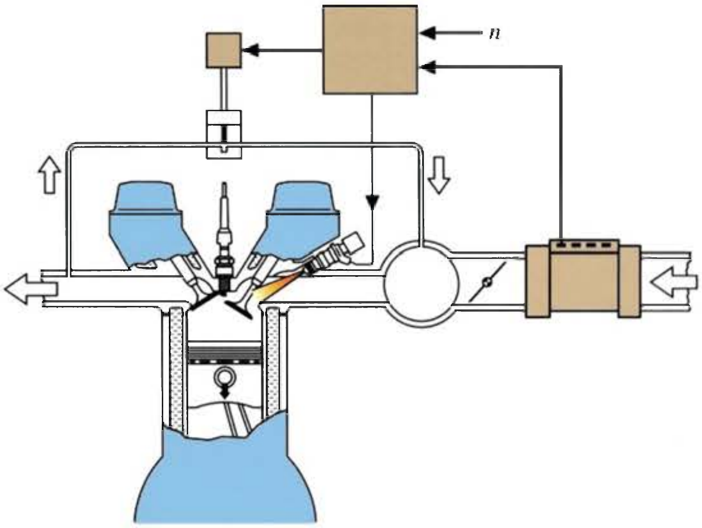


Figure 10.6 Exhaust gas recirculation system

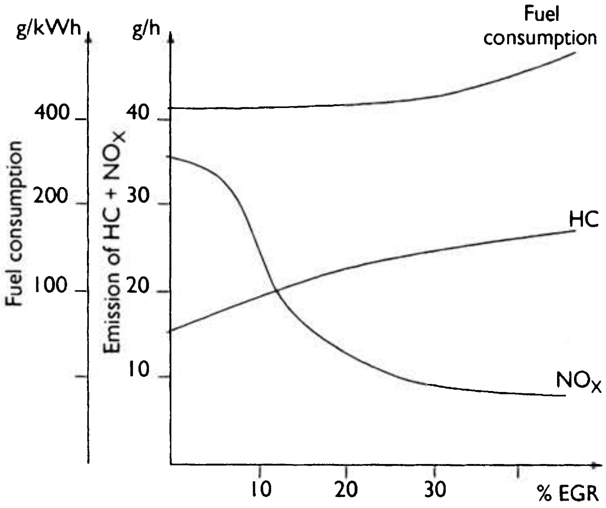


Figure 10.7 Effect of various rates of EGR

of recirculation. However, valves are now available that reduce this particular problem.

10.2.9 Ignition system

Key fact

The quality of a spark will determine its ability to ignite the mixture.

The ignition system can affect exhaust emissions in two ways; first, by the quality of the spark produced, and secondly, the timing of the spark. The quality of a spark will determine its ability to ignite the mixture. The duration of the spark in particular is significant when igniting weaker mixtures. The stronger the spark the less the likelihood of a misfire, which can cause massive increases in the production of hydrocarbons.

The timing of a spark is clearly critical but, as ever, is a compromise with power, drivability, consumption and emissions. Figure 10.8 is a graph showing the influence of ignition timing on emissions and fuel consumption. The production of carbon monoxide is dependent almost only on fuel mixture and is not significantly affected by changes in ignition timing. Electronic and digital ESA ignition systems have made significant improvements to the emission levels of today's engines.

10.2.10 Thermal after-burning

Prior to the more widespread use of catalytic converters, thermal after-burning was used to reduce the production of hydrocarbons. In fact, hydrocarbons do continue to burn in the exhaust manifold and recent research has shown that the type of manifold used, such as cast iron or pressed steel, can have a noticeable effect on the reduction of HC. At temperatures of about 600 °C, HC and CO are burnt or oxidized into H_2O and CO_2. If air is injected into the exhaust manifold just after the valves, then the after-burning process can be encouraged.

10.2.11 Catalytic converters

Stringent regulations in most parts of the world have made the use of a catalytic converter almost inevitable. The three-way catalyst (TWC) is therefore used to

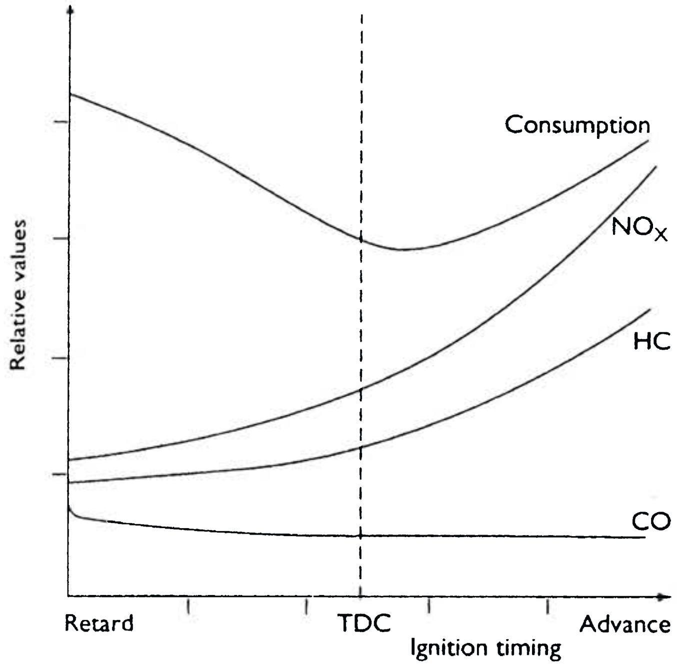


Figure 10.8 Influence of ignition timing on emissions and fuel consumption

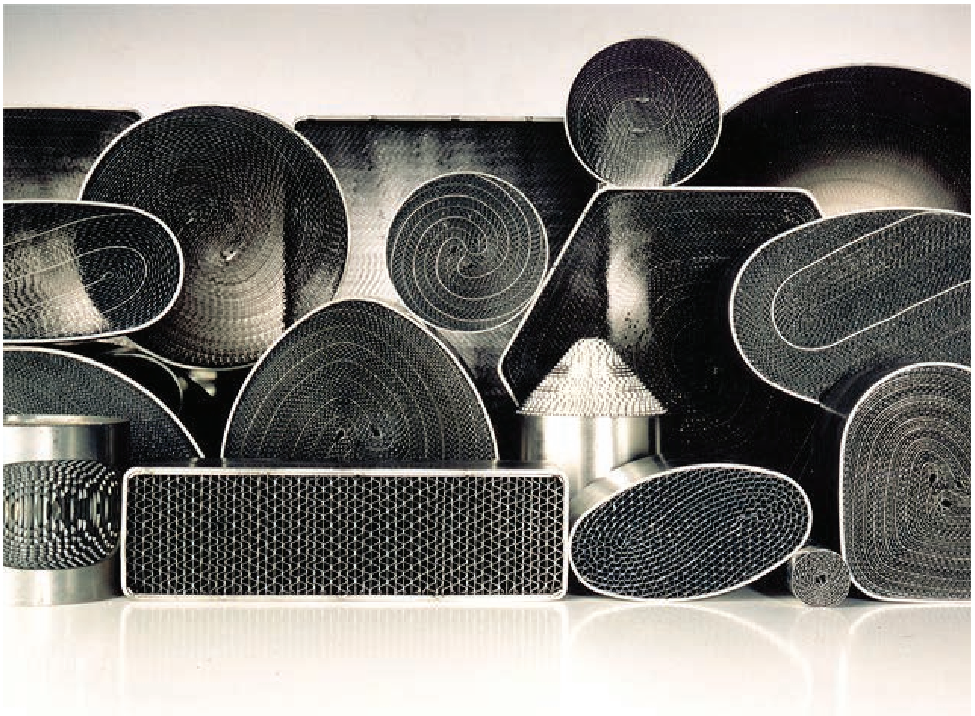


Figure 10.9 Catalytic converter metal substrate designs

Figure 10.10 Ceramic substrates

good effect by most manufacturers. It is a very simple device and looks similar to a standard exhaust box. Note that, in order to operate correctly, however, the engine must be run at, or very near to, stoichiometry. This is to ensure that the right 'ingredients' are available for the catalyst to perform its function. Figures 10.9 and 10.10 show two different substrates used in a catalytic converter to hold the catalytic materials.

There are many types of hydrocarbons but the following example illustrates the main reaction. Note that the reactions rely on some CO being produced by the engine in order to reduce the NOx. This is one of the reasons that manufacturers have been forced to run engines at stoichiometry. This legislation has tended to stifle the development of lean-burn techniques. The fine details of the emission regulations can in fact, have a very marked effect on the type of reduction techniques used. The main reactions in the 'cat' are as follows:

$$2CO + O_2 \rightarrow 2CO_2$$
$$2C_2H_6 + 2CO \rightarrow 4CO_2 + 6H_2O$$
$$2NO + 2CO \rightarrow N_2 + 2CO_2$$

The ceramic monolith type of base, when used as the catalyst material, is a magnesium aluminium silicate and, due to the several thousand very small channels, provides a large surface area. This area is coated with a wash coat of aluminium oxide, which further increases its effective surface area by a factor of about seven thousand. Noble metals are used for the catalysts. Platinum promotes the oxidation of HC and CO, and rhodium helps the reduction of NOx. The converter shown is the latest metal substrate type with a built-in manifold. The whole three-way catalytic converter only contains about 3–4 g of the precious metals.

The main catalytic converter is designed as a modern 2-layer converter with air-gap insulated central part. The position of the catalytic converter close

Definition

Catalyst: A substance that increases the rate of a chemical reaction without undergoing any permanent chemical change itself.

Key fact

The whole three-way catalytic converter only contains about 3–4 g of the precious metals.

to the engine ensures a fast response time (light-off) in the cold start phase. The fabricated manifold design both cuts the overall weight of the vehicle and also favours the lower thermal mass of the light-off catalytic converter. This innovative system thus already complies with future exhaust emission values.

Key fact

The ideal operating temperature range is from about 400 to 800 °C.

The ideal operating temperature range is from about 400 to 800 °C. A serious problem to counter is the delay in the catalyst reaching this temperature. This is known as the 'catalyst light-off time'. Various methods have been used to reduce this time as significant emissions are produced before 'light-off' occurs. Electrical heating is one solution, as is a form of burner, which involves lighting fuel inside the converter. Another possibility is positioning the converter as part of the exhaust manifold and down pipe assembly. This greatly reduces light-off time but gas flow problems, vibration and excessive temperature variations can be problems that reduce the potential life of the unit.

Catalytic converters can be damaged in two ways. The first is by the use of leaded fuel, which causes lead compounds to be deposited on the active surfaces, thus reducing the effective area, and, secondly, by engine misfire, which can cause the catalytic converter to overheat due to burning inside the unit. BMW, for example, uses a system on some vehicles where a sensor monitors the output of the ignition HT system and, if the spark is not present, will not allow fuel to be injected.

A further possible technique to reduce emissions during the warm-up time of the catalyst is to use a small electrically heated pre-converter as shown in Figure 10.11. Initial tests of this system show that the emissions of hydrocarbons during the warm-up phase can be reduced significantly. The problem yet to be solved is that about 30 kW of heat is required during the first 30 s to warm up the pre-converter. This will require a current in the region of 250 A; an extra battery may be one solution.

Key fact

For a catalytic converter to operate at its optimum conversion rate a narrow band within 0.5% of lambda value one is essential.

For a catalytic converter to operate at its optimum conversion rate in order to oxidize CO and HC whilst reducing NOx, a narrow band within 0.5% of lambda value one is essential. Lambda sensors in use at present tend to operate within about 3% of the lambda mean value. When a catalytic converter is in prime condition this is not a problem due to storage capacity within the converter for CO and O2. Damaged converters, however, cannot store a sufficient quantity of these gases and hence become less efficient. The damage, as suggested earlier in this section, can be due to overheating or 'poisoning' due to lead or even silicon. If the control can be kept within 0.5% of lambda the converter will continue to be effective even if damaged to some extent. Sensors are becoming available that can work to this tolerance. A second sensor fitted after the converter can be used to ensure ideal operation.

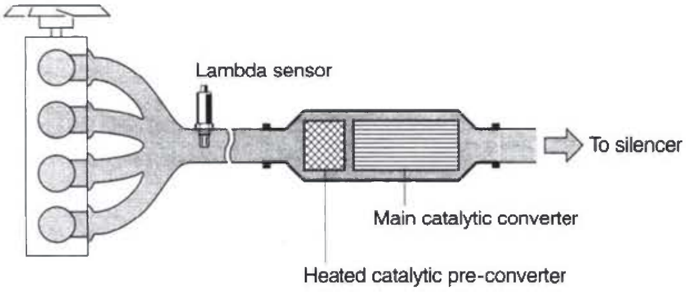


Figure 10.11 Electrically heated catalytic pre-converter

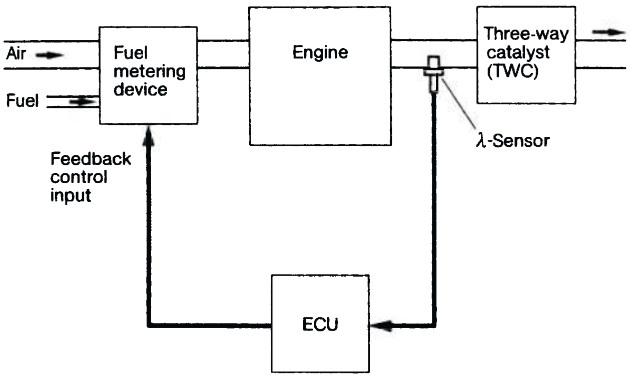


Figure 10.12 Fuel metering with closed loop control

10.2.12 Closed loop lambda control

Current regulations have almost made mandatory closed loop control of the air–fuel mixture in conjunction with a three-way catalytic converter. It was under discussion that a lambda value of 1 should become compulsory for all operating conditions, but this was not agreed.

Lambda control is a closed loop feedback system in that the signal from a lambda sensor in the exhaust can directly affect the fuel quantity injected. The lambda sensor is described in more detail in Chapter 2. Figure 10.12 shows a block diagram of a lambda control system.

A graph to show the effect of lambda control and a three-way catalyst (TWC) is shown in Figure 10.13.

The principle of operation is as follows: the lambda sensor produces a voltage that is proportional to the oxygen content of the exhaust, which is in turn proportional to the air–fuel ratio. At the ideal setting, this voltage is about 450 mV. If the voltage received by the ECU is below this value (weak mixture) the quantity of fuel injected is increased slightly. If the signal voltage is above the threshold (rich mixture) the fuel quantity is reduced. This alteration in the air–fuel ratio must not be too sudden as it could cause the engine to buck. To prevent this, the ECU contains an integrator, which changes the mixture over a period of time.

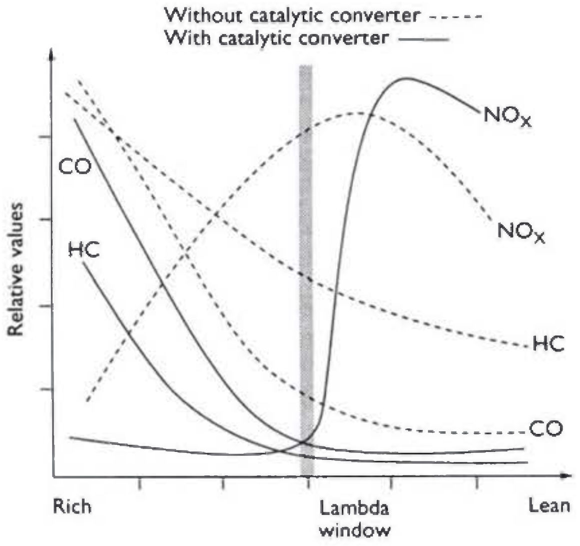


Figure 10.13 The effect of lambda control and a three way catalyst (TWC)

A delay also exists between the mixture formation in the manifold and the measurement of the exhaust gas oxygen. This is due to the engine's working cycle and the speed of the inlet mixture, the time for the exhaust to reach the sensor and the sensor's response time. This is sometimes known as 'dead time' and can be as much as one second at idle speed but only a few hundred milliseconds at higher engine speeds.

Definition

TWC: Three way catalyst.

Due to the dead time the mixture cannot be controlled to an exact value of 1. If the integrator is adjusted to allow for engine speed then it is possible to keep the mixture in the lambda window (0.97–1.03), which is the region in which the TWC is at its most efficient.

10.3 Engine management systems

10.3.1 Motronic M3

I have chosen the Motronic M3 system to outline key engine management features because it is well-known and reliable. It is also a good 'vehicle' for explaining different aspects of this technology (pun intended!).

The combination of ignition and injection control has several advantages. The information received from various sensors is used for computing both fuelling and ignition requirements. Perhaps more importantly, ignition and injection are closely linked. The influence they have on each other can easily be taken into account to ensure that the engine is working at its optimum, under all operation conditions. Overall, this type of system is less complicated than separate fuel and ignition systems and, in many cases, the ECU is able to work in an emergency mode by substituting missing information from sensors with pre-programmed values. This will allow limited but continued operation in the event of certain system failures.

Definition

ROM: Read only memory.

The ignition system is integrated and is operated without a high tension distributor. The ignition process is controlled digitally by the ECU. The data for the ideal characteristics are stored in ROM from information gathered during both prototyping and development of the engine. The main parameters for ignition advance are engine speed and load, but greater accuracy can be achieved by taking further parameters into account, such as engine temperature. This provides both optimum output and close control of anti-pollution levels. Performance and pollution level control means that the actual ignition point must, in many cases, be a trade-off between the two.

The injection system is multipoint and, as is the case for all fuel systems, the amount of fuel delivered is primarily determined by the amount of air drawn into the engine. The method for measuring these data is indirect in the case of this system as a pressure sensor is being used to determine the air quantity.

Electromagnetic injectors control the fuel supply into the engine. The injector open period is determined by the ECU. This will obtain very accurate control of the air–fuel mixture under all operating conditions of the engine. The data for this are stored in ROM in the same way as for the ignition.

Ignition system operation

The main source of reference for the ignition system is from the crankshaft position sensor. This is a magnetic inductive pick-up sensor positioned next to a

Figure 10.14 Motronic system components

flywheel ring containing 58 teeth. Each tooth takes up a 6° angle of the flywheel with one 12° gap positioned 114° before top dead centre (TDC) for the number one cylinder.

Typical resistance of the sensor coil is 800 Ω. The air gap between the sensor and flywheel ring is about 1 mm. The signal produced by the flywheel sensor is shown in Figure 10.15. It is essentially a sine wave with one cycle missing, which corresponds to the gap in the teeth of the reluctor plate.

The information provided to the ECU is engine speed from the frequency of the signal, and engine position from the number of pulses before or after the missed pulses.

Figure 10.16 shows a block diagram layout of how the ignition system is controlled. At ignition system level the ECU must be able to:

- Determine and create advance curves.
- Establish constant energy.
- Transmit the ignition signal direct to the ignition coil.

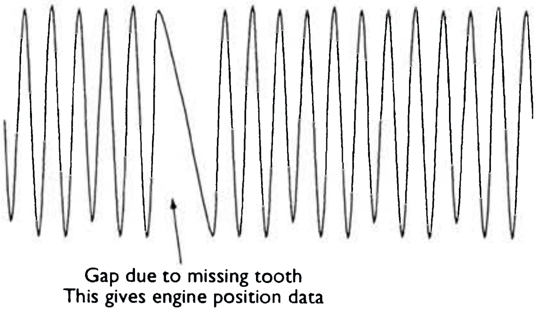


Figure 10.15 Crankshaft sensor signal

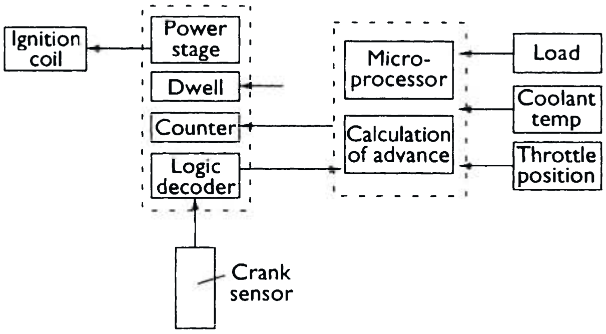


Figure 10.16 Simplified layout of the control of the ignition system

The basic ignition advance angle is obtained from a memorized cartographic map. This is held in a ROM chip within the ECU. The parameters for this are:

- Engine rpm – given by the flywheel sensor.
- Inlet air pressure – given by the manifold absolute pressure sensor.

The above two parameters (speed and load) give the basic setting but to ensure optimum advance angle the timing is corrected by:

- Coolant temperature.
- Air temperature.
- Throttle position.

The ignition is set to a predetermined advance during the starting phase. Figure 10.17 shows a typical advance map and a dwell map used by the Motronic

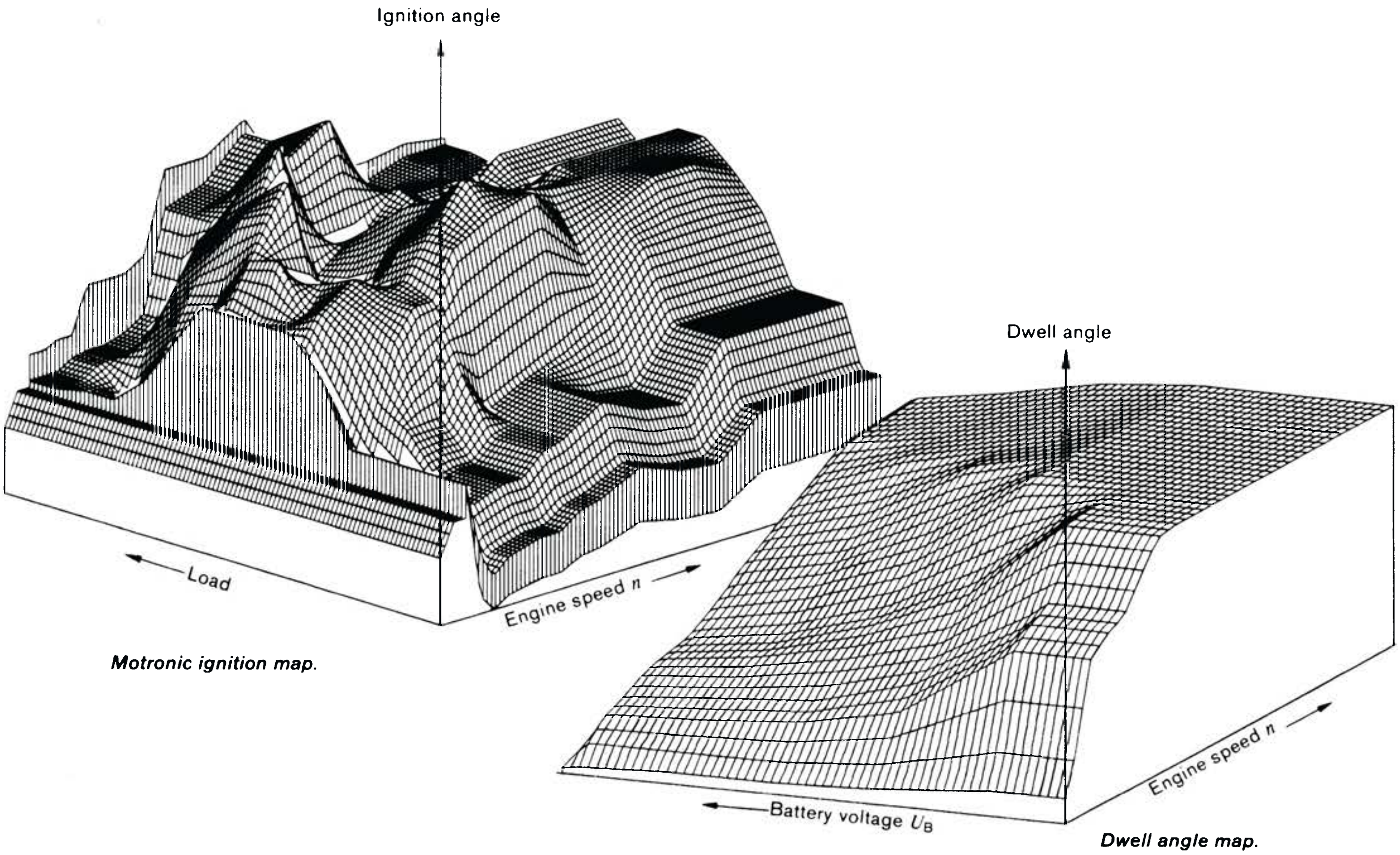


Figure 10.17 Engine timing and dwell maps

system. These data are held in ROM. For full ignition control, the electronic control unit has first to determine the basic timing for three different conditions:

- Under idling conditions, ignition timing is often moved very quickly by the ECU in order to control idle speed. When timing is advanced, engine speed will increase within certain limits.
- Full load conditions require careful control of ignition timing to prevent combustion knock. When a full load signal is sensed by the ECU (high manifold pressure) the ignition advance angle is reduced.
- Partial throttle is the main area of control and, as already stated, the basic timing is set initially by a programme as a function of engine speed and manifold pressure.

Corrections are added according to:

- Operational strategy.
- Knock protection.
- Phase correction.

The ECU will also control ignition timing variation during overrun fuel cut-off and reinstatement and also ensure anti-jerk control. When starting, the ignition timing plan is replaced by a specific starting strategy. Phase correction is when the ECU adjusts the timing to take into account the time taken for the HT pulse to reach the spark plugs. To ensure good drivability the ECU can limit the variations between the two ignition systems to a maximum value, which varies according to engine speed and the basic injection period.

Key fact

Phase correction is when the ECU adjusts the timing to take into account the time taken for the HT pulse to reach the spark plugs.

The anti-jerk function operates when the basic injection period is less than 2.5 ms and the engine speed is between 720 and 3200 rpm. This function operates to correct the programmed ignition timing in relation to the instantaneous engine speed and a set filtered speed; this is done to stabilize the engine rotational characteristics as much as possible.

In order to maintain constant high tension (HT) energy, the dwell period must increase in line with engine speed. To ensure the ignition primary current reaches its maximum at the point of ignition, the ECU controls the dwell by the use of another memory map, which takes battery voltage into account.

The signal from the flywheel sensor is virtually a sinusoid created as the teeth pass the winding. The zero value of this signal occurs as the sensor ‘sees’ the apex of each tooth. A circuit within the ECU (a Schmitt trigger) converts the signal into a square wave. The passage of the missing teeth gives a longer duration signal. The ECU detects the gap in the teeth and, from this, can determine the first TDC. The second TDC in the cycle is determined by counting 29 teeth, which is half a revolution. The ECU, having determined the ignition angle then controls the coil every half engine revolution. Using the reference signal, the ECU switches the coil on at a point determined by a number of teeth corresponding to the dwell, before the point determined by timing value, where the coil is switched off.

The ignition module is only used as a simple switch to control the coil primary windings. It consists of a Darlington-type amplifier. This switching function is carried out within the ECU on some systems, this choice very much depends on the location of the ECU compared with the ignition coil. Also, the heat generated by the switching of heavy current may be better separate from the main ECU. A final consideration is whether the interference caused by the switching could cause problems within the ECU.

The ‘distributorless’ ignition coil is made up of two primary windings and two secondary windings. The primary windings have a common 12 V supply and

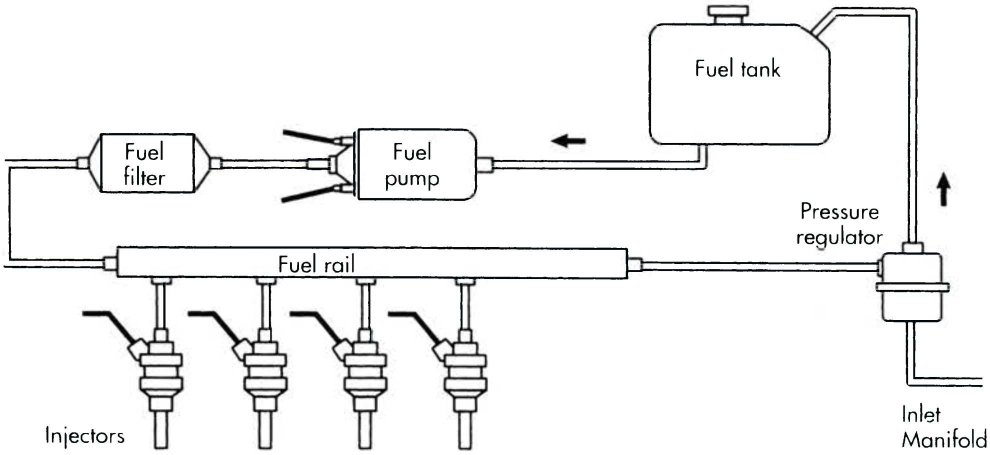


Figure 10.18 The main components of a fuel supply system

are switched to earth in turn in the normal manner. The primary resistance is of the order of 0.5 and the secondary resistance is 14.5 kΩ. The system works on the lost spark principle in that: cylinders 1 and 4 fire together as do 2 and 3. The disadvantage of this system is that one cylinder of each pair has the spark jumping from the plug earth electrode to the centre. However, owing to the very high energy available for the spark, this has no significant effect on performance.

The HT cables used are resistive. Spark plugs used for this system are standard but vary between types of engine. A gap of around 0.8 mm is the norm.

Fuel supply

Fuel is collected from the tank by a pump either immersed in it or outside, but near the tank. The immersed type is quieter in operation has better cooling and no internal leaks. The fuel is directed forwards to the fuel rail or manifold, via a paper filter. Figure 10.18 shows the fuel supply system.

Fuel pressure is maintained at about 2.5 bar above manifold pressure by a regulator mounted on the fuel rail. Excess fuel is returned to the tank. The fuel is usually picked up via a swirl pot in the tank to prevent aeration of the fuel. Each of the four inlet manifold tracts has its own injector.

Key fact

Fuel pressure is maintained at about 2.5 bar above manifold pressure by a regulator mounted on the fuel rail.

The fuel pump is a high-pressure type and is a two-stage device. A low-pressure stage, created by a turbine, draws fuel from the tank and a high-pressure stage, created by a gear pump, delivers fuel to the filter. It is powered by a 12 V supply from the fuel pump relay, which is controlled by the ECU as a safety measure.

The fuel pump characteristics are:

- Delivery – 120 litres per hour at 3 bars.
- Resistance – 0.8 Ω (static).
- Voltage – 12 V.
- Current – 10.5 A.

The rotation of the turbine draws fuel in via the inlet. The fuel passes through the turbine and enters the pump housing where it is pressurized by rotation of the pump and the reduction of the volume in the gear chambers. This pressure opens a residual valve and fuel passes to the filter. When the pump stops, pressure is maintained by this valve, which prevents the fuel returning. If, due to a faulty regulator or a blockage in the line, fuel pressure rises above 7 bar an over-pressure valve will open, releasing fuel back to the tank. Figure 9.34 (in Chapter 9) shows this type of pump.

The fuel filter is placed between the fuel pump and the fuel rail. It is fitted to ensure that the outlet screen traps any paper particles from the filter element. The filter will stop contamination down to between 8 and 10 μm. Replacement of the filter varies between manufacturers but 80 000 km (50 000 miles) is often recommended.

Key fact

The filter will stop contamination down to between 8 and 10 μm.

The fuel rail, in addition to providing a uniform supply to the injectors, acts as an accumulator. Depending on the size of the fuel rail some systems also use an extra accumulator. The volume of the fuel rail is large enough to act as a pressure fluctuation damper, ensuring that all injectors are supplied with fuel at a constant pressure.

Injectors and associated components

One injector (Figure 10.19) is used for each cylinder although very high performance vehicles may use two. The injectors are connected to the fuel rail by a rubber seal. The injector is an electrically operated valve manufactured to a very high precision. The injector comprises a body and needle attached to a magnetic core. When the winding in the injector housing is energized, the core or armature is attracted and the valve opens, compressing a return spring. The fuel is delivered in a fine spray to wait behind the closed inlet valve until the induction stroke begins. Providing the pressure across the injector remains constant, the quantity of fuel admitted is related to the open period, which in turn is determined by the time the electromagnetic circuit is energized.

Figure 10.19 Fuel injector

The injectors typically have the following characteristics:

- Supply voltage – 12 V.
- Resistance – 16 Ω.
- Static output – 150 cc per minute at 3 bar.

The purpose of the fuel pressure regulator is to maintain differential pressure across the injectors at a pre-determined constant. This means the regulator must adjust the fuel pressure in response to changes in manifold pressure. It is made of two compressed cases containing a diaphragm, spring and a valve. Figure 10.20 shows a fuel pressure regulator similar to those used on this and many other injection systems.

Figure 10.20 Fuel pressure regulator

The calibration of the regulator valve is determined by the spring tension. Changes in manifold pressure vary the basic setting. When the fuel pressure is sufficient to move the diaphragm, the valve opens and allows fuel to return to the tank. The decrease in pressure in the manifold, also acting on the diaphragm for example, idle speed, will allow the valve to open more easily, hence maintaining a constant differential pressure between the fuel rail and the inlet manifold. This is a constant across the injectors and hence the quantity of fuel injected is determined only by the open time of the injectors. The differential pressure is maintained at about 2.5 bar.

Key fact

The calibration of the regulator valve is determined by the spring tension.

Key fact

The differential pressure is maintained at about 2.5 bar.

The air supply circuit will vary considerably between manufacturers but an individual manifold from a collector housing, into which the air is fed via a simple butterfly valve, essentially supplies each cylinder. The air is supplied from a suitable filter. A supplementary air circuit is utilized during the warm-up period after a cold start and to control idle speed.

Fuel mixture calculation

The quantity of fuel to be injected is determined primarily by the quantity of air drawn into the engine. This is dependent on two factors:

- Engine rpm.
- Inlet manifold pressure.

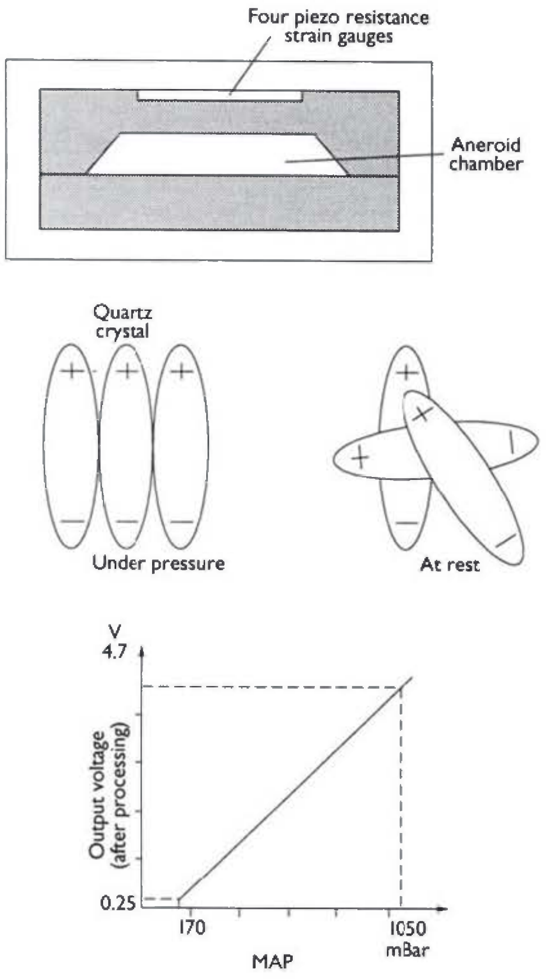


Figure 10.21 Pressure sensor and its voltage output

This speed load characteristic is held in the ECU memory in ROM look-up tables.

A sensor connected to the manifold by a pipe senses the manifold absolute pressure. It is a piezoelectric-type sensor, where the resistance varies with pressure. The sensor is fed with a stabilized 5 V supply and transmits an output voltage according to the pressure. The sensor is fitted away from the manifold and hence a pipe is required to connect it. A volumetric capacity is usually fitted in this line to damp down pressure fluctuations. The output signal varies between about 0.25 V at 0.17 bar to about 4.75 V at 1.05 bar. Figure 10.21 shows a pressure sensor and its voltage output.

The density of air varies with temperature such that the information from the MAP sensor on air quantity will be incorrect over wide temperature variations. An air temperature sensor is used to inform the ECU of the inlet air temperature such that the ECU may correct the quantity of fuel injected. As the temperature of air decreases its density increases and hence the quantity of fuel injected must also be increased.

The sensor is a negative temperature coefficient (NTC) resistor. The resistance value decreases as temperature increases and vice versa. The output characteristic of this sensor is non-linear. Further details about this type of sensor and one solution to the non-linear response problem are examined in Chapter 2.

In order to operate the injectors, the ECU needs to know – in addition to air pressure – the engine speed to determine the injection quantity. The same

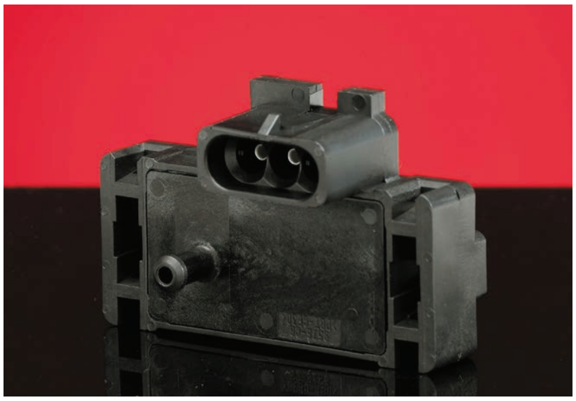


Figure 10.22 Replacement MAP sensor supplied by Delphi (Source: Delphi Media)

flywheel sensor used by the ignition system provides this information. All four injectors operate simultaneously, once per engine revolution, injecting half of the required fuel. This helps to ensure balanced combustion. The start of injection varies according to ignition timing.

A basic open period for the injectors is determined by using the ROM information relating to manifold pressure and engine speed. Two corrections are then made, one relative to air temperature and another depending on whether the engine is idling, at full or partial load.

The ECU then carries out another group of corrections, if applicable:

- after-start enrichment;
- operational enrichment;
- acceleration enrichment;
- weakening on deceleration;
- cut-off on overrun;
- reinstatement of injection after cut-off;
- correction for battery voltage variation.

Under starting conditions, the injection period is calculated differently. This is determined mostly from a set figure, which is varied as a function of temperature.

The coolant temperature sensor is a thermistor and is used to provide a signal to the ECU relating to engine coolant temperature. The ECU can then calculate any corrections to fuel injection and ignition timing. The operation of this sensor is the same as the air temperature sensor.

The throttle potentiometer is fixed on the throttle butterfly spindle and informs the ECU of the throttle position and rate of change of throttle position. The sensor provides information on acceleration, deceleration and whether the throttle is in the full load or idle position. Figure 10.23 shows the throttle potentiometer and its electrical circuit. It comprises a variable resistance and a fixed resistance. As is common with many sensors, a fixed supply of 5 V is provided and the return signal will vary approximately between 0 and 5 V. The voltage increases as the throttle is opened.

Key fact

The throttle potentiometer is fixed on the throttle butterfly spindle and informs the ECU of the throttle position and rate of change of throttle position.

The operation functions employed by this system can be examined under a number of headings or phases, as follows.

Starting phase

Entry to the starting phase occurs as soon as the ECU receives a signal from the flywheel sensor. The ignition advance is determined relative to the engine speed

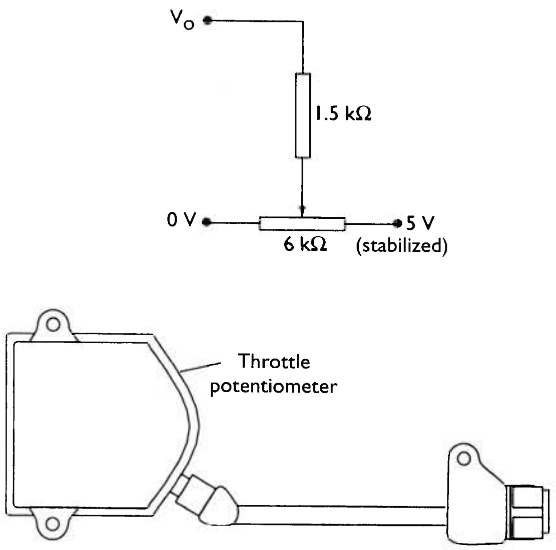


Figure 10.23 Throttle potentiometer and its electrical circuit

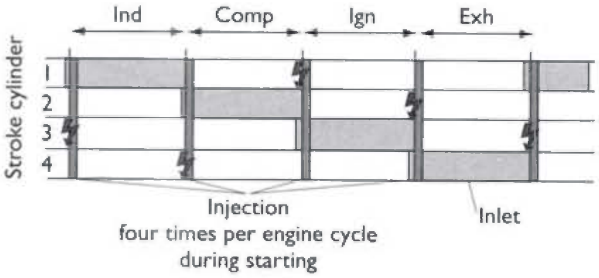


Figure 10.24 Injection and ignition timing relative to engine position

and the water temperature. The ECU operates the injectors four times per engine cycle (twice per crankshaft revolution) in order to obtain the most uniform mixture and to avoid wetting the plugs during the starting phase. Figure 10.24 shows the injection and ignition timing relative to engine position. Injection ceases 24° after the flywheel TDC signal. The ECU sets an appropriate injection period, corrected in relation to water temperature if starting from cold and air temperature if starting from hot. Exit from this starting phase is when the engine speed passes a threshold determined by water temperature.

After-start enrichment phase

Enrichment is necessary to avoid stalling after starting. The amount of enrichment is determined by water and air temperature and decreases under control of the ECU. If the engine is cold or an intermediate temperature, the initial mixture is a function of water temperature. If the engine is hot, the initial mixture is a function of air and water temperature. Figure 10.25 is a representation of the decreasing mixture enrichment after a cold start. If the engine happens to stop within a certain period of time just after a cold start, the next post-start enrichment will be reduced slightly.

Key fact

During warm up, the ignition timing is corrected in relation to water temperature.

Engine cold running phase

During warm up, the ignition timing is corrected in relation to water temperature. Timing will also alter depending on engine speed and load. During the warm-up phase, the injector open period is increased by the coolant temperature signal

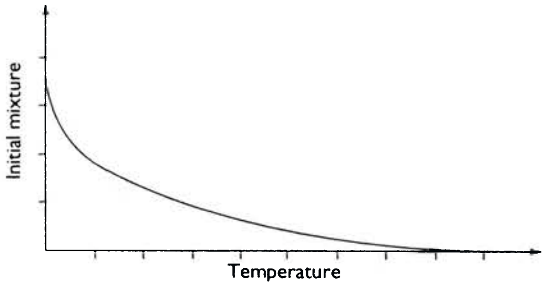


Figure 10.25 Decreasing mixture enrichment after a cold start

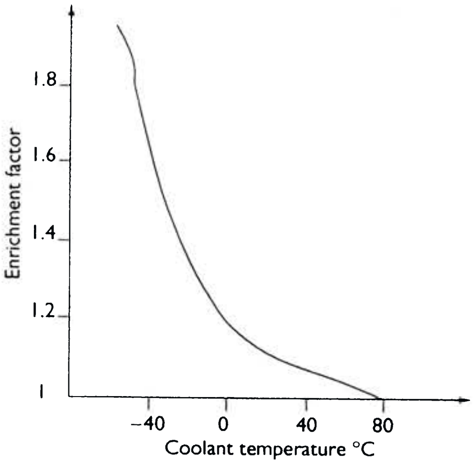


Figure 10.26 Enrichment factor during warm up

to make up for fuel losses and to prevent the engine speed dropping. The enrichment factor is reduced as the resistance of the temperature sensor falls, finally ceasing at 80 °C. Figure 10.26 shows the enrichment factor during warm up. The enrichment factor is determined by engine speed and temperature at idle and at other times by the programmed injection period relative to engine speed as well as the water temperature. To overcome the frictional resistance of a cold engine it is important to increase the mixture supply. This is achieved by using a supplementary air control device, which allows air to bypass the throttle butterfly.

Idling phase

Air required for idling bypasses the throttle butterfly by a passage in the throttle housing. A volume screw is fitted for adjustment of idle speed. Idle mixture adjustment is carried out electronically in response to the adjusting of a potentiometer, either on the ECU or as a separate unit. The ignition and injection functions for idle condition are set using information from the throttle potentiometer that the throttle is at the idle position, and engine speed is set by information from the flywheel sensor.

Full load phase

Under full load conditions the ignition timing is related to engine speed and full load information from the throttle potentiometer. The injection function in order to achieve maximum power must be set such that the mixture ratio is increased to 1 : 1. The information from the throttle potentiometer triggers a programme in the ECU to enrich the mixture in relation to engine speed in order to ensure maximum power over the speed range but also to minimize the risk of knocking.

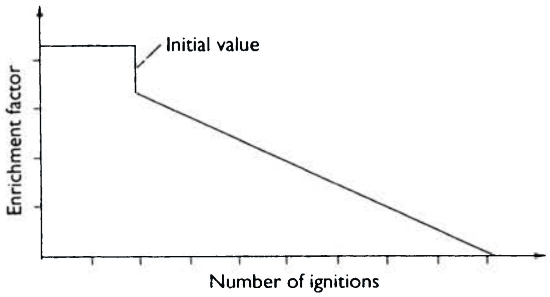


Figure 10.27 Acceleration enrichment phase

It is also important not to increase fuel consumption unnecessarily and not to allow significant increases in exhaust emissions.

Acceleration phase

When a rapid acceleration is detected by the ECU from the rate of change of the throttle potentiometer signal, enrichment occurs over a certain number of ignitions. The enrichment value is determined from water temperature and pressure variations in the inlet manifold. The enrichment then decreases over a number of ignitions. Figure 10.27 shows the acceleration enrichment phase. The enrichment is applied for the calibrated number of ignitions and then reduced at a fixed rate until it is non-existent. Acceleration enrichment will not occur if the engine speed is above 5000 rpm or at idle. Under very strong acceleration it is possible to have unsynchronized injection. This is determined from the water temperature, a ROM map of throttle position against engine speed and a battery voltage correction.

Deceleration phase

If the change in manifold pressure is greater than about 30 mbar the ECU causes the mixture to be weakened relative to the detected pressure change.

Key fact

Injection cut-off occurs when the throttle is closed and the engine speed is above a threshold related to water temperature (about 1500 rpm).

Injection cut-off on deceleration phase

This is designed to improve fuel economy and to reduce particular emissions of hydrocarbons. It will occur when the throttle is closed and when the engine speed is above a threshold related to water temperature (about 1500 rpm). When the engine speed falls to about 1000 rpm, injection recommences with the period rising to the value associated with the current engine speed and load. Figure 10.28 shows the strategy used to control injection cut-out and reinstatement.

Knock protection phase

Ignition timing is also controlled to reduce jerking and possible knocking during cut-off and reinstatement. The calculated advance is reduced to keep the ignition just under the knock limit. The advance correction against knock is a programme relating to injection period, engine speed and water/air temperature.

Engine speed limitation

Injection is cut-off when the engine speed rises above 6900 rpm and is reinstated below this figure. This is simply to afford some protection against over-revving of the engine and the damage that may be caused.

Battery voltage correction

This is a correction in addition to all other functions in order to compensate for changes in system voltage. The voltage is converted every TDC and the correction is then applied to all injection period calculations. On account of the

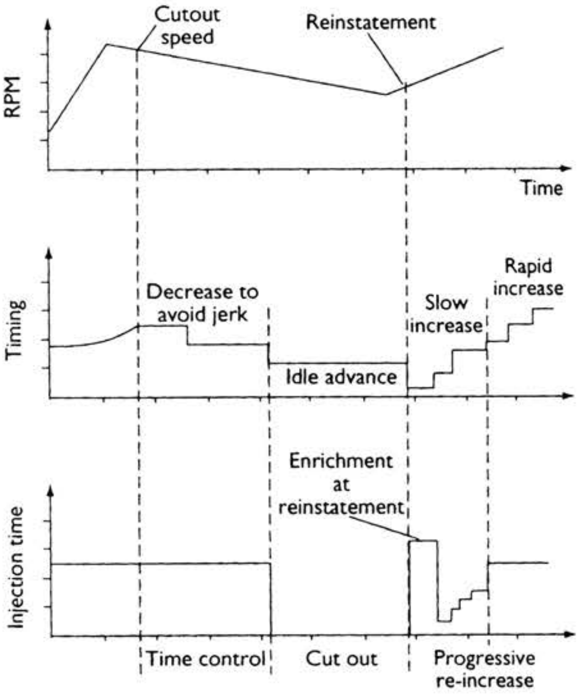


Figure 10.28 Strategy used to control injection cut-out and reinstatement

time taken for full current to flow in the injector winding and the time taken for the current to cease, a variation exists depending on applied voltage. Figure 10.29 shows how this delay can occur; if S_1 is greater than S_2 a correction is required. $S_1 - S_2 = S$ where S represents the time delay due to the inductance of the injector winding.

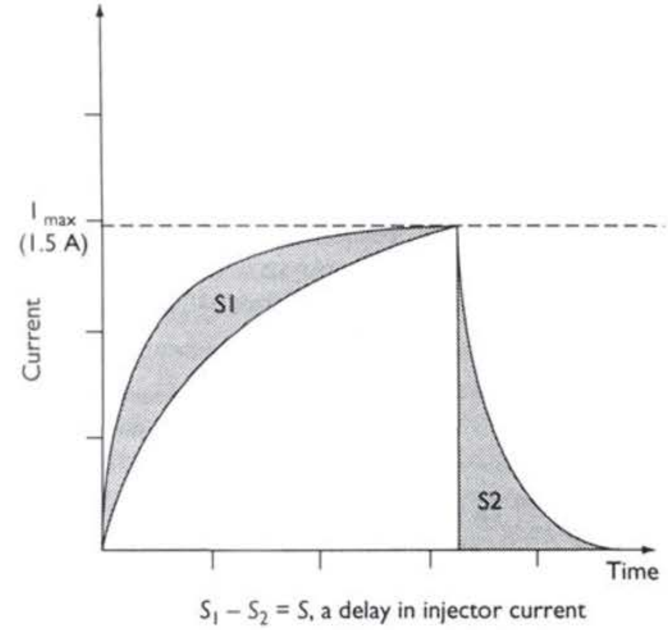


Figure 10.29 injector operation time curve

10.3.2 DI-Motronic

Bosch's high-pressure gasoline direct injection (GDi) (Figures 10.30–10.32) system for petrol engines is based on a pressure reservoir and a fuel rail, which

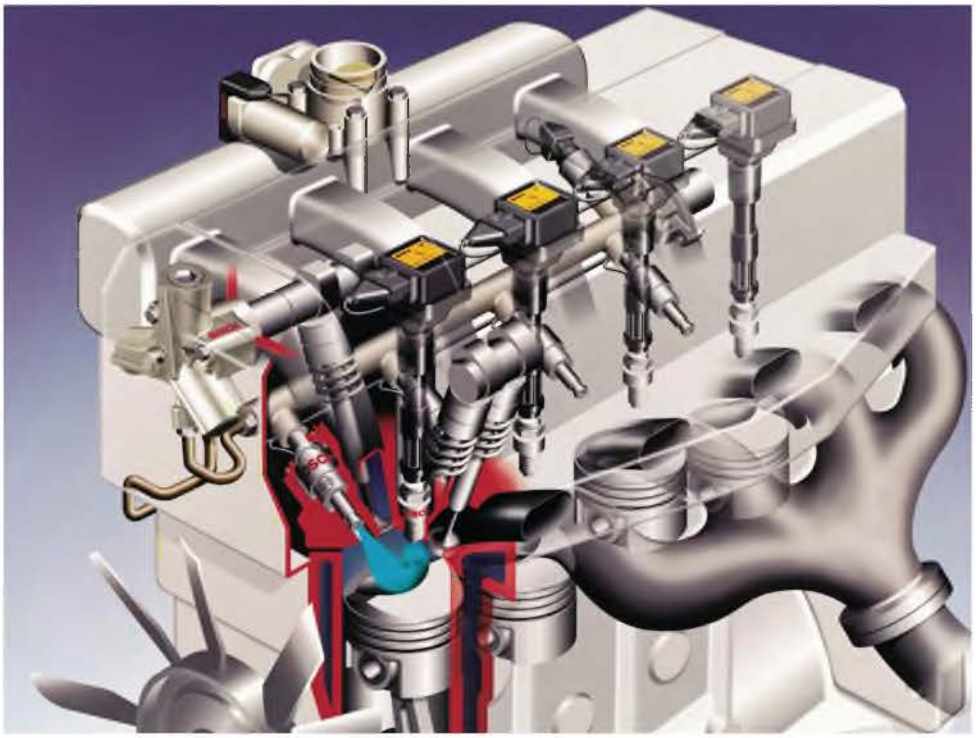


Figure 10.30 Gasoline direct injection on an engine (Source: Bosch Media)

Key fact

Gasoline direct injection (GDi) uses a pressure of up to 120 bar so fuel can be injected directly into the combustion chamber.

a high-pressure pump charges to a regulated pressure of up to 120 bar. The fuel can therefore be injected directly into the combustion chamber via electromagnetic injectors.

The air mass drawn in is adjusted through an electronically controlled throttle valve and is measured with the help of an air mass meter. For mixture control, a wide-band oxygen sensor is used in the exhaust, before the catalytic converters.

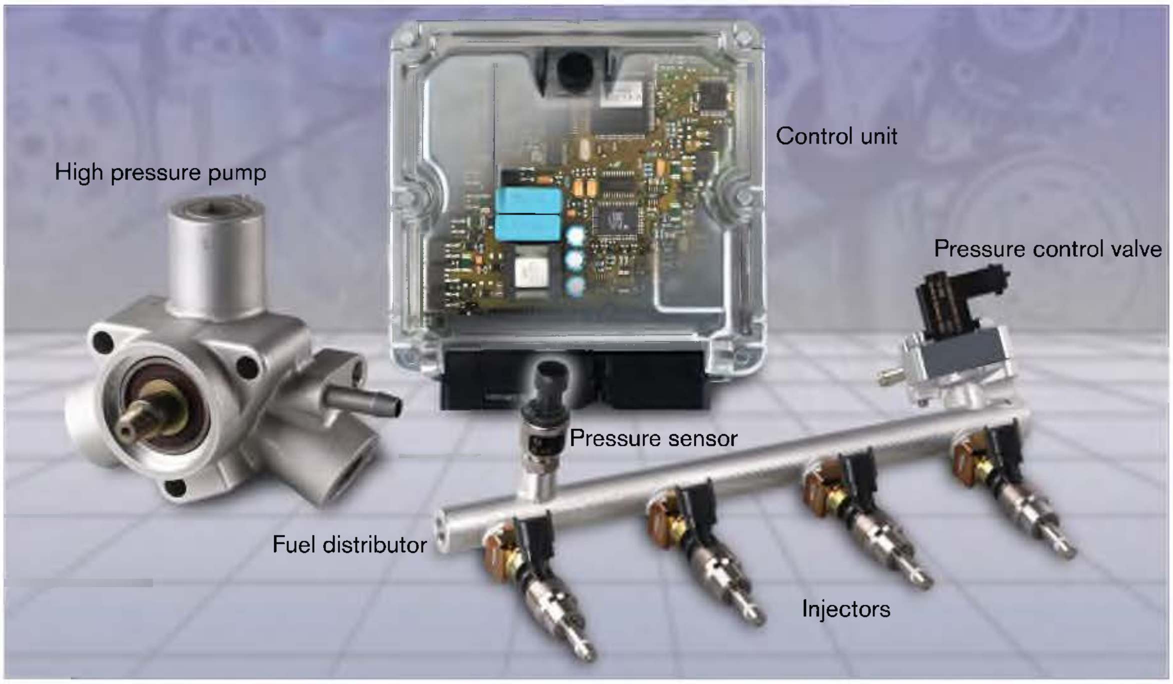


Figure 10.31 Components of gasoline direct injection (Source: Bosch Media)

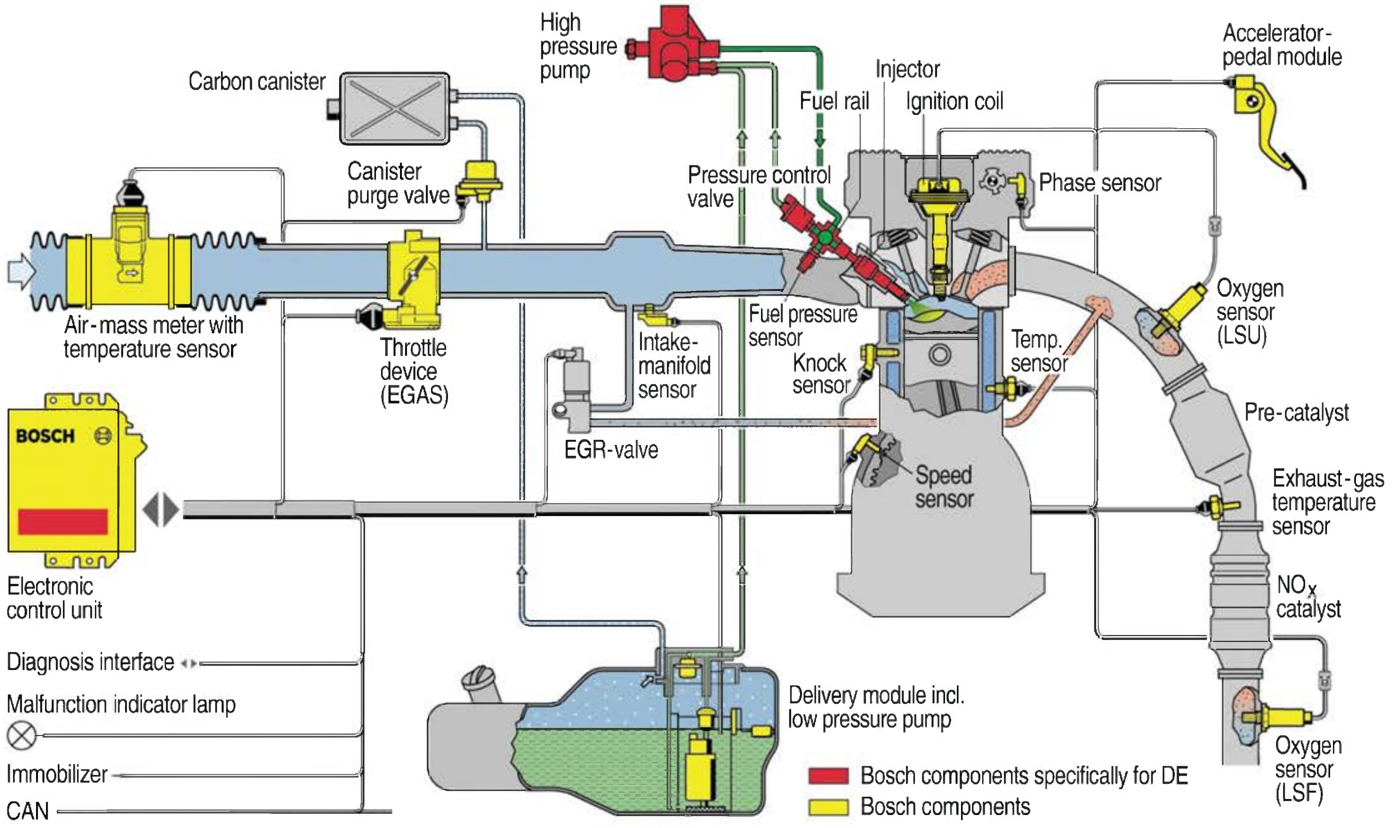


Figure 10.32 Bosch DI-Motronic (Source: Bosch Media)

This sensor can measure a range between lambda = 0.8 and infinity. The electronic engine control unit regulates the operating modes of the engine with gasoline direct injection in three ways:

1. Stratified charge operation – with lambda values greater than 1.
2. Homogenous operation – at lambda = 1.
3. Rich homogenous operation – with lambda = 0.8.

Compared to the traditional manifold injection system, the entire fuel amount must be injected in full-load operation in a quarter of the time. The available time is significantly shorter during stratified charge operation in part-load. Especially at idle, injection times of less than 0.5 milliseconds are required due to the lower fuel consumption. This is only one-fifth of the available time for manifold injection.

The fuel must be atomized very finely in order to create an optimal mixture in the brief moment between injection and ignition (Figures 10.33 and 10.34). The fuel droplets for direct injection are on average smaller than 20 μm (micro metres i.e. a millionth of a metre). This is one-fifth of the droplet size reached with the traditional manifold injection and one-third of the diameter of a single human hair. It improves efficiency considerably. However, even more important than fine atomization is even fuel distribution in the injection beam. This is done to achieve fast and uniform combustion.

Conventional spark ignition engines have a homogenous (well-mixed up!) air/fuel mixture at a 14.7 to 1 ratio, corresponding to a value of lambda = 1. Direct injection engines, however, operate according to the stratified charge concept in the part-load range and function with high excess air. In return, very low fuel consumption is achieved.

Key fact

The oxygen sensor on the Bosch GDi system can measure a range between lambda = 0.8 and infinity.

Definition

A micro metre (μm) is a millionth of a metre.

Key fact

The fuel droplets for direct injection are on average smaller than 20 μm.

Figure 10.33 Injector for direct injection (Source: Bosch Media)

Figure 10.34 Fuel droplet size is important (Source: Bosch Media)

Definition

Stratified: Arranged in approximately horizontal layers.

Definition

Homogenous: Of uniform composition throughout.

Fuel injection just before the ignition point and injection directly into the combustion chamber is to create a stratified (layered) mode (Figures 10.35 and 10.36). The result is a combustible air/fuel mixture cloud on the spark plug, cushioned in a thermally insulated layer, composed of air and residual gas. This raises the efficiency level because heat loss is avoided on the combustion chamber walls. The engine operates with an almost completely opened throttle valve, which avoids additional charge losses.

With stratified charge operation, the lambda value in the combustion chamber is between about 1.5 and 3. In the part-load range, gasoline direct injection achieves the greatest fuel savings with up to 40% at idle compared to conventional petrol injection processes.

With increasing engine load, and therefore increasing injection quantities, the stratified charge cloud becomes even richer and emission characteristics become worse. Like diesel engine combustion, soot may form. In order to prevent this, the DI-Motronic engine control converts to a homogenous cylinder

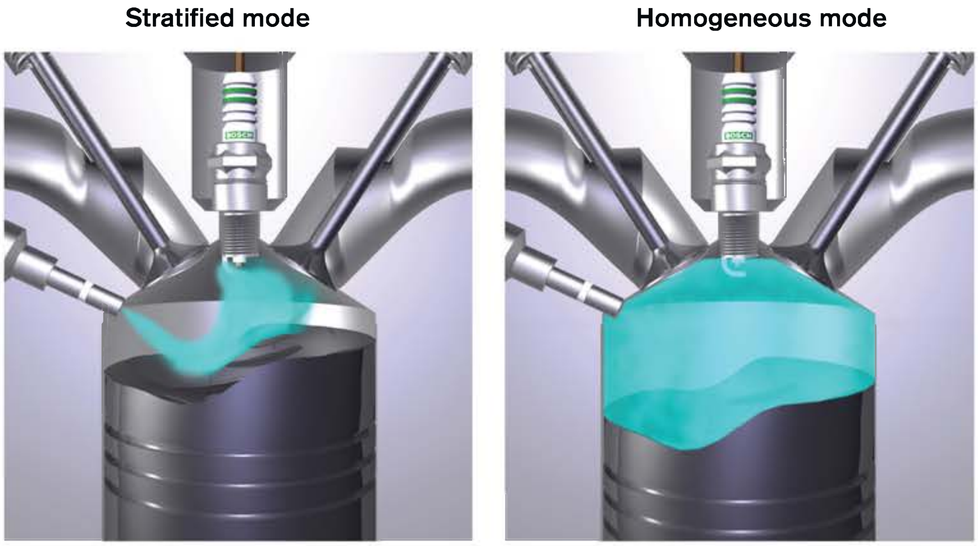


Figure 10.35 Operating modes (Source: Bosch Media)

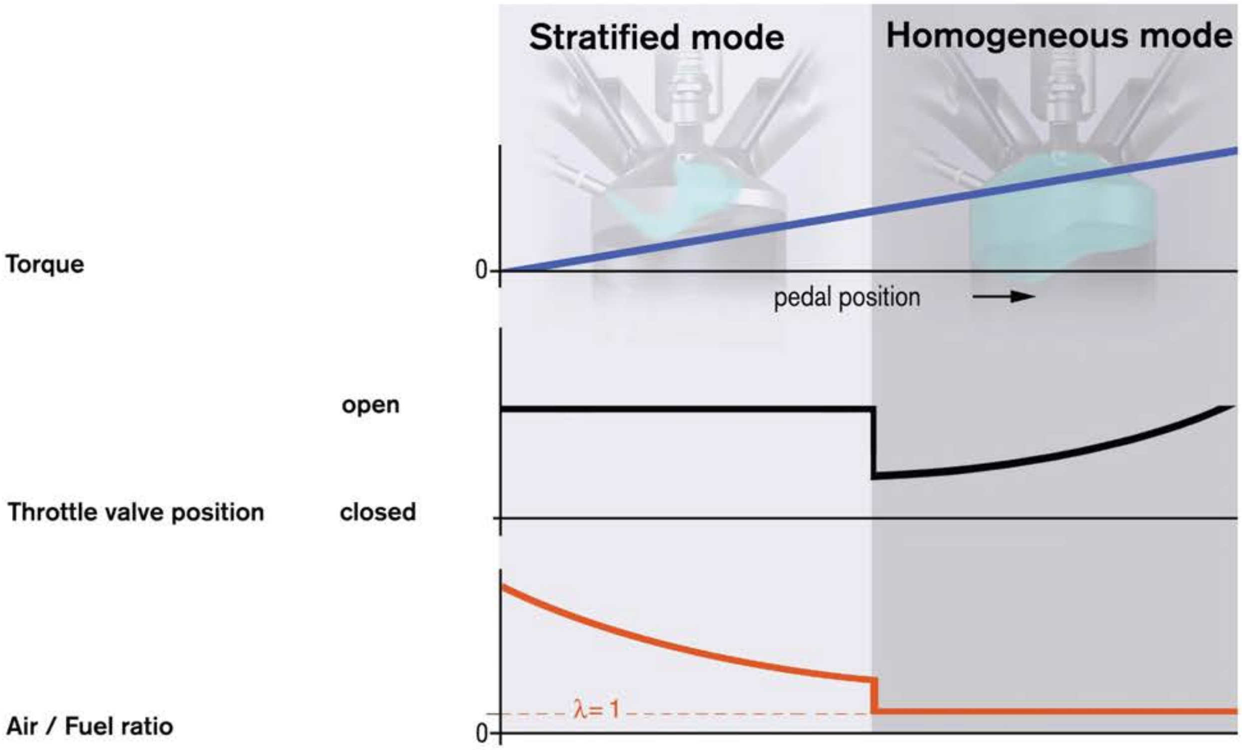


Figure 10.36 Switching between operating modes depending on engine load (Source: Bosch Media)

charge at a pre-defined engine load. The system injects very early during the intake process in order to achieve a good mixture of fuel and air at a ratio of lambda = 1.

As is the case for conventional manifold injection systems, the amount of air drawn in for all operating modes, is adjusted through the throttle valve according

Figure 10.37 ECU, rail and injectors (Source: Bosch Media)

to the desired torque specified by the driver. The Motronic ECU calculates the amount of fuel to be injected from the drawn-in air mass and performs an additional correction via lambda control. In this mode of operation, a torque increase of up to 5% is possible. Both the thermodynamic cooling effect of the fuel vaporizing directly in the combustion chamber, and the higher compression of the engine with gasoline direct injection, play a role in this.

For these different operating modes, two central demands are raised for engine control:

1. The injection point must be adjustable between 'late' (during the compression phase) and 'early' (during the intake phase) depending on the operating point.
2. The adjustment for the drawn-in air mass must be detached from the throttle pedal position in order to permit un-throttled engine operation in the lower load range. However, throttle control in the upper load range must also be permitted.

With optimal use of the advantages, the average fuel saving is up to 15%.

In stratified charge operation the nitrogen oxide (NOx) segments in the very lean exhaust cannot be reduced by a conventional, three-way catalytic converter. The NOx can be reduced by approximately 70% through exhaust returns before the catalytic converter. However, this is not enough to fulfil the ambitious emission limits of the future. Therefore, emissions containing NOx must undergo special treatment. Engine designers are using an additional NOx accumulator catalytic converter in the exhaust system. The NOx is deposited in the form of nitrates (HNO_3) on the converter surface, with the oxygen still contained in the lean exhaust (Figure 10.38).

Key fact

On a GDi system, NOx emissions are reduced by an accumulator catalytic converter in the exhaust system.

The capacity of the NOx accumulator catalytic converter is limited. As soon as it's exhausted, the catalytic converter must be regenerated. In order to remove the deposited nitrates, the DI-Motronic briefly changes over to its third operating mode (rich homogenous operation with lambda values of about 0.8). The nitrate together with the carbon monoxide (CO) is reduced in the exhaust to non-harmful nitrogen and oxygen. When the engine operates in this range, the engine torque is adjusted according to the accelerator pedal position via the throttle valve opening. Engine management has the difficult task of changing between the two different operating modes, in a fraction of a second, in a way not noticeable to the driver.

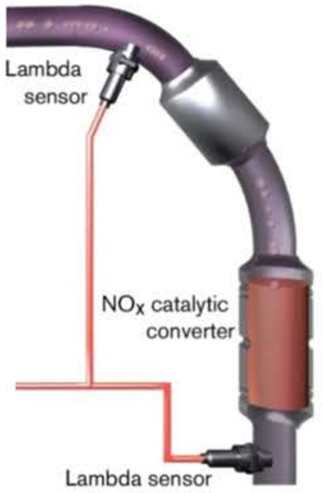


Figure 10.38 NOx catalytic converter

The continuing challenge, set by legislation, is to reduce vehicle emissions to very low levels. Bosch is a key player in the development of engine management systems. The DI Motronic system is now used by many manufacturers.

10.3.3 ME-Motronic principles

The system components of ME Motronic (Figure 10.40) are similar to the previously outlined DI-Motronic. The M system uses a throttle cable and a bypass, ME is fully gas-by-wire, and DI uses a high pressure pump. The M-Motronic and ME-Motronic are described as torque-based engine management systems. This is because they work on the concept that the throttle pedal position is converted by the microprocessor into a torque set-point value. The aim of this torque structure is to sort the very many different functions that require a torque. For example, idle speed control and transmission shit control may need different torques.

Key fact

The M-Motronic and ME-Motronic are torque-based engine management systems.

This torque calculation process is outlined in Figure 10.39. In this way contradictions can be resolved by prioritizing certain functions. The most

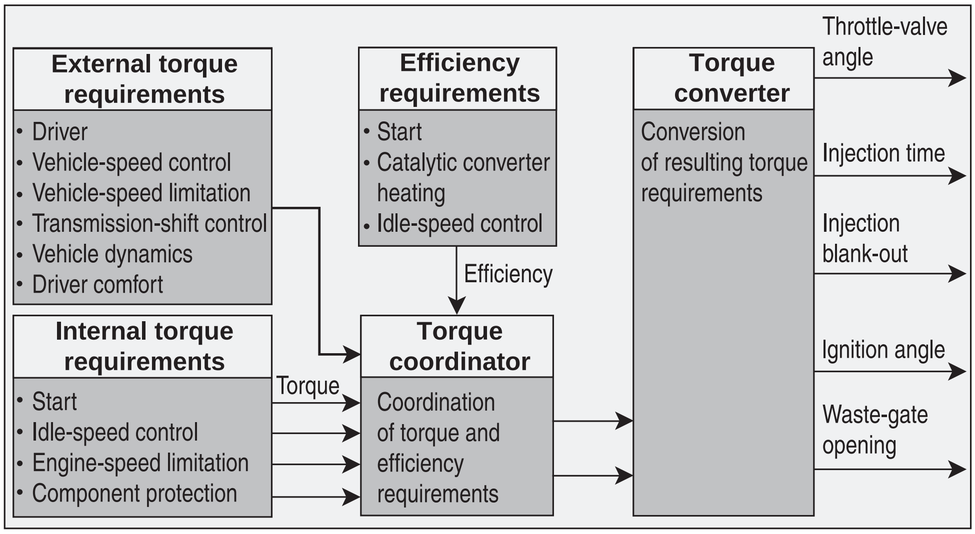


Figure 10.39 Torque-based system structure

appropriate setting can therefore be used taking the driver's requirements into account as well as those of other systems.

10.4 Other aspects of engine management

10.4.1 Introduction

Research is going on all the time into different ways of reducing emissions in order to keep within the current and expected legislation. The one potential change is that CO_2 is included in the legislation. This will, in effect, make fuel consumption as big an issue as noxious emissions. Some general areas under control of the engine management are outlined below.

10.4.2 Variable valve timing

Many engines now employ variable valve timing (VVT) to optimise the inlet valve timing with respect to engine speed and load conditions (Figures 10.41–10.43). Some also control the exhaust camshaft.

As air enters the engine through the inlet manifold this forms a column of moving air that possesses kinetic energy. The pulsating nature of the engines air consumption creates pressure waves in this air column. The energy in these pressure waves can be harnessed to assist in charging the cylinder, increasing the volumetric efficiency of the engine. In order to do this, the valve opening point must be optimised according to the engine condition, and with variable valve timing this can be achieved to increase engine torque and power at various points in the operating speed range.

Key fact

VVT optimises the valve opening point to increase engine torque and power at various points in the operating speed range.

There are various technologies available to provide the required phase angle between the cam drive and the camshaft for variable valve timing. It can be generated via a hydraulic mechanism in the cam wheel that is controlled via a valve assembly from the engine ECU. Cam wheel actuators can employ a 'helix' or pressure differential actuation principles. In addition, some engines have employed valve mechanisms with alternative cam profiles where the engine switches over to a different cam lift profile at certain engine speeds.

1 Activated-charcoal canister,
2 Check valve,
3 Canister-purge valve,
4 Intake-manifold pressure sensor,
5 Fuel rail/Injector,
6 Ignition coil/Spark plug,
7 Phase sensor,
8 Electric secondary-air injection pump,
9 Secondary-air injection valve,
10 Air-mass meter,
11 Throttle-valve assembly (ETC),
12 EGR valve,
13 Knock sensor,
14 RPM sensor,
15 Temperature sensor,
16 Lambda oxygen sensor,
17 Electronic control unit (ECU),
18 Diagnosis interface,
19 Diagnosis lamp,
20 Vehicle immobilizer,
21 Tank pressure sensor,
22 In-tank pump assembly,
23 Accelerator-pedal module,
24 Battery.

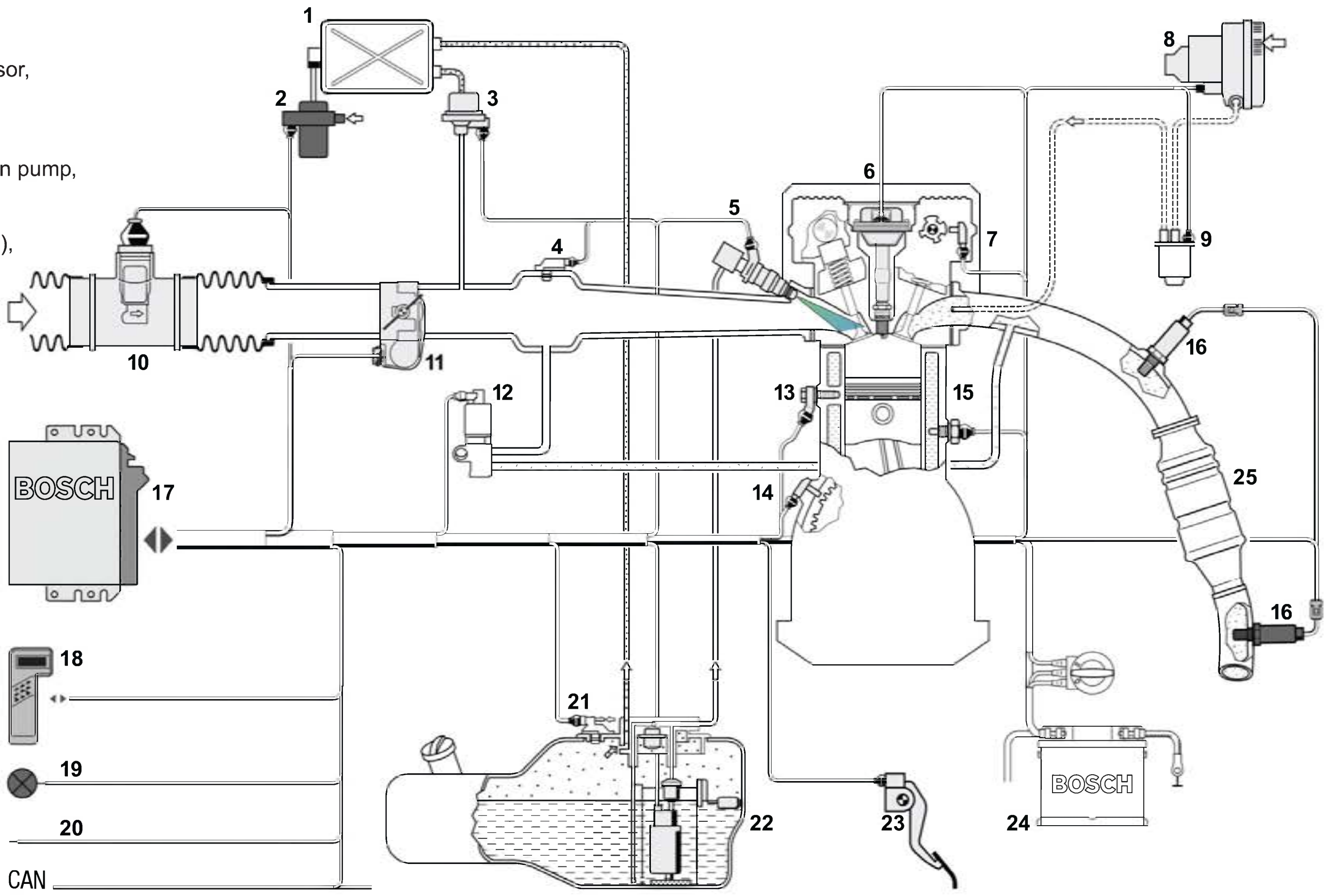


Figure 10.40 ME-Motronic engine management

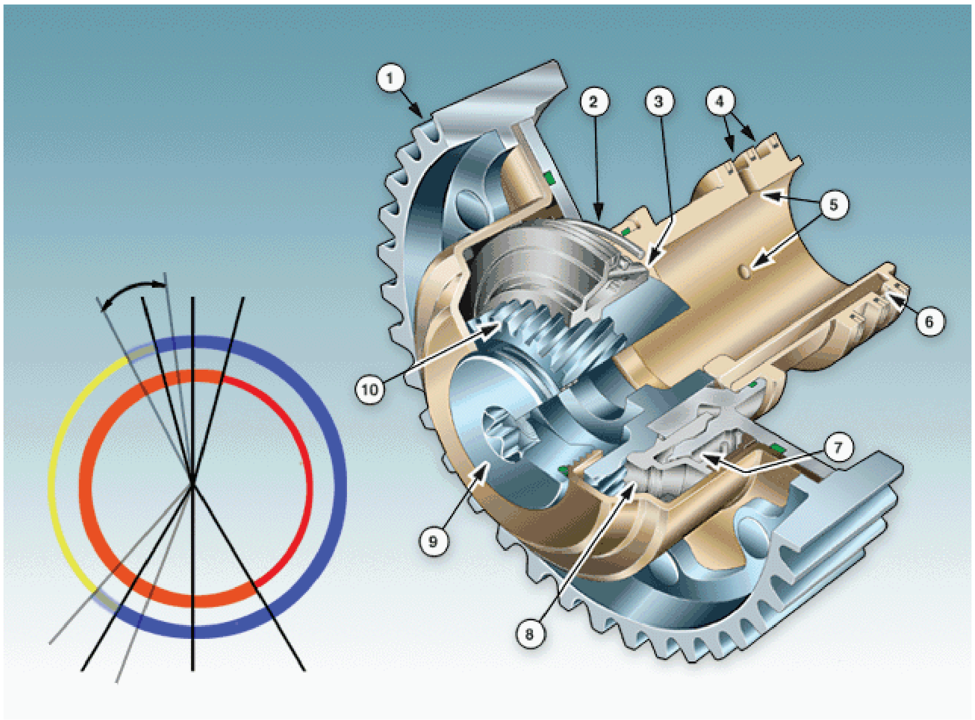


Figure 10.41 Timing is changed under electronic control: 1-Camshaft gear, 2-Spring, 3-Outer helical teeth, 4-Adapter with ring grooves, 5-Oil supply to front chamber via hollow bolt (not shown), 6-Oil supply to rear chamber, 7-Rear chamber, 8-Front chamber, 9-Blanking plug, 1-Inner helical teeth

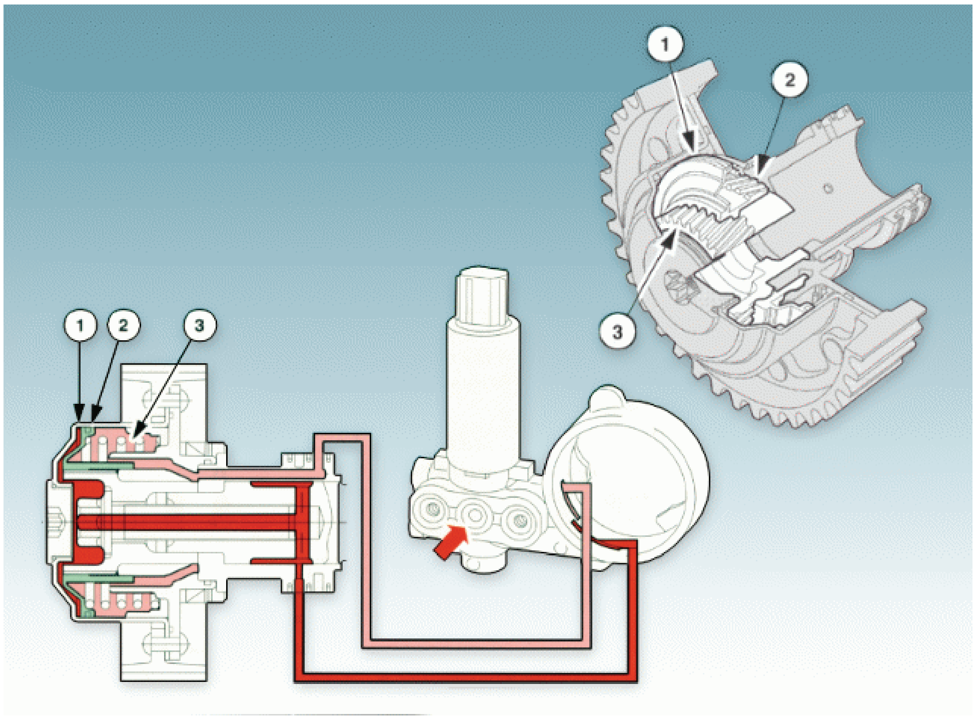


Figure 10.42 Electro hydraulic variable cam control method: 1-Adjusting piston with inner and outer helical teeth, 2-Outer helical teeth connected to camshaft pulley, 3-Inner helical teeth connected to the camshaft

Figure 10.43 Variable cam timing (Source: Ford media)

10.4.3 Lean burn engines

Key fact

Any engine running at a lambda value greater than one is a form of lean burn.

Any engine running at a lambda value greater than one is a form of lean burn. In other words, the combustion takes place with an excess of air. Fuel consumption is improved and CO_2 emissions are lower than with a conventional 'lambda equals one and catalyst system'. However, with the same comparison, NOx emissions are higher due to the excess air factor. Rough running can also be a problem with lean burn (Figure 10.44), due to the problems encountered lighting lean mixtures. A form of charge stratification is a way of improving this.

10.4.4 Two-stroke engines

The two-stroke engine could be an answer to emission problems, but experts have differing views. The main reason for this is that the potential improvements for the four-stroke system have by no means been exhausted. The claimed advantages of the two-stroke engine are lower weight, lower fuel consumption and higher power density. These, however, differ depending on engine design. The major disadvantages are less smooth running, shorter life and higher NOx emissions.

An Australian company, Orbital, have made a considerable contribution to two-stroke technology. A simple shutter control is used in their system and, in a

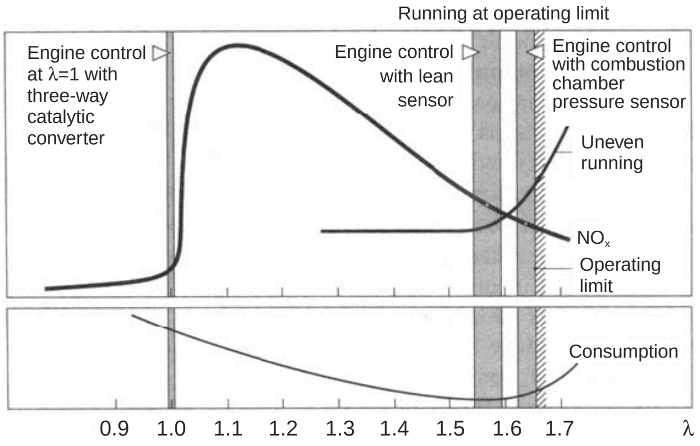


Figure 10.44 Lean burn engine

published paper, a one-litre two-stroke engine was compared with a one-litre four-stroke engine. The two-stroke engine weighs 30% less, has lower consumption and low NOx levels while being comparable in all other ways. The engine can use direct injection to stratify the charge.

However, this technology appears to have stalled because of other technical problems.

10.4.5 Combustion control system

Saab developed a system known as Saab Combustion Control (SCC) to reduce fuel consumption and significantly reduce exhaust emissions. Engine performance is not affected. The key to the operation of the SCC is the use of exhaust gases.

By circulating a significant proportion of exhaust gas back into the combustion process, the fuel consumption can be reduced by up to 10% (Figure 10.45). This technology almost halved the carbon monoxide and hydrocarbon emissions, and cut the nitrogen oxide emissions by 75%.

Figure 10.45 Combustion control spark plug injector (Source: Saab Media)

Unlike standard direct injection systems, the SCC system reaps benefits without disturbing the ideal air-to-fuel ratio (14.7 : 1). This ratio is necessary for a conventional three-way catalytic converter to work properly. The most important aspects of the SCC system are:

- Air-assisted fuel injection with turbulence generator – the injector unit and spark plug are combined into one unit known as the spark plug injector (SPI). Fuel is injected directly into the cylinder with the help of compressed air and another blast of air creates turbulence in the cylinder just before the fuel is ignited. This assists combustion and shortens the combustion time.
- Variable valve timing – variable cams are used so that the SCC system can vary the opening and closing of the inlet and exhaust valves. This allows exhaust gases to be mixed into the combustion air in the cylinder. This is the key aspect that gets the benefits of direct injection while keeping lambda 1 under most operating conditions. The exact recirculation percentage depends on the operating conditions, but up to 70% of the cylinder contents during combustion can consist of exhaust gases.
- Variable spark plug gap with high spark energy – the spark plug gap is variable from 1–3.5 mm. The spark is created from the centre electrode of the SPI to a fixed earth electrode, with a 3.5 mm gap, or to an earth electrode actually on the piston. Very high spark energy (about 80 mJ) is necessary to ignite an air/fuel mixture that is mixed with 70% of exhaust gases.

The best way to understand the SCC process is to start with the expansion or power stroke (the following numbers refer to Figure 10.46).

1. The power stroke operates in the normal way – air/fuel mixture burns, increases the pressure, and this forces the piston down.
2. As the piston reaches the end of the power stroke, the exhaust valves open and most of the exhaust is discharged through the exhaust ports. Remaining exhaust gases are discharging as the piston rises on the exhaust stroke.
3. Fuel is injected into the cylinder via the SPI just before the piston reaches TDC. The inlet valves open at the same time. Exhaust, mixed with fuel, is discharged through both the exhaust and inlet ports.
4. At the start of the inlet stroke, the exhaust and inlet valves open and the mixture of exhaust and fuel is drawn back from the exhaust manifold into the

cylinder. A significant proportion of the exhaust/ fuel mixture now flows up into the inlet ports.

5. As the piston continues to move down, the exhaust valves close but the inlet valves stay open. The exhaust/fuel mixture that flowed into the inlet manifold is now drawn back into the cylinder.
6. As the piston nears BDC, all the exhaust/fuel mixture is drawn back into the cylinder. Towards the end of the inlet stroke, only air is drawn in.
7. As the piston moves upwards during the compression stroke, the inlet valves close and the mixture of exhaust, air and fuel is compressed. About half way up the compression stroke, the SPI delivers a blast of air into the cylinder. This creates turbulence that facilitates combustion and therefore shortens the combustion time.
8. Just before the piston reaches TDC, a spark from the electrode of the SPI ignites the mixture (a) and the next expansion stroke begins (b).

The three-way catalytic converter is still the most important exhaust emission control element. This is because it can catalyse up to 99% of the harmful components in the exhaust gases. However, the catalytic converter has no

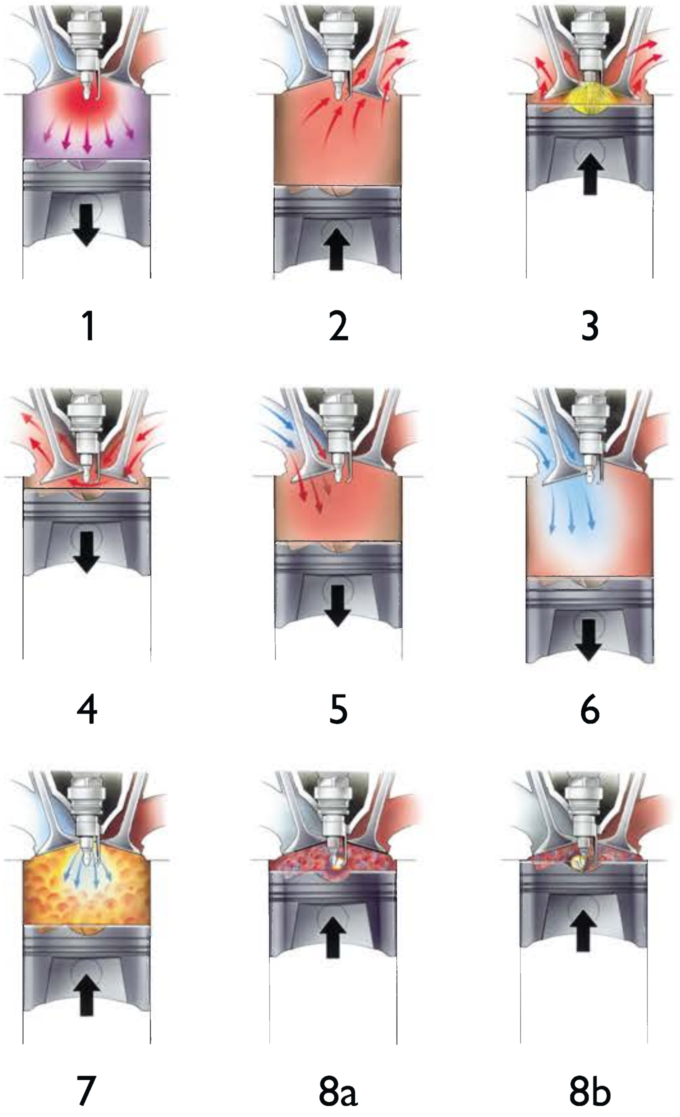


Figure 10.46 Stages of combustion control (Source: Saab Media)

influence on the carbon dioxide (CO_2) emissions, which are directly proportional to the fuel consumption.

Direct injection of petrol is a good way of lowering fuel consumption. Because a precise amount of fuel is injected directly into the cylinder, the consumption can be controlled more accurately. However, only the area around the plug is ignitable because the remainder of the cylinder is filled with air. With standard direct injection systems, this reduces fuel consumption but results in higher nitrogen oxide emissions. The resulting exhaust gases are not ideally suited to a conventional three-way catalytic converter. For this reason, a special catalytic converter with a 'nitrogen oxide trap' has to be used. These are more expensive because they have higher levels of precious metals. In addition, they are more temperature-sensitive and need cooling when under heavy load. This is often achieved by injecting extra fuel. To regenerate the NOx trap when it is 'full up', the engine also has to be run briefly on a richer fuel/air mixture.

The SCC system also contributes towards reducing pumping losses. These usually occur when an engine is running at low load with the throttle almost closed. Under these conditions, the piston in the cylinder operates under a partial vacuum during the induction stroke. The extra energy required for pulling down the piston results in increased fuel consumption. In an SCC engine the cylinder is supplied with just the amount of fuel and air needed at any particular time. The remainder of the cylinder is filled with exhaust gases. This means that the piston does not need to draw in extra air and pumping losses are reduced. The exhaust gases account for 60–70% of the combustion chamber volume, while 29–39% is air; the fuel occupies less than 1%. In general, a higher proportion of exhaust gas is used when the engine is running at low load.

Under low load conditions, the spark is fired from the centre electrode of the plug injector to a fixed earth electrode with a gap of 3.5 mm. Under high load conditions, the spark is fired later (retarded). The gas density in the combustion chamber, under these conditions, is too high for the spark to jump 3.5 mm. A pin on the piston is used instead. The spark will jump to the electrode on the piston when the gap is less than 3.5 mm.

10.4.6 Active cooling

Key fact

An active cooling system uses electronic control to manage and optimize engine temperature.

An active cooling system uses electronic control to manage and optimize engine temperature. The main system components are an electronic valve, an electronically controlled fan and an electrical water pump. Engine temperature is controlled by the efficient management of coolant and air within and around the engine. The advantages of this system are:

- Reduced fuel consumption.
- Lower emissions.
- Reduced engine wear.

Better cabin comfort is achieved by boosting of the heating at lower engine speeds, and heating in the cabin is maintained in cold weather after the engine has been switched off.

Development of Valeo's fully electronically controlled thermal management system, THEMIS, started, in 1995, to work towards satisfying the Euro IV and Euro V emission levels and the Corporate Average Fuel Economy (CAFE) regulation for North America.

Figure 10.47 Pumptronic – electric cooling pump (Source: Valeo)

The complete architecture consists of:

- Pumptronic® or electronic water pump. This system uses brushless motor technology, wet-rotor and rare earth magnets. This results in a global efficiency of over 55%.
- Fantronic® or continuous variable speed fan system. This uses an embedded pulse width modulation driver in the motor, which is cooled by the fan blades themselves.
- Multi-way proportional electronic valve.
- Engine temperature sensor.
- Electronic control unit.
- Optimized heat exchangers (coolant radiators and heater cores).

In addition to improved fuel efficiency, reduced emission levels, enhanced cabin comfort and improved engine reliability, it is possible to have fail-safe modes, self-diagnosis options and servicing diagnosis. Fuel consumption and emissions were tested according to the European and US test cycles in laboratory conditions.

Coolant does not flow during warm up, to allow the engine to heat up quickly; this limits thermal losses. Emissions of HC decrease by 10% and CO by 0–20% during the test cycles. NOx remains unchanged. A higher coolant temperature (110/115 °C vs. 95 °C) is possible on low and medium loads. This results in more efficient combustion, a 2–5% fuel economy and proportionate reduction of CO_2 emissions. The following benefits are also evident:

- Boosting water flow in cold weather provides 30 minutes of heating even after the engine has stopped.
- When cabin heating is not required, there is no water flow in the heater core to optimize climate control systems.
- Knocking and local boiling in the cylinder head are reduced. At high engine load, the ECU lowers engine temperature to an average 90 °C for maximum performance.
- No thermal shocks or heat peak when the engine stops. The electric water pump boosts water flow to ensure a steady reduction of temperature when necessary.
- Potential trouble can be anticipated. In the case of a rapid rise of temperature, the controller boosts the water flow and/or the fan system.

The system tunes and controls the operation of the various components continuously. If one component is not working correctly, the system can

Figure 10.48 Fantronic – electrically operated cooling fan (Source: Valeo)

Figure 10.49 Electronic control valve (Source: Valeo)

compensate by over-boosting another component. This is known as a fail-safe mode and the driver is informed via a warning light. Overall, this active cooling system also reduces the power consumed by the water pump. The first applications are expected in 2005.

Key fact

If one component of the active cooling system is not working correctly, it can compensate by over-boosting another component.

10.4.7 Engine trends – spark ignition

In Europe during 2003, vehicles with compression ignition (CI) engines started to outsell the spark ignition (SI) versions. However, because of this competition, as well as that from alternative fuel vehicles, engineers are making more developments to the SI engine. More power, reduced consumption and emissions, together with more efficient packaging are the key challenges being met. Some of the innovations under development and/or in use are considered briefly here.

Key fact

In Europe during 2003, vehicles with compression ignition (CI) engines started to outsell the spark ignition (SI) versions.

Variable compression ratios

A higher compression ratio results in greater thermal efficiency. However, it also makes the engine run hotter and the components are under greater stress. Being able to vary the compression ratio to achieve improvements under certain speed and load conditions is an innovative approach. Saab has done considerable work in this area.

Electromechanical valve train

Full control of valve operation means engine management can take greater control. However, operating valves independently is difficult – so the camshaft will be with us for some time yet. Lotus engineers have made significant advancements using hydraulic operating mechanisms.

High efficiency superchargers

New developments in supercharging mean that the charger itself takes less energy from the engine. Of particular interest are electrically driven superchargers because they allow full electronic control.

Cylinder deactivation

This technique has been tried on and off for a number of years. The capacity of, say, a 3-litre V8 is reduced when used around town, with the consequent reduction in consumption and emissions. GM uses this system on their XV8 engine. It is called displacement on demand. It is also used on F1 engines.

High pressure direct injection

Gasoline direct injection is now commonplace. However, work is on-going to increase the fuel pressure, as this results in more possibilities for controlling the cylinder charge.

Reduced-current draw-fuel pumps

A simple but effective technique, which can result in lower emissions and consumption, is to reduce the electrical current consumed. A fuel pump has been developed by Visteon, which can increase fuel economy by up to 0.2 mpg.

Intelligent valve control

Honda have produced an engine for the RSX that uses intelligent valve control. The valve lift and phase can be controlled electronically. The result is impressive economy and low emissions.

Gas-by-wire

This concept has been in use by BMW for some time. The idea is that the driver's instructions, via the throttle pedal, are interpreted and the throttle is moved to achieve optimum performance. For example, for full acceleration the driver ‘floors’ the pedal – which opens the throttle fully on a traditional system – but opens the throttle more progressively on a gas-by-wire system.

Air-assisted direct fuel injection

One important aspect of direct fuel injection is that the charge in the cylinder can be stratified. In other words, the region around the plug is at the ideal ratio, but a large part of the cylinder is then made up of air or, better, recirculated exhaust gases. Ford now have an engine that can run as lean as 60 : 1.

‘W’ engine configuration

An interesting cylinder configuration, quite appropriately developed by VW, is the ‘double V’ or ‘W’ concept. This allows a W12 engine to be as compact as a V8. The result is very smooth operation and a relatively low mass which, as with any reduction in mass, improves efficiency.

10.4.8 Transonic combustion

On average only about 15% of the energy from the fuel in the tank is converted into movement of the car according to the U.S. Department of Transportation: Transportation Research Board. The rest of the energy is lost to engine and driveline inefficiencies and idling (see Figure 10.50). The engine is where most thermal efficiency loss takes place. Combustion irreversibility results in large amounts of waste heat escaping through the cylinder walls and unrecoverable exhaust energy. Normal engines run with rich air-to-fuel ratios, which also result in fuel being trapped in the crevice as well as partially combusting near the cylinder walls. This energy loss is the core of automotive inefficiency.

Key fact

On average only about 15% of the energy from the fuel in the tank is converted into movement of the car.

Transonic combustion (TSCi™) is a new combustion process for the petrol/gasoline internal combustion engine. The new combustion process utilizes direct injection of fuel into the cylinder as a supercritical fluid based on the patented concept of injection-ignition. Supercritical fuel promotes rapid mixing with the contents of the cylinder which, after a short delay, results in spontaneous ignition at multiple locations. Multiple ignition sites and rapid combustion combine to result in optimum heat release and high cycle efficiency. Other advantages are the elimination of droplet burning and increased combustion stability that results from multiple ignition sources.

Key fact

The transonic combustion process utilizes direct injection of fuel into the cylinder as a supercritical fluid.

Transonic combustion brings together the injection and ignition processes to become Injection-Ignition. The characteristics of the system address all of the

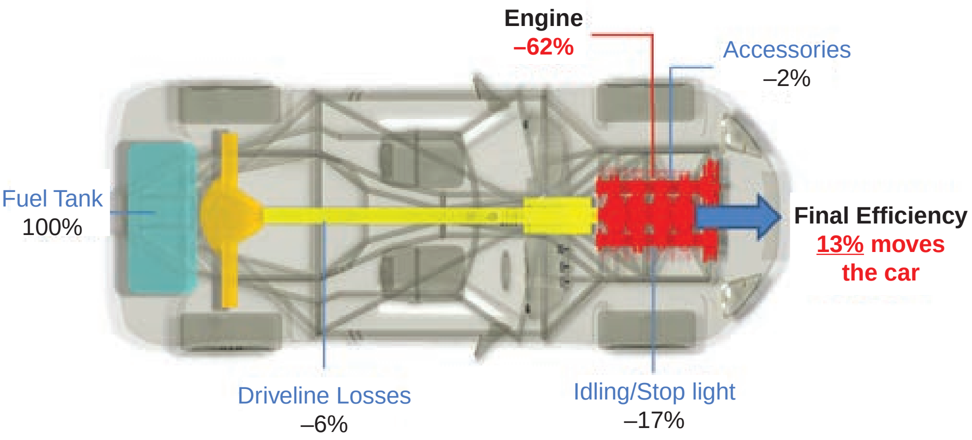


Figure 10.50 Losses during urban driving a petrol/gasoline car (Source: US DOT: Transport Research Board)

Figure 10.51 Transonic combustion process (Source: Transonic Combustion, Inc.)

issues identified above as limiting the efficiency of the gasoline engine; it is capable of operating over a wide range of air/fuel ratios and so does not require a throttle for load control. TSCi™ has inherently short combustion delay and rapid combustion that result in heat release phasing for optimal efficiency. It can be operated at an optimal compression ratio since it is not dependent on high octane petrol/gasoline. The ignition mechanism renders the combustion system fuel neutral in the sense it is not reliant on either octane or cetane values.

A characteristic of the transonic combustion process is that injection of the fuel is delayed to the extent that the heat release predominantly takes place after TDC of the engine power stroke. In order to achieve this, the combustion process must have rapid air-fuel mixing, followed by a short delay period and combustion. Such characteristics can be achieved by injecting the fuel in the form of a supercritical fluid.

Definition

Transonic: Moving from sub-sonic to super-sonic.

Definition

Supercritical fluid: Any substance at a temperature and pressure above its critical point; it is not a solid, liquid or a gas.

Independent engine test results have been attained over a range of speed/load operating conditions to show fuel consumption reductions of 20% at critical engine test conditions. The results have been correlated with thermo-dynamic modelling to provide a detailed understanding and correlation of the fundamental thermodynamics.

The potential of the technology is to provide real world fuel consumption reductions of 20% to 25% in petrol/gasoline fuelled passenger cars with corresponding reductions of greenhouses gases.

The TSCi™ system comprises fuel injectors, fuel pump, fuel conditioning module and engine control algorithms. The technology is optimized for high compression ratio engines.

Visit: www.tscombustion.com for more details.

10.4.9 Formula 1 engine technology

In 2010 the FIA's World Motor Sport Council approved new F1 engine regulations. This will result in the demise of the current 2.4 litre V8 engines. They will be replaced by four cylinder 1.6 litre engines from 2013. It is expected that these environmentally-friendly units will result in a 35% reduction in fuel consumption. However, the same level of performance as the V8s will remain. Kinetic energy recovery systems (KERS) and additional energy management will be used to achieve this.

High pressure fuel injection, up to 500 bar, will be used and the rev limits will be reduced from the current 18 000 to a maximum of 12 000 rpm.

Hybrid F1 engines then should be good and, of course, the technical 'spin-offs' to normal automotive systems will be interesting. Watch this space.

Safety first

Caution/Achtung/Attention – Burning fuel and high voltages can seriously damage your health!

10.4.10 Diagnosing engine management systems

Table 10.1 is based on information available from 'Autodata' in its excellent range of books and data sources. It relates in particular to a specific fuel system but it is also a good guide to many others. The numbers relate to the order in which the systems should be checked.

Table 10.1 Engine management symptoms and faults

Symptom	Possible fault
Engine will not start	Engine and battery earth connections Fuel filter and fuel pump Air intake system for leaks Fuses/fuel pump/system relays Fuel injection system wiring and connections Coolant temperature sensor Auxiliary air valve/idle speed control valve Fuel pressure regulator and delivery rate ECU and connector Limp home function – if fitted
Engine difficult to start when cold	Engine and battery earth connections Fuel injection system wiring and connections Fuses/fuel pump/system relays Fuel filter and fuel pump Air intake system for leaks Coolant temperature sensor Auxiliary air valve/idle speed control valve Fuel pressure regulator and delivery rate ECU and connector Limp home function – if fitted
Engine difficult to start when warm	Engine and battery earth connections Fuses/fuel pump/system relays Fuel filter and fuel pump Air intake system for leaks Coolant temperature sensor Fuel injection system wiring and connections Air mass meter Fuel pressure regulator and delivery rate Air sensor filter ECU and connector Knock control – if fitted
Engine starts then stops	Engine and battery earth connections Fuel filter and fuel pump Air intake system for leaks Fuses/fuel pump/system relays Idle speed and CO content Throttle potentiometer Coolant temperature sensor Fuel injection system wiring and connections ECU and connector Limp home function – if fitted

(Continued)

Table 10.1 *(Continued)*

Symptom	Possible fault
Erratic idling speed	Engine and battery earth connections Air intake system for leaks Auxiliary air valve/idle speed control valve Idle speed and CO content Fuel injection system wiring and connections Coolant temperature sensor Knock control – if fitted Air mass meter Fuel pressure regulator and delivery rate ECU and connector Limp home function – if fitted
Incorrect idle speed	Air intake system for leaks Vacuum hoses for leaks Auxiliary air valve/idle speed control valve Idle speed and CO content Coolant temperature sensor
Misfire at idle speed	Engine and battery earth connections Air intake system for leaks Fuel injection system wiring and connections Coolant temperature sensor Fuel pressure regulator and delivery rate Air mass meter Fuses/fuel pump/system relays
Misfire at constant speed	Air flow sensor
Hesitation when accelerating	Engine and battery earth connections Air intake system for leaks Fuel injection system wiring and connections Vacuum hoses for leaks Coolant temperature sensor Fuel pressure regulator and delivery rate Air mass meter ECU and connector Limp home function – if fitted
Hesitation at constant speed	Engine and battery earth connections Throttle linkage Vacuum hoses for leaks Auxiliary air valve/idle speed control valve Fuel lines for blockage Fuel filter and fuel pump Injector valves ECU and connector Limp home function – if fitted
Hesitation on overrun	Air intake system for leaks Fuel injection system wiring and connections Coolant temperature sensor Throttle potentiometer Fuses/fuel pump/system relays Air sensor filter Injector valves Air mass meter

(Continued)

Table 10.1 *(Continued)*

Symptom	Possible fault
Knock during acceleration	Knock control – if fitted Fuel injection system wiring and connections Air mass meter ECU and connector
Poor engine response	Engine and battery earth connections Air intake system for leaks Fuel injection system wiring and connections Throttle linkage Coolant temperature sensor Fuel pressure regulator and delivery rate Air mass meter ECU and connector Limp home function – if fitted
Excessive fuel consumption	Engine and battery earth connections Idle speed and CO content Throttle potentiometer Throttle valve/housing/sticking/initial position Fuel pressure regulator and delivery rate Coolant temperature sensor Air mass meter Limp home function – if fitted
CO level too high	Limp home function – if fitted ECU and connector Emission control and EGR valve – if fitted Fuel injection system wiring and connections Air intake system for leaks Coolant temperature sensor Fuel pressure regulator and delivery rate
CO level too low	Engine and battery earth connections Air intake system for leaks Idle speed and CO content Coolant temperature sensor Fuel injection system wiring and connections Injector valves ECU and connector Limp home function – if fitted Air mass meter Fuel pressure regulator and delivery rate
Poor performance	Engine and battery earth connections Air intake system for leaks Throttle valve/housing/sticking/initial position Fuel injection system wiring and connections Coolant temperature sensor Fuel pressure regulator/fuel pressure and delivery rate Air mass meter ECU and connector Limp home function – if fitted

The following procedure is very generic but with a little adaptation can be applied to any fuel injection system. Refer to the manufacturer's recommendations if in any doubt.

1. Check battery state of charge (at least 70%).
2. Hand and eye checks (all fuel and electrical connections secure and clean).
3. Check for spark at plug lead (if poor or no spark jump to stage 15).
4. Check fuel pressure supplied to rail (for multipoint systems it will be about 2.5 bar but check specifications).
5. If the pressure is NOT correct jump to stage 11.
6. Is injector operation OK? – continue if NOT (suitable spray pattern or dwell reading across injector supply).
7. Check supply circuits from main relay (battery volts minimum).
8. Continuity of injector wiring (0–0.2 Ω and note that many injectors are connected in parallel).
9. Sensor readings and continuity of wiring (0–0.2 Ω for the wiring sensors will vary with type).
10. If no fuel is being injected and all tests so far are OK (suspect ECU).
11. Fuel supply – from stage 5.
12. Supply voltage to pump (within 0.5 V of battery – pump fault if supply is OK).
13. Check pump relay and circuit (note that, in most cases, the ECU closes the relay but this may be bypassed on cranking).
14. Ensure all connections (electrical and fuel) are remade correctly.
15. Ignition section (if appropriate).
16. Check supply to ignition coil (within 0.5 V of battery).
17. Spark from coil via known good HT lead (jumps about 10 mm, but do not try more).
18. If good spark then check HT system for tracking and open circuits. Check plug condition (leads should be a maximum resistance of about 30 k/m per lead) – stop here in this procedure.
19. If no spark, or it will only jump a short distance, continue with this procedure (colour of spark is not relevant).
20. Check continuity of coil windings (primary 0.5–3 Ω, secondary several kΩ).
21. Supply and earth to 'module' (12 V minimum supply, earth drop 0.5 V maximum).
22. Supply to pulse generator if appropriate (10–12 V).
23. Output of pulse generator (inductive about 1 V AC when cranking, Hall-type switches 0–8 V DC).
24. Continuity of LT wires (0–0.1 Ω).
25. Suspect ECU but only if all of the above tests are satisfactory.

10.5 Advanced engine management technology

10.5.1 Speed density and fuel calculations

Engine management systems that do not use an air flow sensor rely on the speed-density method for determining the required fuel quantity. Accurate measurement of the manifold absolute pressure (MAP) and intake air temperature is essential with this technique.

The volume flow rate of air taken in to an engine at a given speed can be calculated:

$$A_v = \left[\left(\frac{RPM}{60}\right)\left(\frac{D}{2}\right)V_e\right] - EGR_v$$

where: A_v = air volume flow rate (litres/s), EGR_v = exhaust gas recirculation volume (litres/s), D = displacement of the engine (litres), V_e = volumetric efficiency (as a % from look up tables).

The density of air in the inlet manifold is related to its temperature and pressure. If these are measured accurately then density can be calculated. A basic gas law states that in a fixed volume:

$$d_a = d_0\left(\frac{p_i}{p_0} \times \frac{T_0}{T_i}\right)$$

where: d_a = density, p_i = intake pressure, T_i = intake temperature, p_0, d_0 and T_0 are known values relating to pressure, density and temperature under 'sea level standard day' (SLSD) conditions.

The mass of the air can be calculated:

$$M_a = d_a \times V$$

where: M_a = mass of air (kg), d_a = density of the air (kg/litre), V = volume of air (litres).

The mass flow rate can now be calculated:

$$A_m = d_a \times A_v$$

where: A_m = air mass flow rate (kg/s).

Finally, by substitution and simplification, air mass flow can be calculated by:

$$A_m = d_a\left[\left(\frac{RPM.D.V_e}{120}\right) - EGR_v\right]$$

Further to this calculation, the basic fuel quantity can be determined as follows:

$$F = \frac{A_m}{AFR}$$

where: F = fuel quantity (kg), AFR = desired air/fuel ratio.

To inject the required quantity of fuel, the final calculation is that of injector pulse width:

$$T = \frac{F}{R_f}$$

where: T = time, R_f = fuel injector(s) delivery rate.

Note that the actual injection period will also depend on a number of other factors such as temperature and throttle position. The total fuel quantity may also be injected in two halves.

10.5.2 Ignition timing calculation

Data relating to the ideal ignition timing for a particular engine are collected from dynamometer tests and operational tests in the vehicle. These data are stored in the form of look-up tables in ROM. These look-up tables hold data relative to the speed and load of the engine. The number of look-up values is determined by the computing power of the microcontroller, in other words the number of bits, as this determines the size of memory that can be addressed.

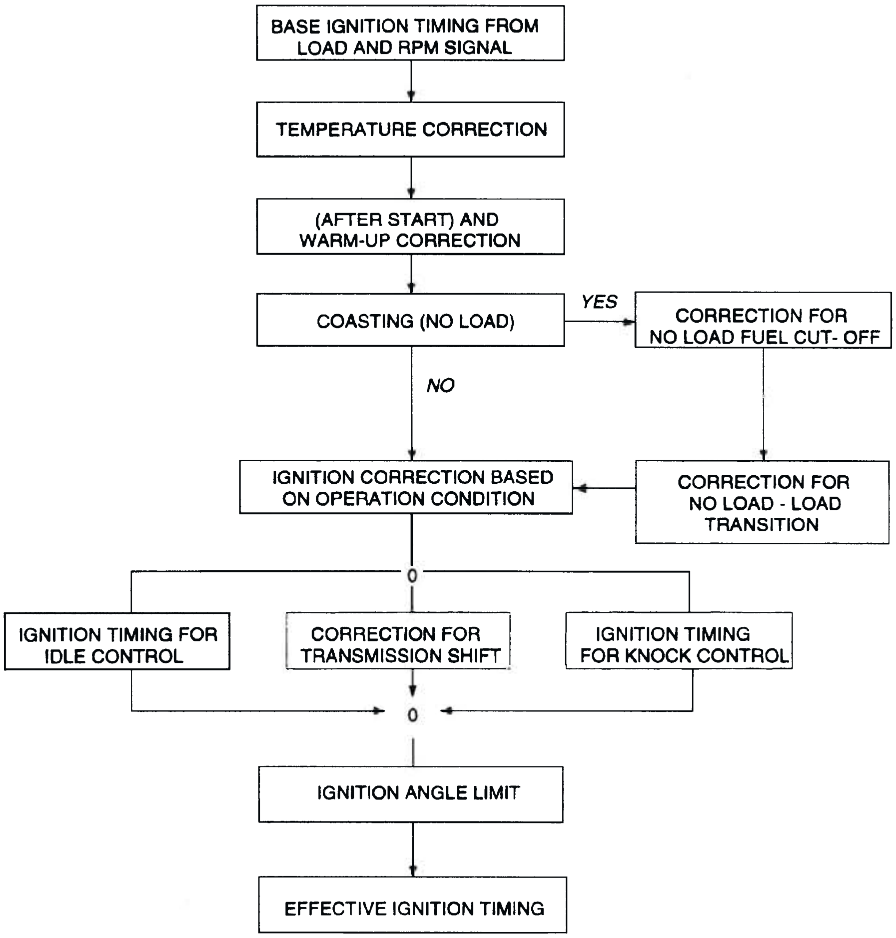


Figure 10.52 Determination of effective ignition timing

Inputs from speed and load sensors are converted to digital numbers and these form the reference to find the ideal timing value. A value can also be looked up for the temperature correction. These two digital numbers are now added to give a final figure. Further corrections can be added in this way for conditions such as overrun and even barometric pressure if required.

This ‘timing number’ is used to set the point at which the coil is switched off; that is, the actual ignition point. The ECU receives a timing pulse from the ‘missing flywheel tooth’ and starts a ‘down counter’. The coil is fired (switched off) when the counter reaches the ‘timing number’. The computing of the actual ‘timing number’ is represented by Figure 10.52.

To prevent engine damage caused by detonation or combustion knock, but still allow the timing to be set as far advanced as possible, a knock sensor is used. The knock sensor (accelerometer) detects the onset of combustion knock, but the detection process only takes place in a ‘knock window’. This window is just a few degrees of crankshaft rotation either side of top dead centre compression for each cylinder. This window is the only time knock can occur and is also a quiet time as far as valve opening and closing is concerned. The sensor is tuned to respond to a particular frequency range of about 5–10 kHz, which also helps to eliminate erroneous signals. The resonant frequency of this type of accelerometer is greater than about 25 kHz.

The signal from the knock sensor is filtered and integrated in the ECU. A detection circuit determines a yes/no answer to whether the engine knocked

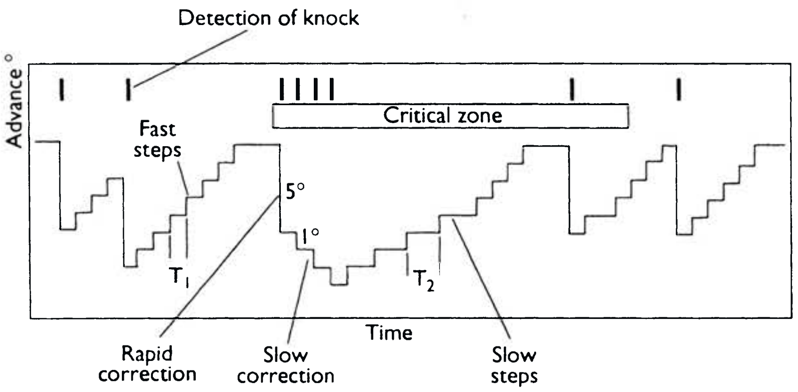


Figure 10.53 How timing is varied in response to combustion knock

or not. When knock is detected on a particular cylinder, the timing for that cylinder is retarded by a set figure, often 2°, each time the cylinder fires, until the knocking stops. The timing is then advanced more slowly back towards the look-up value. Figure 10.53 represents this process in more detail.

10.5.3 Dwell calculation

In order for an ignition system to produce constant energy the dwell angle must increase as the engine speed increases. Ideal dwell values are held in a look-up table; engine speed forms one axis and battery voltage the other. If battery voltage falls, the dwell angle is increased to compensate. The 'dwell number' is used in a similar way to the 'timing number' in the previous section except that this time, the 'dwell number' is used to determine the switch-on point of the coil during operation of the down counter.

10.5.4 Injection duration calculation

The main criteria for the quantity of fuel required for injection are engine speed and load. Further corrections are then added. Figure 10.54 represents the process carried out in a digital electronic control unit to calculate injection duration. The process of injection duration calculation is summarized as follows.

- A basic open period for the injectors is determined from the ROM information relating to engine speed and load.
- Corrections for air and engine temperature.
- Idling, full or partial load corrections.
- After-start enrichment.
- Operational enrichment.
- Acceleration enrichment.
- Weakening on deceleration.
- Cut-off on over-run.
- Reinstatement of injection after cut-off.
- Correction for battery voltage variation.

Under starting conditions the injection period is calculated differently. This is determined from a set figure varied as a function of temperature.

Figure 10.55 is a flow diagram to represent to overall process of calculating fuel and ignition settings.

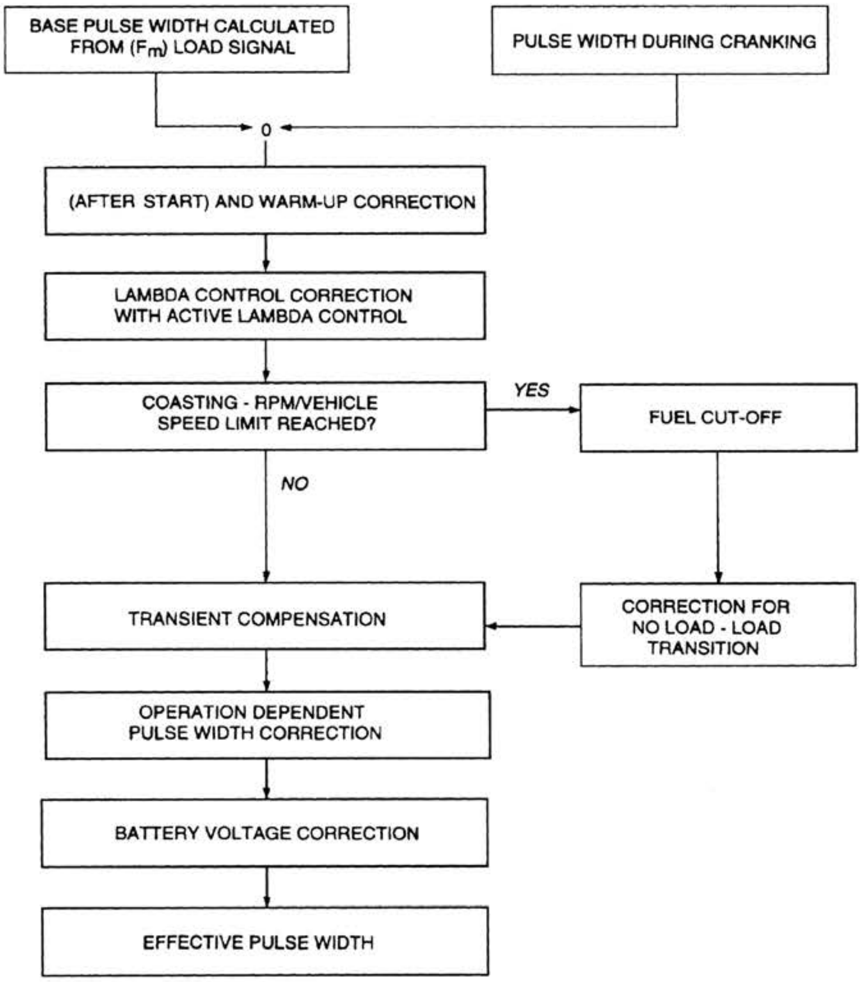


Figure 10.54 Determination of effective injector pulse width

10.5.5 Developing and testing software

There is, of course, more than one way of producing a 'computer' program. Most programs used in the electronic control unit of a vehicle digital control system are specialist applications and, as such, are one-off creations. The method used to create the final program is known as the 'top down structured programming technique'. Following on from a 'need' for the final product, the process can be seen to pass through six definable stages.

1. Requirement analysis seeks to answer the question as to whether a computerized approach is the best solution. It is, in effect, a feasibility study.
2. Task definition is a process of deciding exactly what the software will perform. The outcome of this stage will be a set of functional specifications.
3. Program design becomes more important as the complexity of the task increases. This is because, where possible, it is recommended that the program be split into a number of much smaller tasks, each with its own detailed specification.
4. Coding is the stage at which the task begins to be represented by a computer language. This is when the task becomes more difficult to follow as the language now used is to be understood by the 'computer'.
5. Debugging and validation is the process of correcting any errors or a bug in the program code and then finally ensuring that it is valid. This means

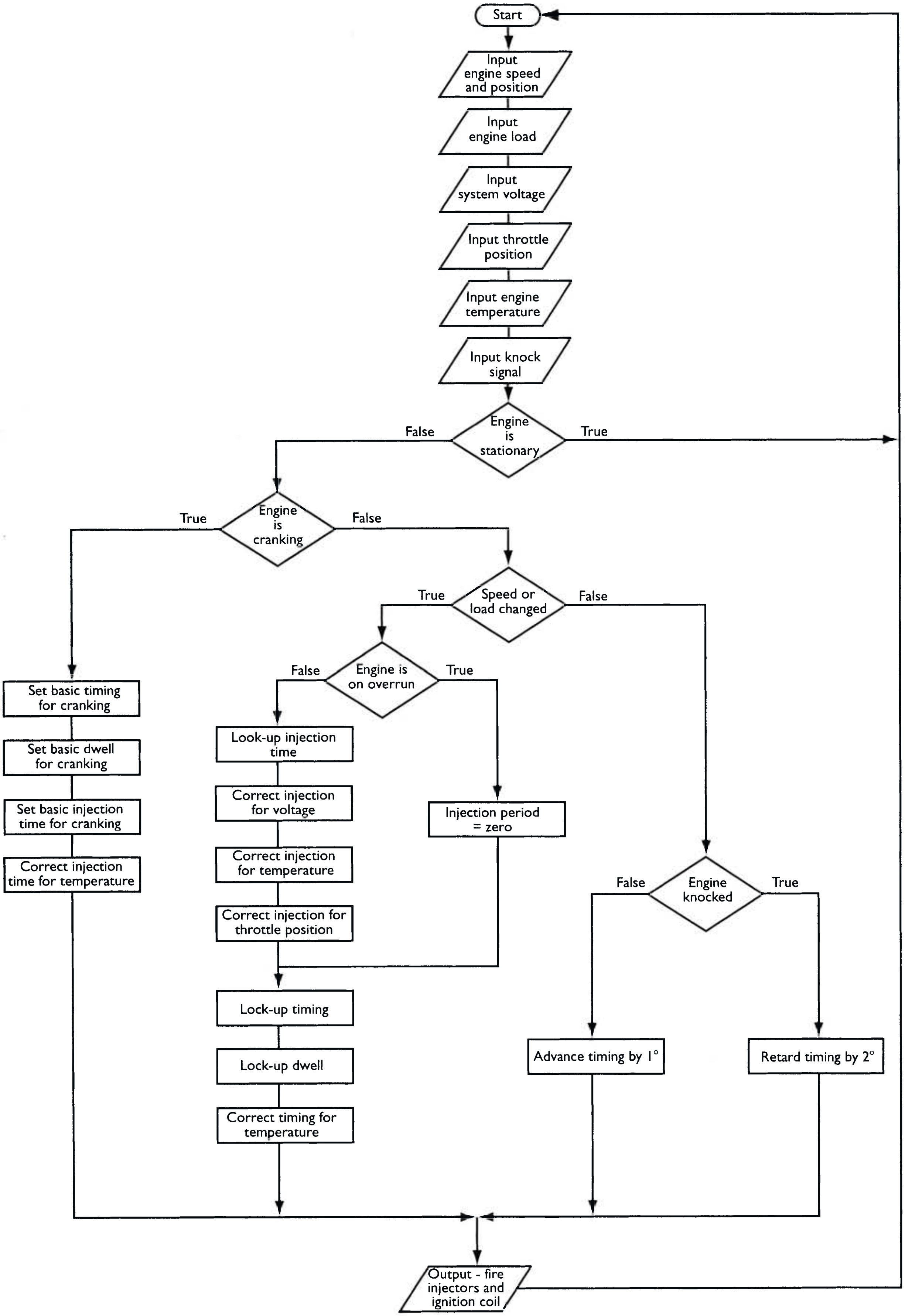


Figure 10.55 Engine management and fuel ignition calculation flow diagram

checking that the desired outputs appear in response to appropriate inputs. In other words, does it work? (As a slight aside, did you know that the original computer bug was actually a moth trapped between the contacts of a relay?) Note that it is very important to get the program right at this stage as it is likely to be incorporated into tens of thousands of specially produced microcontrollers. A serious error can be very expensive to rectify.

6. Operation and maintenance is the stage when the program is actually in use. Occasionally slight errors do not come to light until this stage, such as a slight hesitation during acceleration at high altitude or some other obscure problem. These can be rectified by program maintenance for inclusion in later models.

This section has been included with the intention of filling in the broader picture of what is involved in producing a program for, say, an electronic spark advance system. Many good books are available for further reading on this subject.

10.5.6 Simulation program

Automotive Technology (AT) is a training and diagnostic software program. It works in conjunction with this textbook and on-line learning. All complex electronically controlled systems can be considered as having:

INPUTS – CONTROL – OUTPUTS

The main 'AT' program works in the same way but also incorporates diagnostics. In other words, it will help you learn how complex systems work and how to diagnose faults with them. The program concentrates on engine management, starting and charging. A MultiScope program is included that allows actual tests to be carried out and the results viewed on a scope or a multimeter. The software is fully functional but runs out of fuel! It should be registered if you continue to use it to prevent the tank leaking …

The program allows you to 'drive' the vehicle or directly change inputs to systems such as engine management. The computer (just like the computer in a vehicle), will calculate the outputs of the system. Engine management is the main area covered but other systems are available for use. The system can be set to provide telemetry to the MultiScope as the car is driven round the Silverstone circuit!

The diagnostics part of the program is designed to assist with diagnosing faults in automotive systems. It is ideal to help with the development of diagnostic skills.

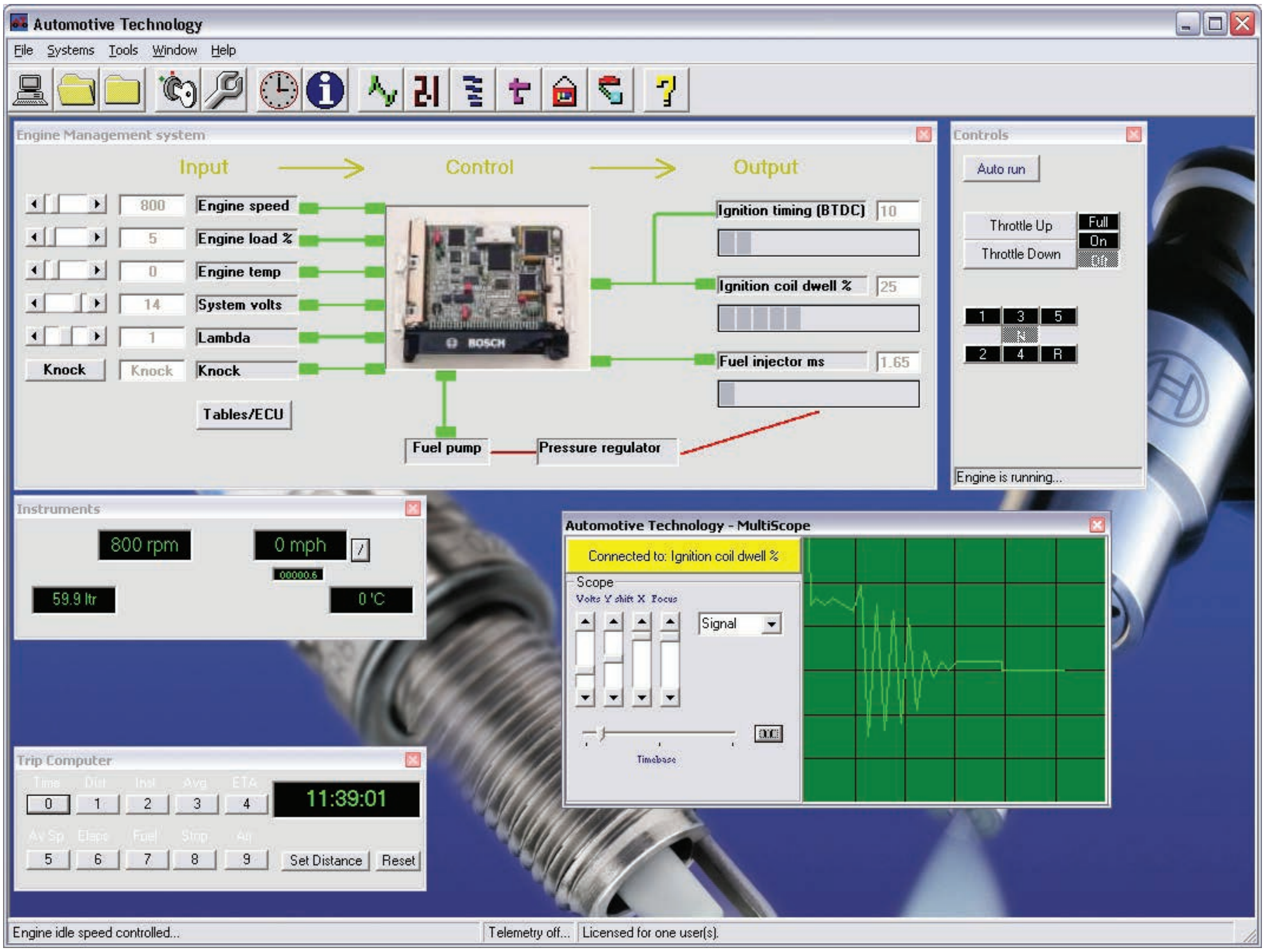
The comprehensive diagnostic routines are part of the program. These can also be printed for use in the workshop. A step by step process helps you track down any fault. The MultiScope program is used to test the operation of sensors and actuators. Faults can be set to allow practice of diagnostic techniques.

WWW

The program (as well as many other useful resources) can be downloaded from: **www.automotive-technology.co.uk**

10.5.7 Hot chipping

Hot chipping is the name often given to the fitting of new processors/memory to improve the performance of a vehicle. It should be noted that the improvements are at the expense of economy, emissions and engine life! Fitting a 'Power Processor', which is a programmable computer specifically designed for high performance engines, is the first step. The fuel map, engine ignition timing map, acceleration fuel and all parameters for fuel management are programmable



Engine speed 800

Engine load % 5

Engine temp 0

System volts 14

Lambda 1

Knock

Ignition timing (BTDC) 10

Ignition coil dwell % 25

Fuel injector ms 1.65

800 rpm

0 mph

000006

59.9 ltr

0 'C

11:39:01

Connected to: Ignition coil dwell %

Engine is running...

Engine idle speed controlled...

Telemetry off...

Licensed for one user(s).

Figure 10.56 AT simulation program

using a PC or laptop computer. Note that a new ECU is needed in most cases but this does allow improvement of other features.

The software even allows changes to be made while you are driving the vehicle. This system is appropriate for virtually any fuel injected engine. A basic calibration is used to get the engine started and running. The user then performs fine tuning.

The systems are capable of closed or open loop operation. Some systems even feature control of nitrous injection with automatic engagement based on throttle position and rpm. Ignition timing is automatically retarded with pre-set parameters.

Tuning a fuel-injected engine requires experience, time and patience. One mistake with the laptop keyboard and an engine can easily be turned into a pile of junk from detonation or a lean condition!

When determining the size of the base fuel map's rpm resolution, the cell widths should be as small as possible. This gives the most tuning set points in the operating range of the engine. If the map is configured to 5000 rpm, any resolution above that figure would be lost, but resolution would be gained where the engine spends its most time, i.e. below 5000 rpm.

If the fuel map is calibrated to 5000 rpm and the calibrated pulse width at that speed is 12 ms, the ECU will keep issuing pulses of 12 ms at any speed above this value. It is beneficial to use as many of the 256 (16 16 look-up table or 28 relating to 8 bits) set points as possible during tuning. This is established by setting the rpm between cells. The largest fuel commands should be at the peak torque and, as the engine speed escalates above peak torque, the pulse width reduces. Most values from the ECU's inputs and outputs will be available 'on-screen', as if from the serial data link on a production ECU.

Most systems use 'interpolative' software, meaning the cells surrounding the actual chosen cell in the fuel map will affect the issued pulse width. Getting the fuel calculations as near to the stoichiometric set point as possible and using very little, if any, oxygen sensor trim is a good technique. This is the approach that the original equipment manufacturers use. While working on the base fuel map, note that with injector pulse widths below 2 ms, you are entering an unstable range. Work with all of the cells around the chosen idle cell because the surrounding cell values are used for interpolation. Large variations in matrix values around the idle cell can lead to surging.

The resolution of the ignition map is referenced from the fuel table and is scaled at a rate of 1.5 to the fuel table. The same theory applies to the spark table, as to the fuel table, in regard to keeping the same timing command beyond its rpm resolution. The amount of retardation required to stop detonation once it is started in the combustion chamber is greater than the amount that would be needed never to allow detonation to start. A trial and error method is required for the best results. The amount of spark advance is affected by engine criteria such as:

- Cylinder-head combustion chamber design.
- Mixture movement.
- Piston design.
- Intake manifold length and material.
- Compression ratio.
- Available fuel.
- Thermal transfer from the cylinder-head to the cooling system.

More information: http://accel-ignition.com

10.5.8 Artificial Intelligence

Artificial intelligence (AI) is the ability of an artificial mechanism to exhibit intelligent behaviour. The term invites speculation about what constitutes the mind or intelligence. Such questions can be considered separately, but the endeavour to construct and understand increasingly sophisticated mechanisms continues.

AI has shown great promise in the area of expert systems, or knowledge-based expert programs, which, although powerful when answering questions within a specific domain, are nevertheless incapable of any type of adaptable, or truly intelligent, reasoning.

No generally accepted theories have yet emerged within the field of AI, due in part to AI being a new science. However, it is assumed that on the highest level, an AI system must receive input from its environment, determine an action or response and deliver an output to its environment. This requires techniques of expert reasoning, common sense reasoning, problem solving, planning, signal interpretation and learning. Finally, the system must construct a response that will be effective in its environment.

The possibilities for AI in vehicle use are unlimited. In fact, it becomes more a question of how much control the driver would be willing to hand over to the car. If, for example, the vehicle radar detects that you tend to follow the car in front too closely, should it cause the brakes to be applied? The answer would probably be no, but if the question was, as the engine seems to surge at idle should the idle speed be increased slightly, then the answer would most likely be yes.

It is not just the taking in of information and then applying a response as this is carried out by all electronic systems to some extent, but in being able to adapt and change. For example, if the engine was noticed to surge when the idle speed was set to 600 rpm, then the ECU would increase the speed to, say, 700 rpm. The adaptability, or a very simple form of AI, comes in deciding to set the idle speed at 700 rpm on future occasions. This principle of modifying the response is the key. Many systems use a variation of this idea to control idle speed and also to adapt air–fuel ratios in response to a lambda sensor signal.

An adaptive ignition system has the ability to adapt the ignition point to the prevailing conditions. Programmed ignition has precise values stored in the memory appropriate for a particular engine. However, due to manufacturing tolerances, engine wear with age and road conditions means that the ideal timing does not always correspond to that held in the ECU memory.

The adaptive ignition ECU has a three-dimensional memory map as normal for looking up the basic timing setting, but it also has the ability to alter the spark timing rapidly, either retarding or advancing, and to assess the effect this has on engine torque. The ECU monitors engine speed by the crankshaft sensor, and if it sees an increase in speed after a timing alteration, it can assume better combustion. If this is the case, the appropriate speed load site on the memory map is updated. The increase in speed detected is for one cylinder at a time; therefore, normal engine speed changes due to the throttle operation do not affect the setting.

The operation of the adaptive ignition system is such as to try and achieve a certain slope on the timing versus torque curve as shown by Figure 10.57. Often the slope is zero (point A) for maximum economy but is sometimes non-zero (point B), to avoid detonation and reduce emissions.

Figure 10.58 shows the adaptive ignition block diagram. The fixed spark timing map produces a 'non-adapted' timing setting. A variation is then added or

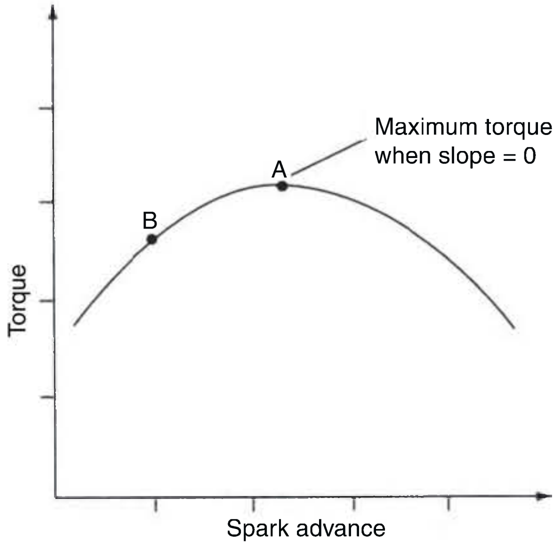


Figure 10.57 Timing versus torque curve

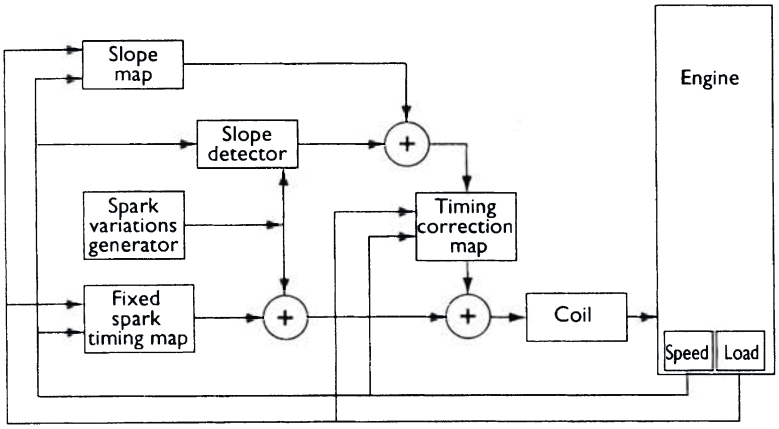


Figure 10.58 Adaptive ignition block diagram

subtracted from this point and the variation is also sent to the slope detector. The slope detector determines whether the engine torque was increased or decreased from the measure of the slope on the torque/timing curve compared with data from the slope map. The difference is used to update the timing correction map. The correction map can be updated every time a spark variation occurs, allowing very fast adaptation even during rapid changes in engine operation. The slope map can be used to aim for either maximum torque or minimum emissions.

10.5.9 Neural computing

The technology behind neural computing is relatively new and is expanding rapidly. The exciting aspect is that neural networks have the capacity to learn rather than having to be programmed. This form of artificial intelligence does not require specific instructions on how a problem can be solved. The user allows the computer to adapt itself during a training period, based on examples of similar problems. After training, the computer is able to relate the problem to the solution, inputs to outputs, and thus offer a viable answer to the 'question'.

The main part of a neural computer is the neural network, a schematic representation of which is shown in Figure 10.59. In this representation the circles represent neurons and the lines represent links between them. A neuron is a simple processor, which takes one or more inputs and produces an output. Each input has an associated 'weight', which determines its intensity or strength.

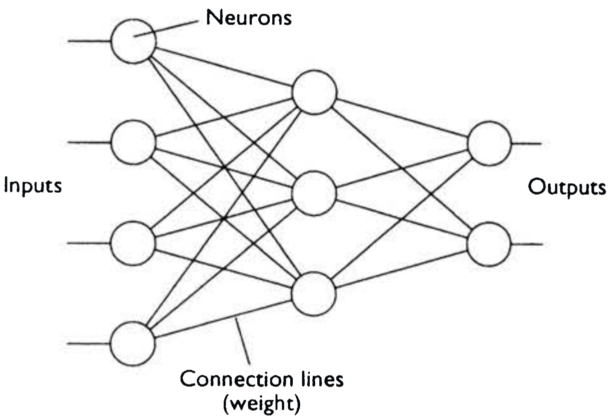


Figure 10.59 Neural network

The neuron simply has to determine the weight of its inputs and produce a suitably weighted output. The number of neurons in a network can range from tens to many thousands.

The way the system learns is by comparing its actual output with an expected output. This produces an error value, which in turn changes the relative weights of the links back through the whole network. This eventually results in an ideal solution, as connections leading to the correct answer are strengthened. This, in principle, is similar to the way a human brain works. The neural computing system has a number of advantages over the conventional method.

- Very fast operation due to 'parallel processing'.
- Reduced development time.
- Ability to find solutions to problems that are difficult to define.
- Flexible approach to a solution, which can be adapted to changing circumstances.
- More robust, as it can handle 'fuzzy' data or unexpected situations. An adaptive fuzzy system acts like a human expert. It learns from experience and uses new data to fine-tune its knowledge.

The advantages outlined make the use of neural nets on automobile systems almost inevitable. Some are even starting to be used in such a way that the engine control system is able to learn the driver's technique and anticipate the next most likely action. It can then set appropriate system parameters before the action even happens!

CHAPTER 11

Lighting

11.1 Lighting fundamentals

11.1.1 Introduction

Vehicle lighting systems are very important, particularly where road safety is concerned. If headlights were suddenly to fail at night and at high speed, the result could be catastrophic. Many techniques have been used, ranging from automatic changeover circuits to thermal circuit breakers, which pulse the lights rather than putting them out as a blown fuse would. Modern wiring systems fuse each bulb filament separately and even if the main supply to the headlights failed, it is likely that dim-dip would still work.

We have come a long way since lights such as the Lucas ‘King of the road’ were in use. These were acetylene lamps! A key point to remember with vehicle lights is that they must allow the driver to:

- See in the dark.
- Be seen in the dark (or conditions of poor visibility).

Sidelights, rear lights, brake lights and others are relatively straightforward. Headlights present the most problems, because on dipped beam they must provide adequate light for the driver, without dazzling other road users. Many techniques have been tried over the years and great advances have been made, but the conflict between seeing and dazzling is difficult to overcome.

11.1.2 Bulbs

Joseph Swan in the UK invented and patented the first light bulb in 1878 and had demonstrated it some ten years earlier. Thomas Edison tends to get the credit for some reason. Much incremental development has taken place since that time. The number, shape and size of bulbs used on vehicles is increasing all the time. Figure 11.1 shows a common selection. Most bulbs for vehicle lighting are either conventional tungsten filament bulbs or tungsten halogen.

Figure 11.1 Selection of bulbs

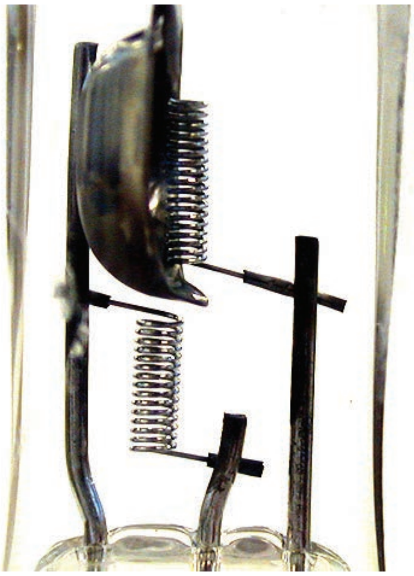


Figure 11.2 A bulb filament is like a spiralled spiral

In the conventional bulb the tungsten filament is heated to incandescence by an electric current. In a vacuum the temperature is about 2300 °C. Tungsten is a heavy metallic element and has the symbol W; its atomic number is 74; and its atomic weight 2.85. The pure metal is steel grey to tin white in colour. Its physical properties include the highest melting point of all metals: 3410 °C. Pure tungsten is easily forged, spun, drawn and extruded, whereas in an impure state it is brittle and can be fabricated only with difficulty. Tungsten oxidizes in air, especially at higher temperatures, but it is resistant to corrosion and is only slightly attacked by most mineral acids. Tungsten or its alloys are therefore ideal for use as filaments for electric light bulbs. The filament is normally wound into a 'spiralled spiral' to allow a suitable length of thin wire in a small space and to provide some mechanical strength. Figure 11.2 shows a typical bulb filament.

If the temperature mentioned above is exceeded even in a vacuum, then the filament will become very volatile and break. This is why the voltage at which a bulb is operated must be kept within tight limits. The vacuum in a bulb prevents the conduction of heat from the filament but limits the operating temperature.

Gas-filled bulbs are more usual, where the glass bulb is filled with an inert gas such as argon under pressure. This allows the filament to work at a higher temperature without failing and therefore produce a whiter light. These bulbs will produce about 17 lm/W compared with a vacuum bulb, which will produce about 11 lm/W.

Figure 11.3 Halogen bulbs

Most vehicles now use tungsten halogen bulbs for their headlights as these are able to produce about 24 lm/W (more for some modern designs). The bulb has a long life and will not blacken over a period of time like other bulbs. This is because in normal gas bulbs, over a period of time, about 10% of the filament metal evaporates and is deposited on the bulb wall. The gas in halogen bulbs is mostly iodine. The name halogen is used because there are four elements within group VIIA of the periodic table, known collectively as the halogens. The name, derived from the Greek hal- and -gen, means 'salt producing'. The four halogens are bromine, chlorine, fluorine and iodine. They are highly reactive and are not found free in nature. The gas is filled to a pressure of several bar.

The glass envelope used for the tungsten halogen bulb is made from fused silicon or quartz. The tungsten filament still evaporates but, on its way to the bulb wall, the tungsten atom combines with two or more halogen atoms forming a tungsten halide. This will not be deposited on to the bulb because of its temperature. The convection currents will cause the halide to move back towards the filament at some point and it then splits up, returning the tungsten to the filament and releasing the halogen. Because of this the bulb will not become blackened, the light output will therefore remain constant throughout its life. The envelope can also be made smaller as can the filament, thus allowing better focusing. Figure 11.3 shows tungsten halogen headlight bulbs.

Next, some common bulbs are outlined.

Festoon

The glass envelope has a tubular shape, with the filament stretched between brass caps cemented to the tube ends. This bulb was commonly used for number-plate and interior roof lighting.

Miniature centre contact (MCC)

This bulb has a bayonet cap consisting of two locating pins projecting from either side of the cylindrical cap. The diameter of the cap is about 9 mm. It has a single central contact (SCC), with the metal cap body forming the second contact, often the earth connection. It is made with various power ratings ranging from 1 to 5 W.

Capless bulb

These bulbs have a semi-tubular glass envelope with a flattened end, which provides the support for the terminal wires, which are bent over to form the two contacts. The power rating is up to 5 W, and these bulbs are used for panel lights, sidelights and parking. They are now very popular due to the low cost of manufacture.

Single contact, small bayonet cap (SBC)

These bulbs have a bayonet cap with a diameter of about 15 mm with a spherical glass envelope enclosing a single filament. A single central contact (SCC) uses the metal cap body to form the second contact. The size or wattage of the bulb is normally 5 W or 21 W. The small 5 W bulb, is used for side or tail lights and the larger 21 W bulb is used for indicators, hazard, reversing and rear fog-lights.

Double contact, small bayonet cap

This bulb is similar in shape and size to the large SCC 15 mm SBC bulb, as described above. It has two filaments, one end of each being connected to an end contact, and both of the other ends are joined to the cap body forming a third contact, which is usually the earth. These caps have offset bayonet pins so that the two filaments, which are of different wattage, cannot be connected the wrong way around. One filament is used for the stop light and the other for the tail light. They are rated at 21 and 5 W (21/5 W) respectively.

11.1.3 External lights

Regulations exist relating to external lights, the following is a simplified interpretation and amalgamation of current regulations. The range of permissible luminous intensity is given in brackets after each sub heading.

Sidelights (up to 60 cd)

A vehicle must have two sidelights each with wattage of less than 7 W. Most vehicles have the sidelights incorporated as part of the headlight assembly.

Rear lights (up to 60 cd)

Again, two must be fitted each with wattage not less than 5 W. Lights used in Europe must be ‘E’ marked and show a diffused light. Their position must be within 400 mm from the vehicle edge and over 500 mm apart, and between 350 and 1500 mm above the ground.

Brake lights (40–100 cd)

There two lights are often combined with the rear lights. They must be between 15 and 36 W each, with diffused light and must operate when any form of first line brake is applied. Brake lights must be between 350 and 1500 mm above the ground and at least 500 mm apart in a symmetrical position. High-level brake lights are now allowed and, if fitted, must operate with the primary brake lights.

Reversing lights (300–600 cd)

No more than two lights may be fitted with a maximum wattage each of 24 W. The light must not dazzle and either be switched automatically from the gearbox or with a switch incorporating a warning light. Safety reversing ‘beepers’ are now often fitted in conjunction with this circuit, particularly on larger vehicles.

Figure 11.4 Vehicle lighting designs.

Day running lights (800 cd max)

Volvo use day running lights as these are in fact required in Sweden and Finland. These lights come on with the ignition and must only work in conjunction with the rear lights. Their function is to indicate that the vehicle is moving or about to move. They switch off when parking or headlights are selected.

Rear fog lights (150–300 cd)

One or two may be fitted but, if only one, then it must be on the offside or centre line of the vehicle. They must be between 250 and 1000 mm above the ground and over 100 mm from any brake light. The wattage is normally 21 W and they must only operate when either the sidelights, headlights or front fog lights are in use.

Front spot and fog lights

If front spot lights are fitted (auxiliary driving lights), they must be between 500 and 1200 mm above the ground and more than 400 mm from the side of the vehicle. If the lights are non-dipping then they must only operate when the headlights are on main beam. Front fog lamps are fitted below 500 mm from the ground and may only be used in fog or falling snow. Spot lamps are designed to produce a long beam of light to illuminate the road in the distance. Fog lights are designed to produce a sharp cut off line such as to illuminate the road just in front of the vehicle but without reflecting back or causing glare.

Figure 11.4 shows a selection of vehicle light designs and some of the groupings used.

11.1.4 Headlight reflectors

Light from a source, such as the filament of a bulb, can be projected in the form of a beam of varying patterns by using a suitable reflector and a lens. Reflectors used for headlights are usually parabolic, bifocal or homifocal. Lenses, which are also used as the headlight cover glass, are used to direct the light to the side of the road and in a downward direction. Figure 11.5 shows how lenses and reflectors can be used to direct the light.

The object of the headlight reflector is to direct the random light rays produced by the bulb into a beam of concentrated light by applying the laws of reflection. Bulb filament position relative to the reflector is important, if the desired beam direction and shape are to be obtained. This is demonstrated in Figure 11.5(a). First, the light source (the light filament) is at the focal point, so the reflected beam will be parallel to the principal axis. If the filament is between the focal point and the reflector, the reflected beam will diverge – that is, spread outwards

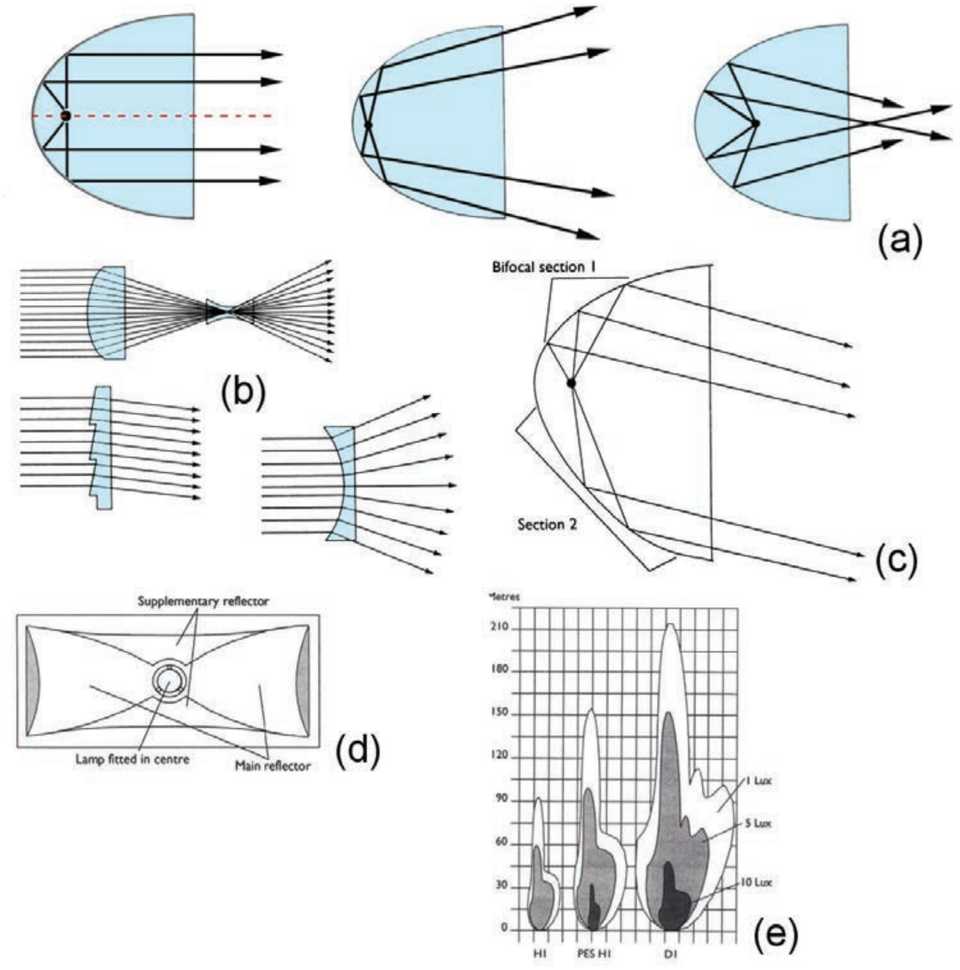


Figure 11.5 Headlight patterns are produced by careful use of lenses and reflectors

along the principal axis. Alternatively, if the filament is positioned in front of the focal point the reflected beam will converge towards the principal axis.

A reflector is basically a layer of silver, chrome or aluminium deposited on a smooth and polished surface such as brass or glass. Consider a mirror reflector that 'caves in' – this is called a concave reflector. The centre point on the reflector is called the pole, and a line drawn perpendicular to the surface from the pole is known as the principal axis. If a light source is moved along this line, a point will be found where the radiating light produces a reflected beam parallel to the principal axis. This point is known as the focal point, and its distance from the pole is known as the focal length.

Parabolic reflector

A parabola is a curve similar in shape to the curved path of a stone thrown forward in the air. A parabolic reflector (Figure 11.5(a)) has the property of reflecting rays parallel to the principal axis when a light source is placed at its focal point, no matter where the rays fall on the reflector. It therefore produces a bright parallel reflected beam of constant light intensity. With a parabolic reflector, most of the light rays from the light-bulb are reflected and only a small amount of direct rays disperses as stray light.

The intensity of reflected light is strongest near the beam axis, except for light cut-off by the bulb itself. The intensity drops off towards the outer edges of the beam. A common type of reflector and bulb arrangement is shown in Figure 11.6

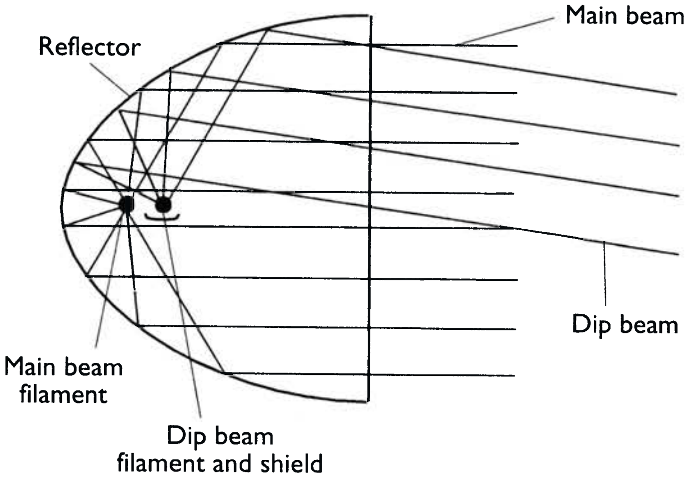


Figure 11.6 Creating a dip beam with a twin filament shielded bulb

where the dip filament is shielded. This gives a nice sharp cut-off line when on dip beam and is used mostly with asymmetric headlights.

Bifocal reflector

The bifocal reflector (Figure 11.5(c)) as its name suggests has two reflector sections with different focal points. This helps to take advantage of the light striking the lower reflector area. The parabolic section in the lower area is designed to reflect light down to improve the near field area just in front of the vehicle. This technique is not suitable for twin filament bulbs, it is therefore only used on vehicles with a four-headlight system. With the aid of powerful CAD programs, variable focus reflectors can be made with non-parabolic sections to produce a smooth transition between each area.

Homifocal reflector

A homifocal reflector (Figure 11.5(d)) is made up of a number of sections each with a common focal point. This design allows a shorter focal length and hence, overall, the light unit will have less depth.

The effective luminous flux is also increased. It can be used with a twin filament bulb to provide dip and main beam. The light from the main reflector section provides the normal long range lighting and the auxiliary reflectors improve near field and lateral lighting.

Poly-ellipsoidal headlight system (PES)

The poly-ellipsoidal system (PES) as shown in Figure 11.7 was introduced by in 1983. It allows the light produced to be as good, or in some cases better than conventional lights, but with a light-opening area of less than 30 cm^2. This is achieved by using a CAD designed elliptical reflector and projection optics. A shield is used to ensure a suitable beam pattern. This can be for a clearly defined cut-off line or even an intentional lack of sharpness. The newer PES Plus system, which was intended for larger vehicles, further improved the near-field illumination. These lights are only used with single filament bulbs and must form part of a four-headlamp system.

11.1.5 Complex shape reflectors

The surface of the reflector is calculated through advanced computer analysis using a minimum of 50 000 individual points, each specific to the head-lamp model under design. Complex shape reflectors control beam cut-off and pattern

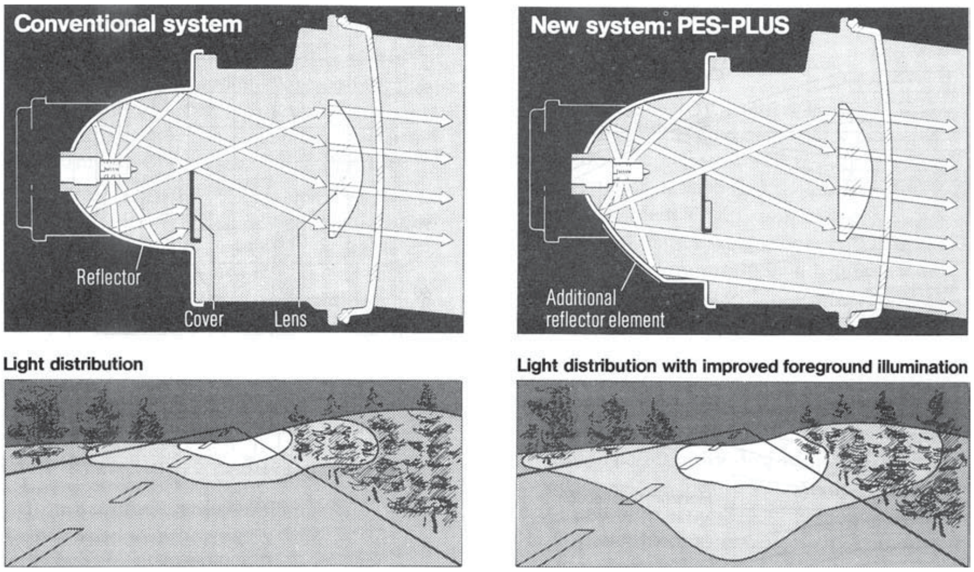


Figure 11.7 Improved poly-ellipsoid low beam

as well as homogeneity. Headlamp lenses can be perfectly clear or with striations purely for decorative purposes.

Jewel aspect signal lamps are based on the complex shape technology widely used in headlamps. Beam pattern is no longer controlled by the lens but by the reflector which, in some cases, may be in conjunction with an intermediary filter. Conventional lens optics using prisms is minimized, giving the impression of greater depth and brightness.

11.1.6 Headlight lenses

A good headlight should have a powerful far-reaching central beam, around which the light is distributed both horizontally and vertically in order to illuminate as great an area of the road surface as possible. The beam formation can be considerably improved by passing the reflected light rays through a transparent

Figure 11.8 Clear lens and a complex shape reflector

block of lenses. It is the function of the lenses partially to redistribute the reflected light beam and any stray light rays, so that a better overall road illumination is achieved with the minimum of glare. A block prism lens is shown as Figure 11.5(b).

Lenses work on the principle of refraction – that is, the change in the direction of light rays when passing into or out of a transparent medium, such as glass (plastic on some very recent headlights). The headlight front cover and glass lens, is divided up into a large number of small rectangular zones, each zone being formed optically in the shape of a concave flute or a combination of flute and prisms. The shape of these sections is such that, when the roughly parallel beam passes through the glass, each individual lens element will redirect the light rays to obtain an improved overall light projection or beam pattern.

The flutes control the horizontal spread of light. At the same time the prisms sharply bend the rays downwards to give diffused local lighting just in front of the vehicle. The action of lenses is shown as Figure 11.5(b).

Many headlights are now made with clear lenses, which means that all the light directionality is performed by the reflector.

11.1.7 Headlight levelling

The principle of headlight levelling is very simple, in that the position of the lights must change depending on the load in the vehicle. Figure 11.9 shows a simple manual aiming device operated by the driver.

An automatic system can be operated from sensors positioned on the vehicle suspension. This will allow automatic compensation for whatever the load distribution on the vehicle. Figure 11.10 shows the layout of this system. The actuators, which actually move the lights, can vary from hydraulic devices to stepper motors.

Automatic static actuators adjust beam height to the optimum position in line with vehicle load conditions. The system includes two sensors (front and rear) which measure the attitude of the vehicle. An electronic module converts data

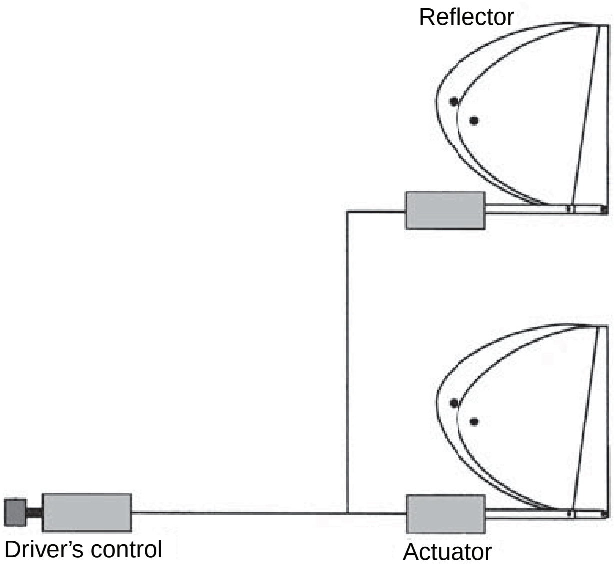


Figure 11.9 Manual headlight levelling

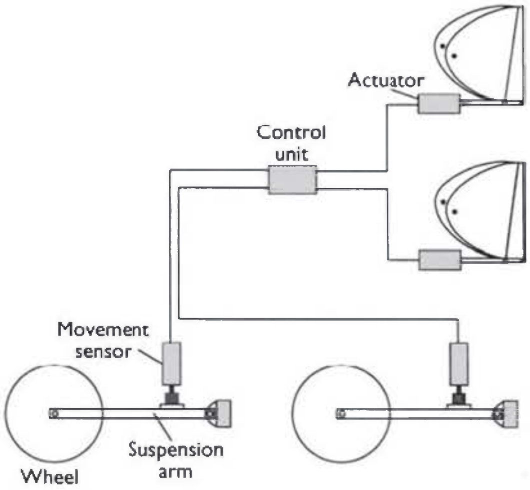


Figure 11.10 Automatic headlight adjustment

from the sensors and drives two electric gear motors (or actuators) located at the rear of the headlamps, which are mechanically attached to the reflectors.

Automatic dynamic adjusters have two sensors, an electronic module and two actuators. The sensors are the same as in the static system but the electronic module is more sophisticated in that it includes electronics that control rapid response actuator stepper motors. Response time to changes in vehicle attitude due to acceleration or deceleration is measured in tenths of a second. Corrective action is continuous and provides enhanced driving comfort, as the beam aim is optimized. In line with regulations, automatic dynamic levelling actuators are mandatory on all vehicles equipped with high intensity discharge (HID) lighting systems.

11.1.8 Headlight beam setting

Many types of beam setting equipment are available and most work on the same principle. The method is the same as using an aiming board but is more

Figure 11.11 Headlights

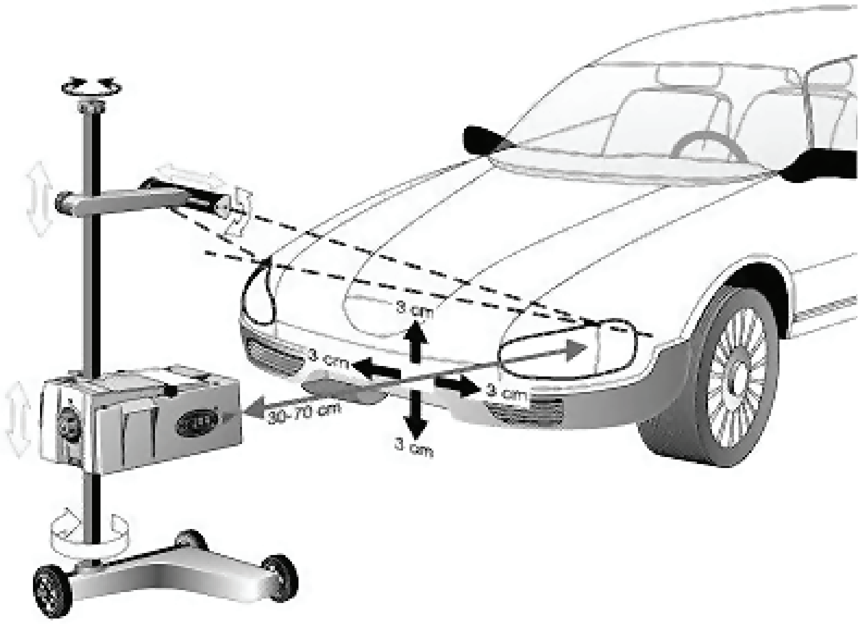


Figure 11.12 Headlamp alignment (Source: Hella)

convenient and accurate due to easier working and because less room is required.

Move the beam setter into position in front of the headlamp to be checked, and align the beam setter box with the middle of the headlamp. It must not be more than 3 cm out of line horizontally or vertically. The distance between the front edge of beam setter box and the headlamp should be between 30 and 70 cm. The beam setter must be readjusted before each headlamp is checked (Figure 11.12).

When adjusting the headlamps, the given inclination for the cut off line (from a data book etc.) must be set on the alignment equipment. The beam is now adjusted until the cut-off line and break off are in the correct position on the screen of the aligner (Figure 11.13).

To set the headlights of a car using an aiming board the following procedure should be adopted.

1. Park the car on level ground square on to a vertical aiming board at a distance of 10 m if possible. The car should be unladen except for the driver.
2. Mark out the aiming board as shown below.
3. Bounce the suspension to ensure it is level.
4. With the lights set on dip beam, adjust the cut-off line to the horizontal mark, which will in most cases be 1 cm below the height of the headlight centre for every 1 m the car is away from the board.* The break-off point should be adjusted to the centre line of each light in turn.

**Note: If the required dip is 1% then 1 cm per 1 m. If 1.2% is required then 1.2 cm per 1 m, etc. Always check data for actual settings.*

The practicality of headlight aiming is represented by Figure 11.14. Adjustment is by moving two screws positioned on the headlights, such that one will cause the light to move up and down the other will cause side-to-side movement.

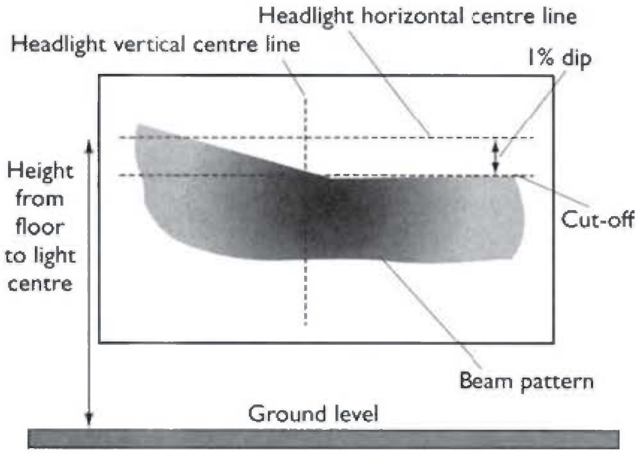


Figure 11.13 Asymmetric dip beam pattern

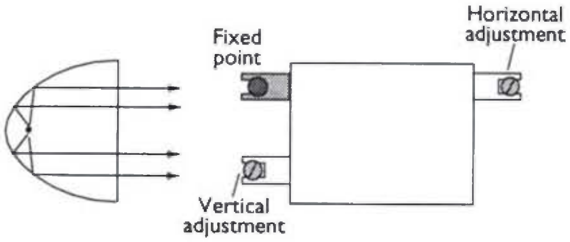


Figure 11.14 Headlight adjustment

11.2 Lighting circuits

11.2.1 Basic lighting circuit

Figure 11.15 shows a simple lighting circuit. Whilst this representation helps to demonstrate the way in which a lighting circuit operates, it is not now used in this simple form. The circuit does, however, help to show in a simple way how various lights in and around the vehicle operate with respect to each other. For example, fog lights can be wired to work only when the sidelights are on. Another example is how the headlights cannot be operated without the sidelights first being switched on.

11.2.2 Dim-dip circuit

Dim-dip headlights were an attempt to stop drivers just using sidelights in semi-dark or poor visibility conditions. The circuit is such that when sidelights and ignition are on together, then the headlights will come on automatically at about one-sixth of normal power.

Dim-dip lights are achieved in one of two ways. The first uses a simple resistor in series with the headlight bulb and the second is to use a ‘chopper’ module, which switches the power to the headlights on and off rapidly. In either case the ‘dimmer’ is bypassed when the driver selects normal headlights. Figure 11.16 is a simplified circuit of dim-dip lights using a series resistor. This is the most cost-effective method but has the problem that the resistor (about 1 Ω) gets quite hot and hence has to be positioned appropriately.

Safety first

If there is any doubt as to the visibility or conditions, switch on dipped headlights – if your vehicle is in good order it will not discharge the battery.

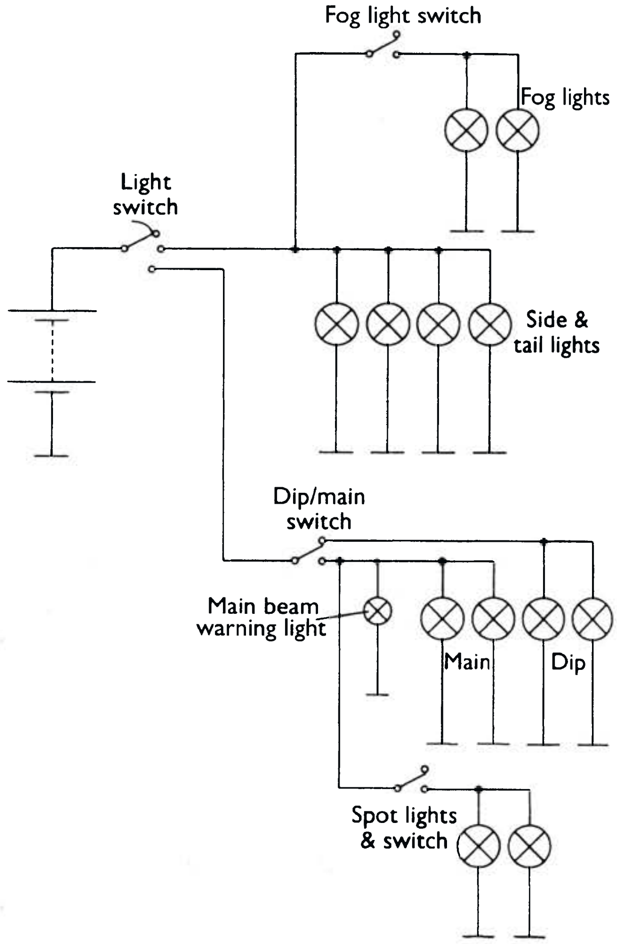


Figure 11.15 Simplified lighting circuit

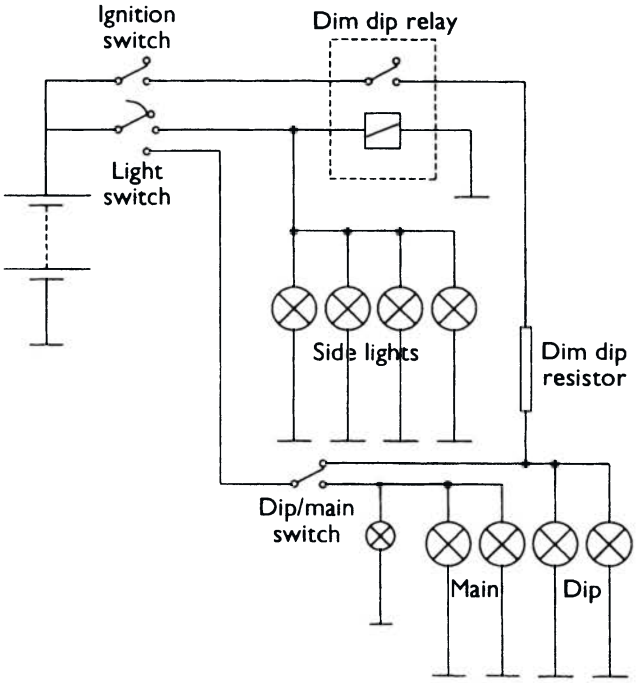


Figure 11.16 Simplified circuit dim-dip lights using a series resistor

11.2.3 General lighting circuit

The circuit shown in Figure 11.17 is the complete lighting system of a British vehicle the late nineties. It is an older design but it is very useful to illustrate how the main circuits work. Operation of the main parts of this circuit is as follows.

Sidelights

Operation of the switch allows the supply on the N or N/S wire (colour codes are discussed in chapter 4) to pass to fuses 7 and 8 on an R wire. The two fuses then supply left sidelights and right sidelights as well as the number plate light.

Dipped beam

When the dip beam is selected, a supply is passed to fuse 9 on a U wire and then to the dim-dip unit, which is now de-energized. This then allows a supply to fuses 10 and 11 on the O/U wire. This supply is then passed to the left light on a U/K wire and the right light on a U/B wire.

Main beam

Selecting main beam allows a supply on the U/W wire to the main/dip relay, thus energizing it. A supply is therefore placed on fuses 21 and 22 and hence to each of the headlight main beam bulbs.

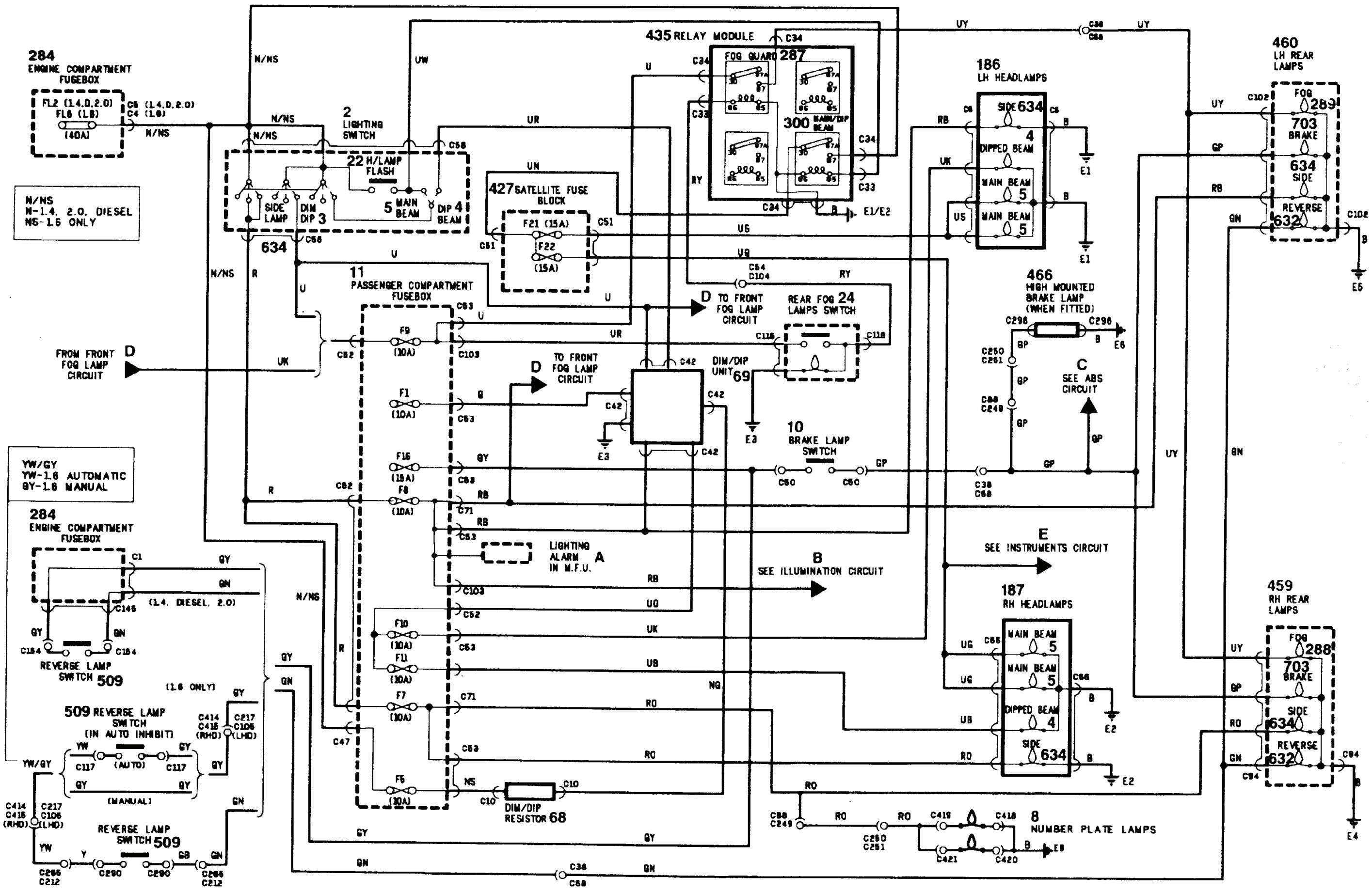


Figure 11.17 Complete vehicle lighting circuit

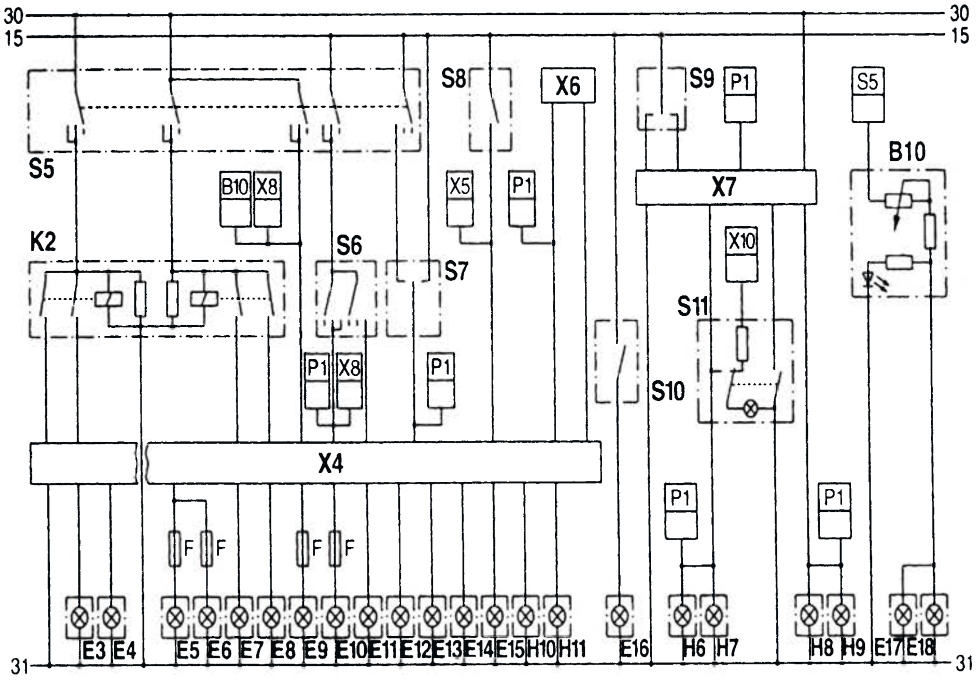


Figure 11.18 Lighting circuit flow diagram

Dim-dip

When sidelights are on there is a supply to the dim-dip unit on the R/B wire. If the ignition supplies a second feed on the G wire then the unit will allow a supply from fuse 5 to the dim-dip resistor on an N/S wire and then on to the dim-dip unit on an N/G wire. The unit then links this supply to fuses 10 and 11 (dip beam fuses).

11.2.4 Flow diagram lighting circuit

Figure 11.18 shows a typical lighting circuit using the 'flow diagram' or schematic technique. The identifiers are listed in the Table 11.1. Note that, when following this circuit, the wires do not pass directly through the 'lamp check module' from top to bottom. There is a connection to the appropriate lamp but this will be through for example, a sensing coil.

Also, note how codes are used to show connections from some components to others rather than a line representing the wire. This is to reduce the number of wires in general but also to reduce crossover points.

11.2.5 Central lighting control circuit

Key fact

All commands to operate the lights are sent via the LIN bus in central control systems.

Figure 11.19 is from a Ford vehicle with adaptive front lighting. The light switch in this case has a supply (30) and an earth/ground (31) connection, but all commands to operate the lights are sent via the LIN bus. The central module supplies outputs to operate the lights and a separate module is used in this case for the adaptive features.

11.2.6 Testing procedure

Table 11.2 lists some common symptoms of a lighting system malfunction together with suggestions for the possible fault. The faults are very generic but will serve as a good reminder.

Table 11.1 Identifiers for Figure 11.18

Identifier	Device
B10	Dimmer for instrument lighting
E3, 4	Fog warning lamps
E5, 6	Main beam headlamps
E7, 8	Fog lamps
E9, 10	Dip beam headlamps
E11, 12	Side-marker lamps
E13	Number plate lamp
E14, 15	Tail lamps
E16	Reverse lamp
E17, 18	Instrument lighting
H6, 7, 8, 9	Indicator lamps
K2	Lighting relay
S5	Headlamp switch
S6	Fog lamp switch
S7	Dip switch
S8	Stop lamp switch
S9	Turn signal switch
S10	Back-up lamp switch
S11	Hazard warning switch
X4	Plug, lamp check module
X6	Plug, check control
X7	Socket, hazard warning relay

Table 11.2 Lighting symptoms and faults

Symptom	Possible fault
Lights dim	High resistance in the circuit Low alternator output Discoloured lenses or reflectors
Headlights out of adjustment	Suspension fault Loose fittings Damage to body panels Adjustment incorrect
Lights do not work	Bulbs blown Fuse blown Loose or broken wiring/connections/fuse Relay not working Corrosion in light units Switch not making contact

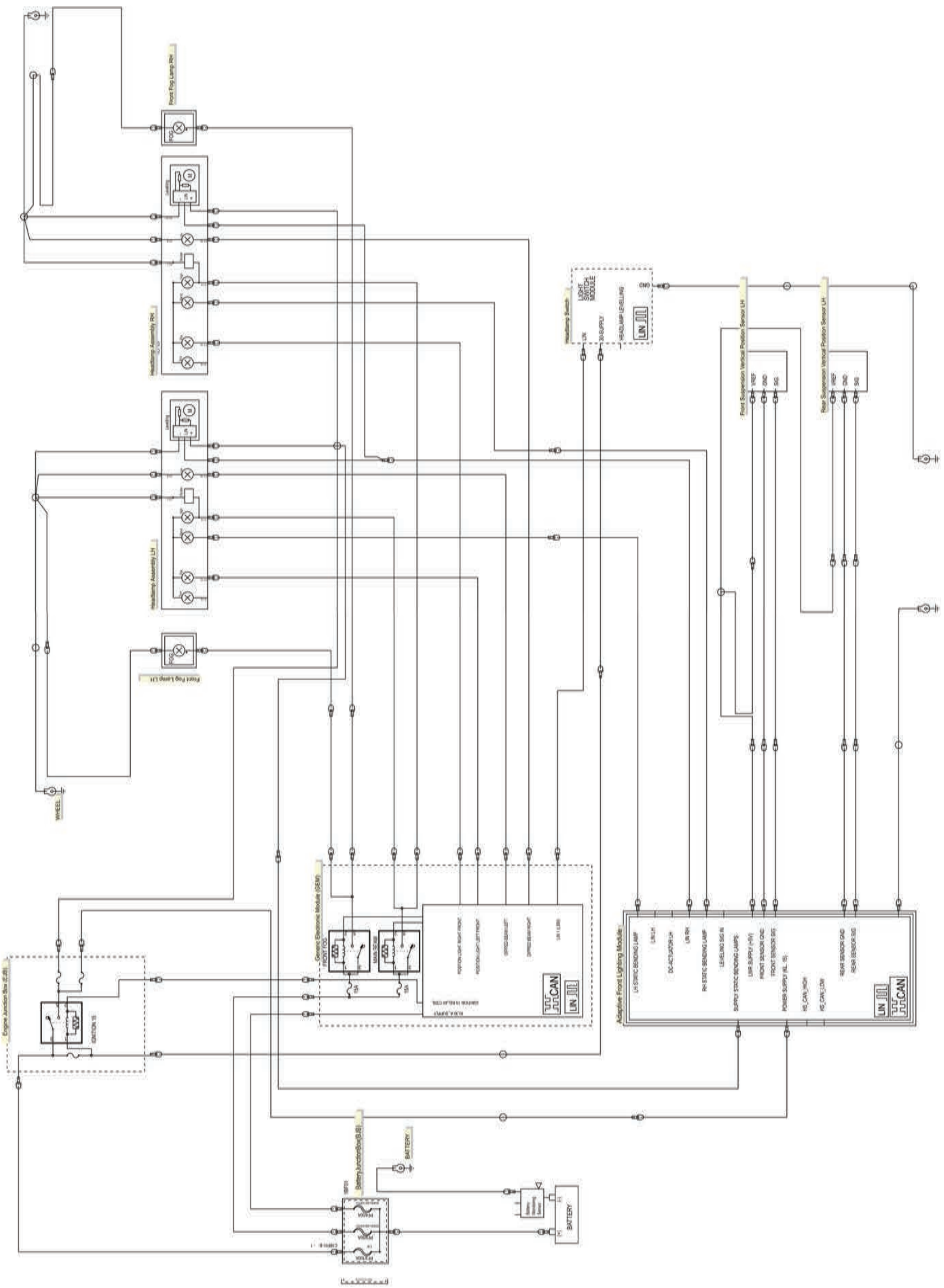


Figure 11.19 Lighting circuit using a central control module (Source: Ford Motor Company)

The process of checking a lighting system circuit is broadly as follows:

1. Check battery (see Chapter 5) – must be 70% charged.
2. Check bulb(s) – visual check or test with ohmmeter.
3. Fuse continuity – (do not trust your eyes) voltage at both sides with a meter or a test lamp.
4. If used, does the relay click (if yes, jump to stage 8) – this means the relay has operated, it is not necessarily making contact.
5. Supply to switch – battery volts.
6. Supply from the switch – battery volts.
7. Supplies to relay – battery volts.
8. Feed out of the relay – battery volts.
9. Voltage supply to the light – within 0.5 V of the battery.
10. Earth circuit (continuity or voltage) – 0 Ω or 0 V.

11.3 Gas discharge, LED and infrared lighting

11.3.1 Gas discharge lamps

Gas discharge headlamps (GDL) are now being fitted to vehicles. They have the potential to provide more effective illumination and new design possibilities for the front of a vehicle. The conflict between aerodynamic styling and suitable lighting positions is an economy/safety trade off, which is undesirable. The new headlamps make a significant contribution towards improving this situation because they can be relatively small. The GDL system (also known as high intensity discharge or HID) consists of three main components.

Key fact

Gas discharge headlamps (GDL) provide more effective illumination and new design possibilities for the front of a vehicle.

Lamp

This operates in a very different way from conventional incandescent bulbs. A much higher voltage is needed. Figure 11.20 illustrates the operating principle of a GD bulb.

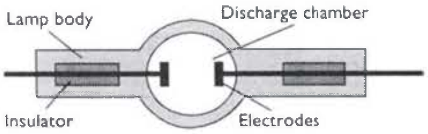


Figure 11.20 Operating principle of a gas discharge bulb

Ballast system

This contains an ignition and control unit and converts the electrical system voltage into the operating voltage required by the lamp. It controls the ignition stage and run up as well as regulating during continuous use and finally monitors operation as a safety aspect. Figure 11.21 shows the lamp circuit and components.

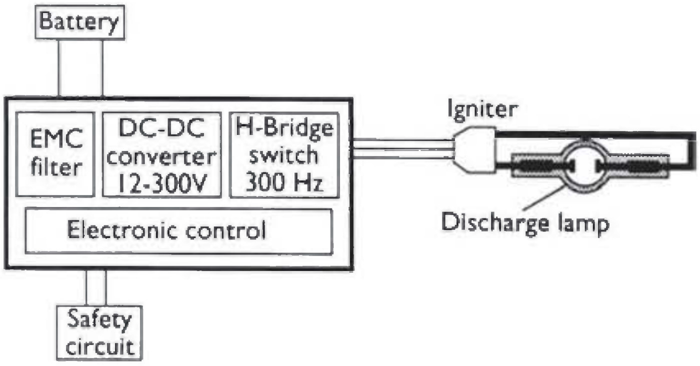


Figure 11.21 Ballast system to control a GDL

Headlamp

The design of the headlamp is broadly similar to conventional units. However, in order to meet the limits set for dazzle, a more accurate finish is needed, hence more production costs are involved.

The source of light in the gas discharge lamp is an electric arc, and the actual discharge bulb is only about 10 mm across. Two electrodes extend into the bulb, which is made from quartz glass. The gap between these electrodes is 4 mm. The distance between the end of the electrode and the bulb contact surface is 25 mm – this corresponds to the dimensions of the standardized H1 bulb.

At room temperature, the bulb contains a mixture of mercury, various metal salts and xenon under pressure. When the light is switched on, the xenon illuminates at once and evaporates the mercury and metal salts. The high luminous efficiency is due to the metal vapour mixture. The mercury generates most of the light and the metal salts affect the colour spectrum. Figure 11.22 shows the spectrum of light produced by the GDL compared with that from a halogen H1 bulb. Table 11.3 highlights the difference in output between the D1 and H1 bulbs (the figures are approximate and for comparison only).

The high output of UV radiation from the GDL means that in some cases for reasons of safety, special filters are required. Figure 11.23 shows the luminance of the GDL again compared with a halogen H1 bulb. The average output of the GDL is three times greater.

To start the D1 (GDL) lamp, the following four stages are run through in sequence.

- Ignition – a high voltage pulse causes a spark to jump between the electrodes, which ionizes the gap. This creates a tubular discharge path.
- Immediate light – the current flowing along the discharge path excites the xenon, which then emits light at about 20% of its continuous value.
- Run-up – the lamp is now operated at increased wattage, the temperature rises rapidly and the mercury and metal salts evaporate. The pressure in the

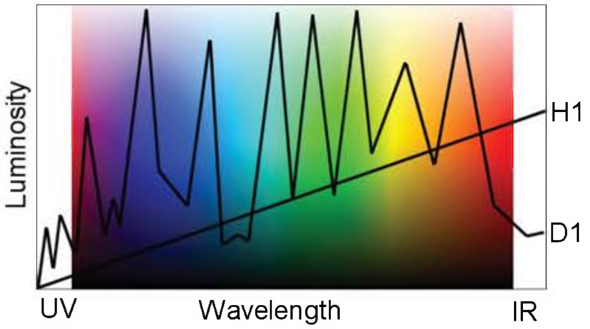


Figure 11.22 Spectrum of light produced by the GDL D1 bulb compared with that from a halogen HI bulb

Table 11.3 GDL and haligen bulbs

Bulb	Light	Heat	UV radiation
H1	8%	92%	<1%
D1	28%	58%	14%

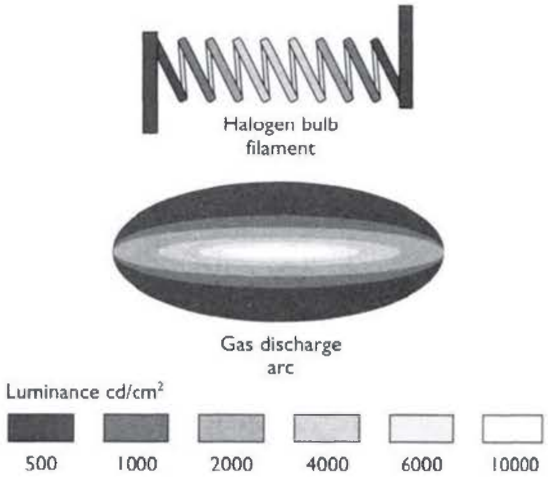


Figure 11.23 Luminance of the GDL compared with a halogen light bulb

lamp increases as the luminous flux increases and the light shifts from the blue to the white range.

- Continuous – the lamp is now operated at a stabilized power rating of 35 W. This ensures that the arc remains still and the output does not flicker. The luminous flux (28 000 lm) and the colour temperature (4500 K) are reached.

In order to control the above stages of operation, a ballast system is required. A high voltage, which can be as much as 20 kV, is generated to start the arc. During run-up, the ballast system limits the current and then also limits voltage. This wattage control allows the light to build up very quickly but prevents overshoot, which would reduce the life of the bulb. The ballast unit also contains radio suppression and safety circuits.

Key fact

A high voltage, which can be as much as 20 kV, is generated to start the arc.

The complete headlamp can be designed in a different way, as the D1 bulb produces 2.5 times the light flux and at less than half the temperature of the conventional H1 bulb. This allows far greater variation in the styling of the headlamp and hence the front end of the vehicle.

If the GDL system is used as a dip beam, the self-levelling lights are required because of the high luminous intensities. However, use as a main beam may be a problem because of the on/off nature. A GDL system for dip beam, which stays on all the time and is supplemented by a conventional main beam (four-headlamp system), is the most appropriate use. Figure 11.5(e) shows the light distribution of the D1 and H1 bulbs used in headlamps.

11.3.2 Xenon lighting

The risk of being injured or killed in a traffic accident on the roads is much higher at night than during the day, in spite of the smaller volumes of traffic. Although only about 33% of accidents occur at dusk or in the dark, the number of persons seriously injured increases by 50%, and the number of deaths by 136% compared with accidents that occur during the day.

Alongside factors such as self-dazzling caused by wet road surfaces, higher speeds because of the reduced traffic density and a reduction of about 25% of the distance maintained to the vehicle in front, causes relating to eye physiology play a very important role.

The eyes age faster than any other sensory organ, and the human eye's powers of vision begin to deteriorate noticeably from as early an age as 30!

The consequence of this, a reduction in visual acuity and contrast sensitivity when the light begins to fade, is a situation that is very rarely noticed by the motorist, as these functional deficits develop only slowly.

However, even of a person with healthy eyes, has reduced vision at night. The associated risk factors include delayed adjustment to changes between light and dark, impaired colour vision and the slow transition from day to night, which, through the habituation effect, can lull the motorist into a false sense of security.

A good xenon headlamp alone is not enough to translate the additional light quantity and quality into increased safety. In order, for example, to avoid the hazard of being dazzled by oncoming traffic, the legally required range of additional equipment includes such items as headlamp cleaning equipment and automatic beam levellers. Only the system as whole is able to provide the clear advantage of higher safety for all road-users, even under the most adverse weather conditions. This means that even in rain, fog and snow, spatial vision is improved and the motorist's orientation abilities are less restricted.

Some 94% of xenon headlamp users are convinced of their positive benefits. Visibility in rain is also judged by 80% to be better, while 75% of those surveyed have perceived an increase in safety for cyclists and pedestrians owing to the wider illumination of the road. The same percentage maintains that, thanks to xenon light, obstacles on the road are more easily recognized.

The xenon bulb is a micro-discharge bulb filled with a mixture of noble gases including xenon. The bulb has no filament, as is the case with a halogen bulb, but the light arc is created between two electrodes. As is the case with other gas discharge bulbs, the xenon bulb has an electronic starter for quick ignition, and requires an electronic ballast to function properly.

The xenon bulb provides more than twice the amount of light of a halogen bulb, while only consuming half the power. Therefore, the driver can see more clearly, and the car has more power for other functions. Moreover, it is environmentally friendly, as less power means less fuel consumption. The clear white light produced by the xenon bulb is similar to daylight. Research has shown that this enables drivers to concentrate better. Furthermore, this particular light colour reflects the road markings and signs better than conventional lighting. The xenon bulb also delivers a marked contribution to road safety in the event of limited visibility due to weather conditions. In practical terms, the life span of the bulb is equal to that of the car, which means that the bulb need only be replaced in exceptional cases.

The light produced by a xenon bulb is, in fact, not blue but white, falling well within the international specifications for white light – the light only appears blue in comparison to the warmer ‘yellow’ light produced by halogen. However, it clearly appears white in comparison to daylight. Technically speaking, it is possible to adapt the light colour produced, but this would lead to a substantial loss of intensity, thereby cancelling out the particular advantages.

The international regulations governing light distribution and intensity on the road are very strict. Xenon light falls well within these boundaries.

Xenon lighting is arguably less irritating than conventional light. This is because the light–darkness borders are much more clearly defined. The increased amount of light produced is mainly used to achieve higher intensity and better distribution of light on the road. Moreover, the verges are also better lit. There are three conditions that must be met. These are contained in the international regulations concerning the use of xenon light: the headlamps must be aligned according to regulations; the vehicle must be fitted with an automatic headlamp levelling

Figure 11.24 Xenon lighting (Source: Hella)

system, so that when the load is increased the headlight beams are automatically adjusted; the headlamp must be fitted with an automatic cleaning system, as dirt deposits on the lens act as a diffuser, thereby projecting the light beyond the prescribed range. These three conditions together with the extensive life span of the xenon bulb greatly reduce the risk of incorrectly aligned headlamps. The use of halogen bulbs entails a much higher risk.

Xenon light sometimes appears to irritate oncoming drivers. In normal circumstances drivers look straight ahead; however, due to the conspicuous colour of xenon light, drivers are more inclined to look into the headlamps. The same phenomenon was experienced during the introduction of halogen headlamps in the 1960s. In those days people also spoke of ‘that irritating white light’. The introduction of xenon headlamps will therefore entail a period in which everybody will become accustomed. Figure 11.24 shows xenon components produced by Hella.

11.3.3 Ultraviolet headlights

The GDL can be used to produce ultraviolet (UV) lights. Since UV radiation is virtually invisible it will not dazzle oncoming traffic but will illuminate fluorescent objects such as specially treated road markings and clothing. These glow in the dark much like a white shirt under some disco lights. The UV light will also penetrate fog and mist, as the light reflected by water droplets is invisible. It will even pass through a few centimetres of snow.

Key fact

The UV light will also penetrate fog and mist, as the light reflected by water droplets is invisible.

Cars with UV lights use a four-headlamp system. This consists of two conventional halogen main/dip lights and two UV lights. The UV lights come on at the same time as the dipped beams, effectively doubling their range but without dazzling.

Two-stage blue filters are used to eliminate visible light. Precise control of the filter colour is needed to ensure UVB and UVC are filtered out, as these can cause eye damage and skin cancer. This leaves UVA, which is just beyond the visible spectrum and is used, for example, in suntan lamps. However, some danger still exists; for example, if a child were to look directly and at close range into the faint blue glow of the lights. To prevent this, the lights will only operate when the vehicle is moving. This is a very promising contribution to road safety but has yet to become mainstream.

11.3.4 LED lighting

Key fact

LEDs have a typical rated life of over 50 000 hours, compared with just a few thousand for incandescent lamps.

Light emitting diode (LED) displays were first produced commercially in 1968. Almost from this time there has been speculation as to possible vehicle applications. LEDs quickly found applications in the interior of the vehicle, particularly in dashboard displays. However, until recently, legislation has prevented the use of LEDs for exterior lighting. A simple change in the legislative language from 'incandescent lamp' to 'light source', made it possible to use lighting devices other than filament bulbs. Figure 11.25 shows a light unit containing LEDs.

Figure 11.25 Light units with LED's

The advantages of LED lighting are clear, the greatest being reliability. LEDs have a typical rated life of over 50 000 hours, compared with just a few thousand for incandescent lamps. The environment in which vehicle lights have to survive is hostile to say the least. Extreme variations in temperature and humidity as well as serious shocks and vibration have to be endured.

LEDs are more expensive than bulbs but the potential savings in design costs due to sealed units being used and the greater freedom of design could outweigh the extra expense.

Most of the major manufacturers now use of LEDs for some lighting applications. In particular LEDs are used for high-level brake lights. This is because of their shock resistance, which will allow them to be mounted on the boot lid. A further advantage is that they turn on quicker than ordinary bulbs. This 'turn-on' time is important; the times are about 130 ms for the LEDs, and 200 ms for bulbs. If this is related to a vehicle brake light at motorway speeds, then the increased reaction time equates to about a car length. This is therefore, a major contribution to road safety.

Key fact

Most of the major manufacturers now use of LEDs for some lighting applications.

Heavy vehicle side marker lights are an area of use where LEDs have proved popular. Many lighting manufacturers produce lights for the after-market. Being able to use sealed units will greatly increase the life expectancy. Side indicator repeaters are a similar issue due to the harsh environmental conditions.

Due to the development and use of gallium nitride (GaN) and indium doped gallium nitride (InGaN), 'super-bright' LEDs can replace incandescent bulbs. In

Figure 11.26 LED lighting (Source: Visteon)

addition to adding another colour to the 'palate', blue is key in working within a matrix of red and green. When combined at various intensities red, green and blue will produce white or any other colour of light. However, while white light can be created in this way, coating an 'InGaN' blue LED with phosphor directly produces a white light. The process is called the phosphor down-conversion method.

Key fact

Coating an 'InGaN' blue LED with phosphor produces a white light.

11.3.5 Infrared lights

Thermal-imaging technology promises to make night driving less hazardous. Infrared thermal-imaging systems are going to be fitted to cars. The Cadillac division of General Motors has a system called 'Night Vision' as an option. After 'Night Vision' is switched on, 'hot' objects, including animals and people show up as white in the thermal image, as shown in Figure 11.27.

The infrared end of the light spectrum was discovered as long ago as 1800 by William Herschel. When investigating light passing through a prism, Herschel found heat was being emitted by rays he could not see. This part of the spectrum is called infrared (from the Latin infra, meaning 'below') because the rays are below the frequency of red light. The infrared spectrum begins at a wavelength of about 0.75 m and extends up to 1 mm. Every object at a temperature above absolute zero (–273.15 °C) emits some kind of infrared radiation.

Key fact

Every object at a temperature above absolute zero (–273.15 °C) emits some kind of infrared radiation.

On the vehicle system a camera unit sits on headlamp-type mountings in the centre of the car, behind the front grille. Its aim is adjusted just like that of headlamps. The mid-grille position was chosen because most front collisions involve offset rather than full head-on impacts. However, the sensor is claimed to be tough enough to withstand 9 mph (14.5 kph) bumper impacts anyway. The sensor is focused 125 m ahead of the car. The outer lens of the sensor is coated with silicon to protect it against scratching. Behind this are two lenses made of black glass called tecalgenite. This is a composite material that transmits infrared easily but visible light will not pass through it.

The device looks a bit like a conventional camera, but instead of film it houses a bank of ferroelectric barium-strontium-titanate (BST) sensor elements; 76 800 of them can be packed onto a substrate measuring 25 mm square. Each element is

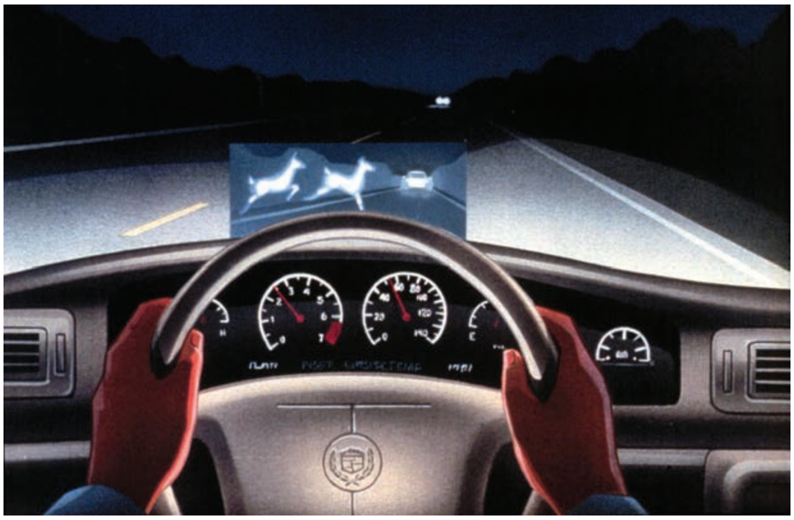


Figure 11.27 Night vision system in use

a temperature dependent capacitor, the capacitance of which changes in direct proportion to how much infrared radiation it senses. This is termed an uncooled focal plane array (UFPA). An electrically-heated element maintains a temperature of 10 °C inside the UFPA, enabling it to operate between ambient temperatures of 40 and 85°C.

Between the lens and the bank of UFPA sensor elements there is a thin silicon disc rotated by an electric motor at 1800 rpm. Helical swirls are etched on some segments of the disc. Infrared radiation is blocked by the swirls but passes straight through the plain segments. The UFPA elements respond to the thermal energy of the objects viewed by the lens. Each sensor's reading switches on and off every 1/30 of a second, thus providing video signals for the system's head-up display (HUD).

The display, built into the dashboard, projects a black-and-white image, which the driver sees near the front edge of the car's bonnet. Objects in the image are the same size as viewed by the UFPA, helping the driver judge distances to them.

Key fact

Objects in the image are the same size as viewed by the sensor, helping the driver judge distances to them.

11.4 Other lighting techniques

11.4.1 Mono-colour signal lamps

With mono-colour technology, in addition to the traditional red functions (stop, tail lamp and fog), the reverse and turn signal functions appear red when not in use, but emit white and amber light respectively when functioning. Several technologies make this possible. In the case of subtractive synthesis lamps, coloured screens are placed in front of the bulb. Their colours are selected so that, in conjunction with the red of the external lens, they colour the light emitted by the lamp in line with the regulations: white for reverse, amber for the turn signal. Complementary colour technology uses a two-colour external lens, which combines red (dominant) and its complementary colour (yellow for the turn signal, blue for reverse). The combination of these two lights – red and yellow for the turn signal, red and blue for reverse – produces the colour of light (white or amber) stipulated by the regulations.

11.4.2 Linear lighting

Linear tail lamps can easily be harmonized with the design of the vehicle by introducing the aspect of very elongated lamps. Each function light is narrow, (35 mm), and can be up to 400 mm long. The lamps use optical intermediary screens, which are so precise that they not only fulfil legal photometric requirements but also create a harmonious overall aspect and very distinct separations between the function lights. This new technology is particularly well suited for the rear of mini-vans and light trucks.

11.4.3 Neon technology

As with LED technology, neon lamps have an almost instantaneous response time (increased safety), take up little space (design flexibility) and last more than 2000 hours, thus exceeding the average use of a CHMSL during the life of the vehicle. Moreover, the neon CHMSL is very homogeneous in appearance and offers unmatched lateral visibility.

Definition

CHMSL: Centre high mounted stop lamp.

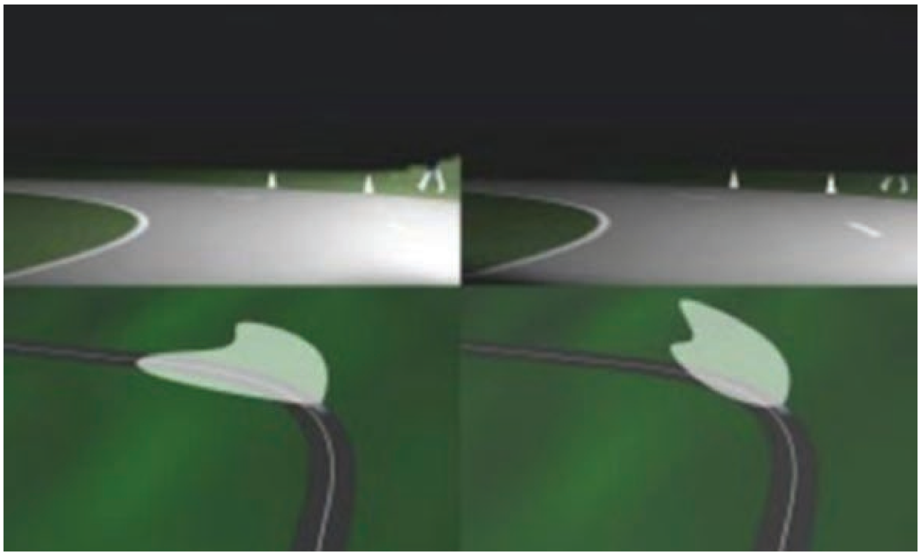


Figure 11.28 Dynamic bending light and normal lighting (Source: Valeo)

11.4.4 Bending Light

The Bending Light system consists of a bi-xenon projector, or reflector headlamp, that can rotate up from its normal position. An additional projector, or reflector, or a combination of the two can be used to deliver more light into a road bend. The actuation of the motorized lighting unit, within each headlamp assembly, is controlled by an electronic control unit, which employs signals from the steering wheel and wheel-speed sensors. A link to a satellite navigation system (GPS) can also be used if required.

The system includes three distinct lighting types:

- Motorway Lighting – typically above 80 km/h (50 mph), the low-beam function of the head-lamp is raised using a signal received from the wheel-speed sensor to actuate a self-levelling system, which increases driver visibility at high speeds.
- Adverse Weather Lighting – provides, under reduced-visibility conditions in fog, rain and snow, additional illumination to help keep track of road edges, while light is removed from the foreground to reduce reflection from the wet road.
- Town Lighting – in well-illuminated urban areas the light beam is lowered and lateral light is increased, improving pedestrian and cyclist identification at crossings as well as reducing dazzle.

Bending Light is an intelligent headlamp system that optimizes the night-time illumination of road curves by directional control of vehicle headlamps. To turn an increased quantity of light into road bends automatically, Bending Light systems adopt several flexible design approaches. Dynamic Bending Light (DBL) uses a Bi-Xenon lamp (projector or reflector type) housed in each headlamp unit, together with an electronic actuator and an electronic control unit. This design facilitates the horizontal rotation of the Bi-Xenon lamp by up to 15° from the normal 'straight-ahead' position (Figures 11.28 and 11.29). This function is controlled by a microcontroller linked to the vehicle's data network with real-time inputs from both the steering angle and speed sensors. Fixed Bending Light (FBL) employs an additional projector or reflector type lamp integrated into the headlamp unit at a 45° angle.

Key fact

Bending Light is an intelligent headlamp system that optimizes the night-time illumination of road curves by directional control of vehicle headlamps.

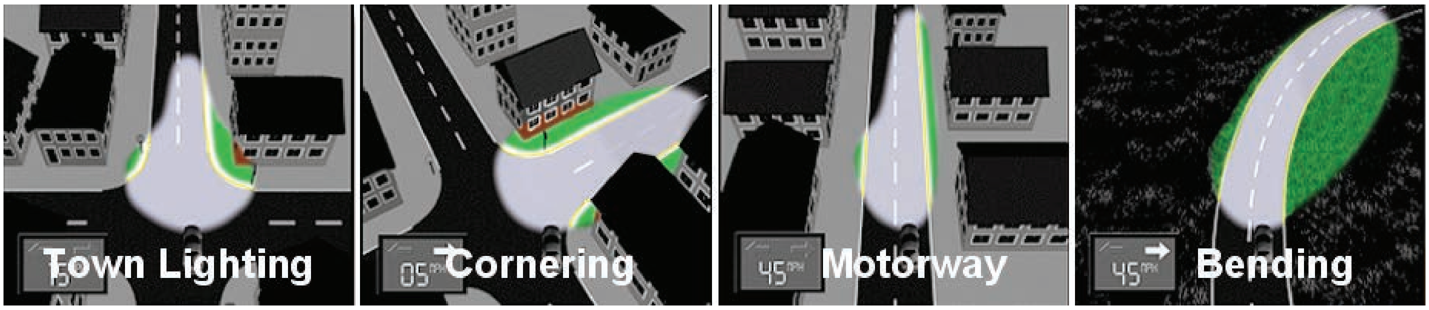


Figure 11.29 Intelligent lighting (Source: Visteon)

11.4.5 Intelligent front lighting

The lighting of modern vehicles has improved continually in the past few decades. The halogen technology set new standards after it was introduced early in the 1970s, as has xenon technology in the 1990s. The advantages of these systems are their high lighting performance and precise light distribution. The intelligent lighting systems of the future, however, will have to offer even more than this in order to make driving safer and more enjoyable.

European drivers, according to a study, would like the front lighting to respond to the various different light conditions they encounter such as daylight, twilight, night-time, and driving in and out of tunnels, and to such weather situations as rain, fog, or falling snow. They would also like better illumination on bends. Drivers would also like better light on motorways. Their list of requirements also includes better light along the edge of the road, and additional light for parking in a narrow space and when reversing.

Turning these requirements into an intelligent front lighting system means development of totally new lighting technologies that can respond in various different ways to all these different situations, some of which call for contradictory patterns of light distribution. For instance, direct lighting of the area immediately in front of the car is desirable when the roadway is dry, but can dazzle oncoming traffic if the road is wet. Light emitted above the cut-off line in fog dazzles the driver him/herself. And a long-range, narrow pattern of light distribution for high-speed motorway driving is unsuitable on twisting country roads, where the need is for a broad illumination in front of the car, possibly augmented by special head-lamps for bends or a ‘dynamic’ long-range lighting system. Despite the wide diversity of all these light-distribution patterns, none must be allowed to dazzle oncoming drivers.

Another theme is the idea of lights that switch on automatically. Unlit vehicles keep turning up at night, for instance in city-centre traffic, because the street lighting is so good that some drivers fail to notice that they are driving without lights. The same phenomenon can be seen where cars drive through tunnels. In both cases, the unlit vehicles represent a major safety risk because other road users can hardly see them.

With the aid of the sensors that are already installed on some vehicles, an intelligent lighting system can recognize the ever-changing light situation and give the appropriate assistance to the driver. For instance, the sunlight sensors that already exist for controlling air-conditioning systems, or speed sensing devices, could also deliver data to an intelligent lighting system.

Additional sensors for ambient light and light density in the field of vision, for identifying a dry or wet road, fog, and whether the road ahead is straight or curved, could also deliver important data. In modern vehicles with digital electronic systems and bus interfaces, these data will not only be useful to the lighting systems but also to the other electronically controlled systems, such as ABS or ASR, and give the driver vital assistance particularly in the most difficult driving situations.

The data transmitted by the various sensors on a vehicle can only be put to use if the vehicle has a 'dynamic' headlamp system that is capable of producing various different light-distribution patterns. This could begin with an automatic, dynamic height-adjustment and headlamps that automatically swivel sideways and could even include variable reflectors providing a whole range of light-distribution patterns.

11.5 Advanced lighting technology

11.5.1 Lighting terms and definitions

Many unusual terms are used when relating to lighting, this section aims to give a simplified description of those used when dealing with vehicle lighting. First terms associated with the light, itself and then terms relating more particularly to vehicle lights. The definitions given are generally related to the construction and use of headlights.

Luminous flux (φ)

The unit of luminous flux is the lumen (lm). Luminous flux is defined as the amount of light passing through an area in one second. The lumen is defined as the light falling on a unit area at a unit distance from a light source, which has a luminous intensity of one candela.

Luminous intensity *I*

Power to produce illumination at a distance. The unit is the candela (cd) it is a measure of the brightness of the light rather than the amount of light falling on an object.

Illumination intensity *E*

This can be defined on a surface as the luminous flux reaching it per unit area. The luminous intensity of a surface such as the road will be reduced if the light rays are at an angle. The unit is the lux (lx) it is equivalent to one lumen per square metre or to the luminance of a surface one metre from a point source of light of one candela. In simple terms it depends on the brightness, distance from, and angle to, a light source.

Brightness or luminance *L*

This should not be confused with illumination. For example when driving at night the illumination from the vehicle lights will remain constant. The brightness or luminance of the road will vary depending on its surface colour. Luminance therefore depends not just on the illumination but also on the light reflected back from the surface.

Range of a headlight

The distance at which the headlight beam still has a specified luminous intensity is its range.

Geometric range

This is the distance to the cut-off line on the road surface when the dip beam is set at an inclination of below the horizontal.

Visual range

This is affected by many factors so cannot be expressed in units but it is defined broadly as the distance within the luminous field of vision, at which an object can still be seen.

Signal identification range

The distance at which a light signal can be seen under poor conditions.

Glare or dazzle

This is difficult to express, as different people will perceive it in different ways. A figure is used however and that is if the luminous intensity is 1 lx at a distance of 25 m, in front of a dipped headlight at the height of the light centre, then the light is said not to glare or dazzle. The old British method said that the lights must not dazzle a person on the same horizontal plane as the vehicle at a distance over 25 feet, whose eye level is more than 3ft 6in above the plane (I presume he was sitting down!). In general headlights when on dipped beam must fall below a horizontal line by about 1% (1.2% or more in some cases) or 1 cm or 1.2 cm/m.

11.5.2 Single light-source lighting

It is possible to use a gas discharge lamp (GDL) as a central source for vehicle lighting. Development of this new headlamp system allows a reduction in headlamp dimensions for the same output or improved lighting with the same dimensions. Using a GDL as a central light source for all the vehicle lights is shown in Figure 11.30.

The principle is that light from the ‘super light source’, is distributed to the headlamps and other lamps by a light-guide or fibre-optic link. The light from the GDL enters the fibre-optics via special lenses and leaves the light-guide in a similar manner as shown in Figure 11.31. A patterned covered lens provides the required light distribution. Shields can provide functions such as indicators, or electro-chromatic switches may even become available.

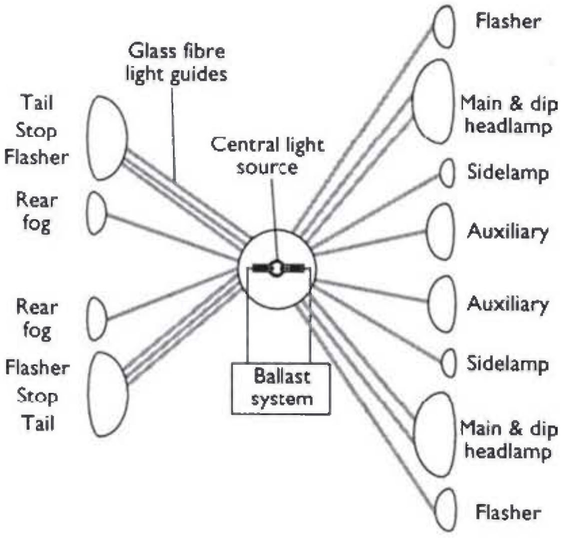


Figure 11.30 GDL as central light source for all the vehicle lights

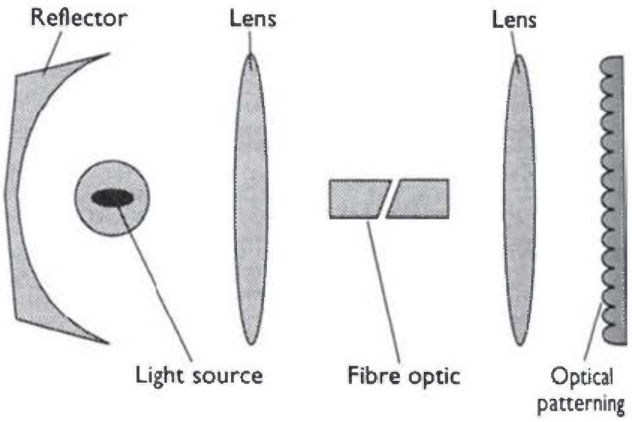


Figure 11.31 The light from the gas discharge lamp (GDL) enters and leaves the light guide via a special lenses

Heat build-up can be a problem in the fibre-optics but an infrared permeable coating on the reflector will help to alleviate this issue. The light-guide system has a very low photometric efficiency (10–20% at best), but the very efficient light source still makes this technique feasible. One of the main advantages is being able to improve the light distribution of the main headlamp. Due to the legal limits with regard to dazzle, conventional lights do not intensely illuminate the area just under the cut-off line. Consequently, several glass fibre bundles can be used to direct the light in an even distribution onto the desired areas of the road.

The central light source can be placed anywhere in the vehicle. Only one source is required but it is thought that a second would be used for safety reasons. A vehicle at present uses some 30 to 40 bulbs, and this number could be reduced markedly. A single light source could be utilized for rear lights on the vehicle, which would allow rear lights with an overall depth of only about 15 mm. This could be supplied with light from a single conventional bulb.

CHAPTER 12

Auxiliaries

12.1 Windscreen washers and wipers

12.1.1 Functional requirements

The requirements of the wiper system are simple. The windscreen must be clean enough to provide suitable visibility at all times. To do this the wiper system must meet the following requirements:

- Efficient removal of water and snow.
- Efficient removal of dirt.
- Operate at temperatures from –30 to 80 °C.
- Pass the stall and snow load test.
- Have a service life in the region of 1.5 million wipe cycles.
- Be resistant to corrosion from acid, alkali and ozone.

Most wiper linkages consist of a series or parallel mechanism (Figure 12.1). Some older types use a flexible rack and wheel boxes similar to the operating mechanism of many sunroofs. The latest versions drive the blades directly from the motors. One of the main considerations for the design of a wiper linkage is the point at which the blades must reverse. This is because of the high forces on the motor and linkage at this time. If the reverse point is set so that the linkage is at its maximum force transmission angle, then the reverse action of the blades puts less strain on the system. This also ensures

Key fact

A key consideration for the design of a wiper linkage is the point at which the blades reverse.

Figure 12.1 Wiper motor and linkage

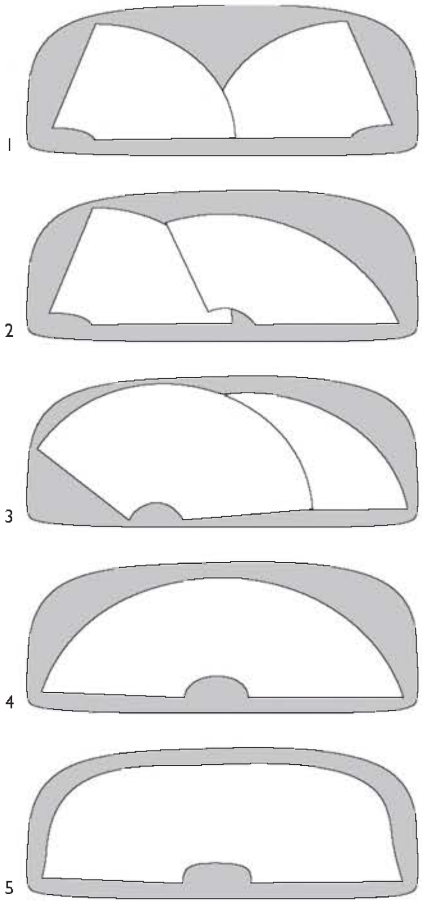


Figure 12.2 Five techniques of moving wiper blades on the screen

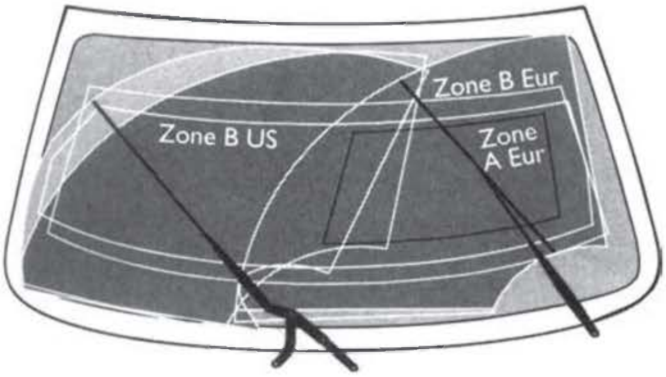


Figure 12.3 Non circular wiping

smoother operation.

The requirements of the wiper system are simple. The windscreen must be clean enough to provide suitable visibility at all times. To do this, the wiper system must meet the following requirements.

- Efficient removal of water and snow.
- Efficient removal of dirt.
- Operate at temperatures from 30 to 80 °C.
- Pass the stall and snow load test.
- Service life in the region of 1 500 000 wipe cycles.
- Resistant to corrosion from acid, alkali and ozone.

In order to meet the above criteria, components of good quality are required for both the wiper and washer system. The actual method used by the blades in cleaning the screen can vary, providing the legally prescribed area of the screen is cleaned. Figure 12.2 shows five such techniques.

Figure 12.3 shows how the front screen is split into ‘zones’ and how a ‘non-circular wiping’ technique can be applied.

12.1.2 Wiper blades

The wiper blades (Figure 12.4) are made of a rubber compound and are held on to the screen by a spring in the wiper arm. The aerodynamic properties of

Figure 12.4 Details of a wiper blade

the wiper blades have become increasingly important due to the design of the vehicle as different air currents flow on and around the screen area. The strip on top of the rubber element is often perforated to reduce air drag. A good quality blade will have a contact width of about 0.1 mm. The lip wipes the surface of the screen at an angle of about 45°. The pressure of the blade on the screen is also important as the coefficient of friction between the rubber and glass can vary from 0.8 to 2.5 when dry and 0.1 to 0.6 when wet. Temperature and velocity will also affect these figures.

Key fact

The aerodynamic properties of the wiper blades have become increasingly important.

12.1.3 Wiper linkages

Most wiper linkages consist of series or parallel mechanisms. Some older types use a flexible rack and wheel boxes similar to the operating mechanism of many sunroofs. One of the main considerations for the design of a wiper linkage is the point at which the blades must reverse. This is because of the high forces on the motor and linkage at this time. If the reverse point is set so that the linkage is at its maximum force transmission angle then the reverse action of the blades puts less strain on the system. This also ensures smoother operation. Figure 12.5 shows two typical wiper linkage layouts, the first figure is shown at the reverse point. Note that the position of the rotary link and the angles of the rods are designed to reduce the loading on the motor at this point.

Figure 12.6 shows one method used on some vehicles together with the cam linkage, which allows off-screen parking.

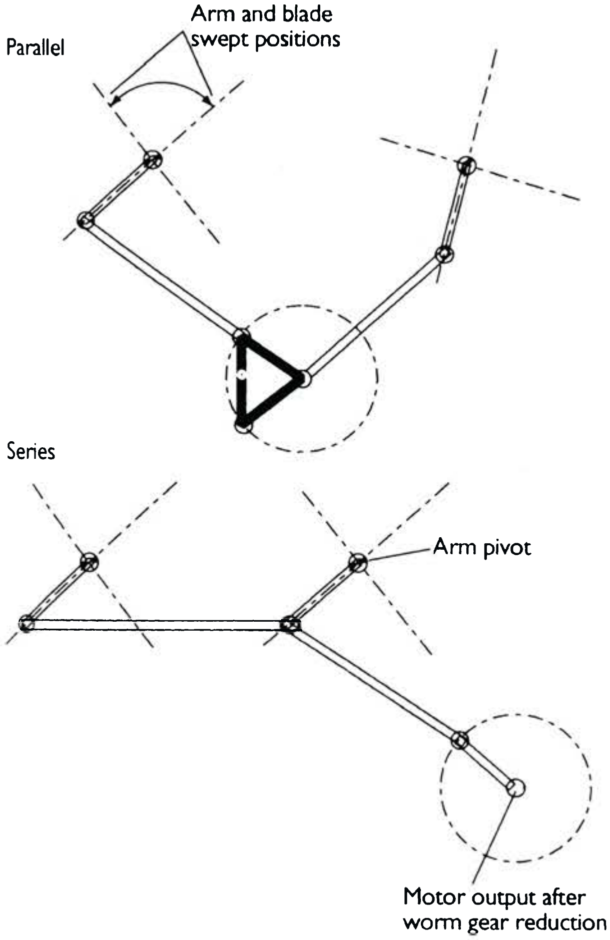


Figure 12.5 Two typical wiper linkage layouts

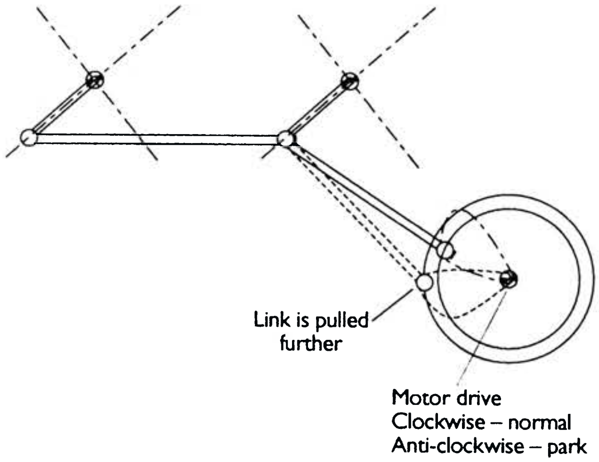


Figure 12.6 Wiper linkage used on some vehicles, together with the cam link which allows off-screen reverse parking.

12.1.4 Wiper motors

Key fact

All modern wiper motors are permanent magnet types and the drive is taken via a worm gear to increase torque and reduce speed.

Key fact

Wiper motors or the associated circuit must have some kind of short circuit and stall protection.

All modern wiper motors are permanent magnet types. The drive is taken via a worm gear to increase torque and reduce speed. Three brushes may be used, or some form of electronic control, to allow two-speed operation. In three brush motors, the normal speed operates through two brushes placed in the usual positions opposite to each other. For a fast speed, the third brush is placed closer to the earth brush. This reduces the number of armature windings between them, which reduces resistance, hence increasing current and therefore speed.

Wiper motors or the associated circuit must have some kind of short circuit protection. This is to protect the motor in the event of stalling, if frozen to the screen for example. A thermal trip of some type is often used or a current sensing circuit in the wiper ECU if fitted.

Figure 12.7 shows a typical wiper motor. Typical specifications for wiper motor speed and hence wipe frequency are 45 rpm at normal speed and 65 rpm at fast

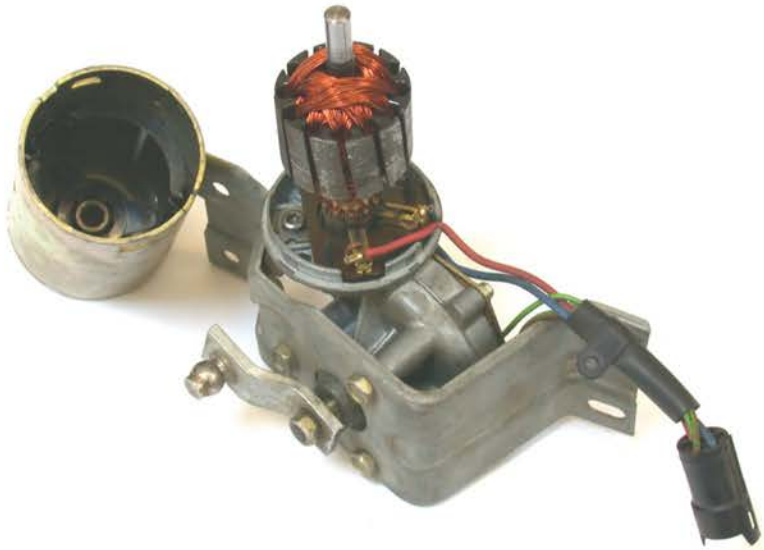


Figure 12.7 Motor brushes and armature

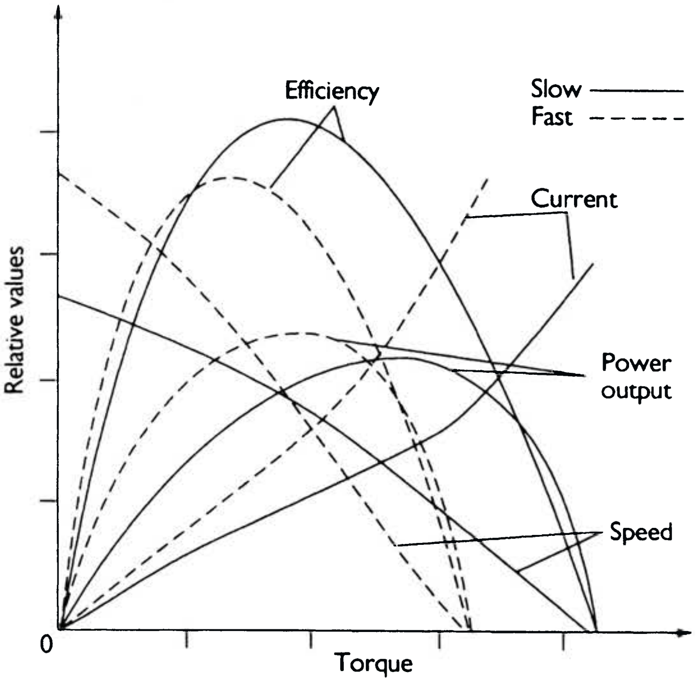


Figure 12.8 Characteristics of a wiper motor; the two sets of curves indicate fast and slow speed

speed. The motor must be able to overcome the starting friction of each blade at a minimum speed of 5 rpm. The characteristics of a typical car wiper motor are shown in Figure 12.8. The two sets of curves indicate fast and slow speed.

Wiper motors, or the associated circuit, often have some kind of short circuit protection. This is to protect the motor in the event of stalling, if frozen to the screen for example. A thermal trip of some type is often used or a current sensing circuit in the wiper ECU, if fitted. The maximum time a motor can withstand stalled current is normally specified. This is usually in the region of about 15 minutes.

12.1.5 Windscreen washers

The windscreen washer system usually consists of a simple DC permanent magnet motor driving a centrifugal water pump. The water, preferably with a cleaning additive, is directed onto an appropriate part of the screen by two or more jets. A non-return valve is often fitted in the line to the jets to prevent water siphoning back to the reservoir. This also allows ‘instant’ operation when the washer button is pressed. The washer circuit is normally linked to the wiper circuit such that when the washers are operated the wipers start automatically and will continue for several more sweeps after the washers have stopped. The circuit is shown in the next section.

Key fact

The windscreen washer is a simple DC permanent magnet motor driving a centrifugal water pump.

A windscreen washer system consists of a simple DC permanent magnet motor, which drives a centrifugal water pump. The water, preferably with a cleaning additive, is directed onto an appropriate part of the screen by two or more jets. A non-return valve is often fitted in the line to the jets to prevent water siphoning back to the tank. This also allows ‘instant’ operation when the washer button is pressed. The washer circuit is normally linked in to the wiper circuit, such that when the washers are operated, the wipers start automatically and will continue, for several more sweeps, after the washers have stopped.

Figure 12.9 Washer motors and pumps (Source: MadeInChina.com)

12.1.6 Washer and wiper circuits

Figure 12.10 shows a circuit for fast, slow and intermittent wiper control. The switches are shown in the off position and the motor is stopped and in its park position. Note that the two main brushes of the motor are connected together via the limit switch, delay unit contacts and the wiper switch. This causes regenerative braking because of the current generated by the motor due to its momentum after the power is switched off. Being connected to a very low resistance loads up the 'generator' and it stops instantly when the park limit switch closes.

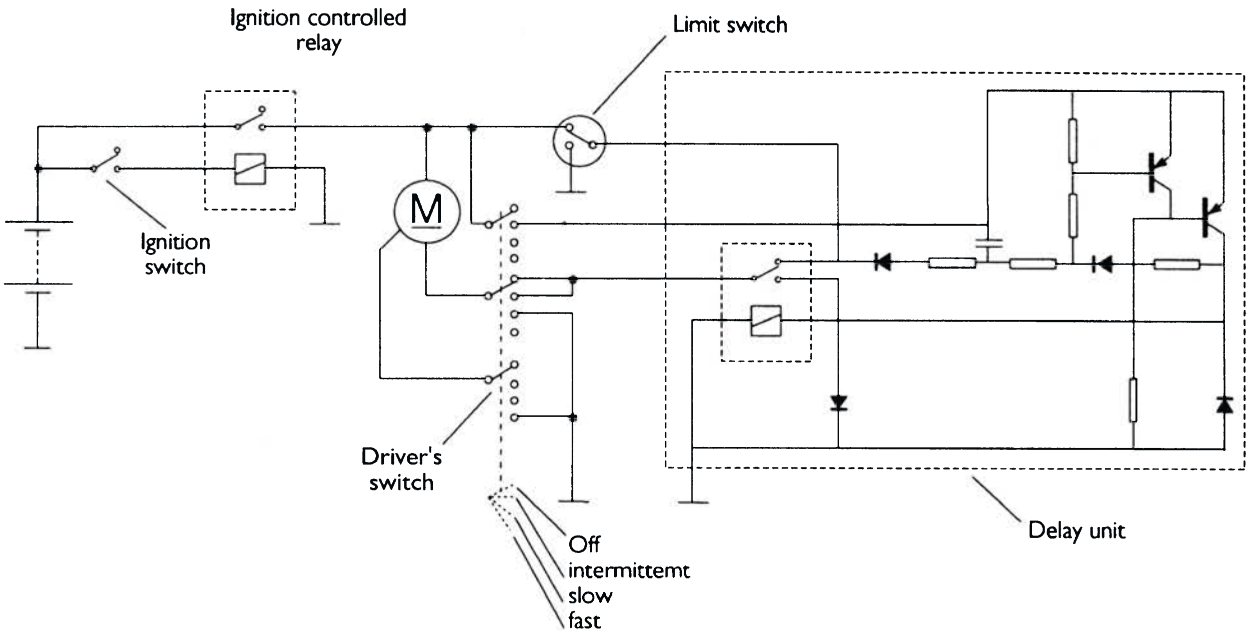


Figure 12.10 Wiper circuit with intermittent/delay operation as well as slow and fast speed

When either the delay contacts or the main switch contacts are operated the motor will run at slow speed. When fast speed is selected the third brush on the motor is used. On switching off, the motor will continue to run until the park limit switch changes over to the position shown. This switch is only in the position shown when the blades are in the parked position.

A simple capacitor-resistor (CR) timer circuit often based around a 555 IC or similar integrated circuit is used to control intermittent wipe. The charge or discharge time of the capacitor causes a delay in the operation of a transistor, which in turn operates a relay with change-over contacts.

Figure 12.11 shows the circuit of a programmed wiper system. The ECU contains two change-over relays to enable the motor to be reversed. Also contained in the ECU is a circuit to switch off the motor supply in the event of the blades stalling. To reset this, the driver’s switch must be returned to the off position.

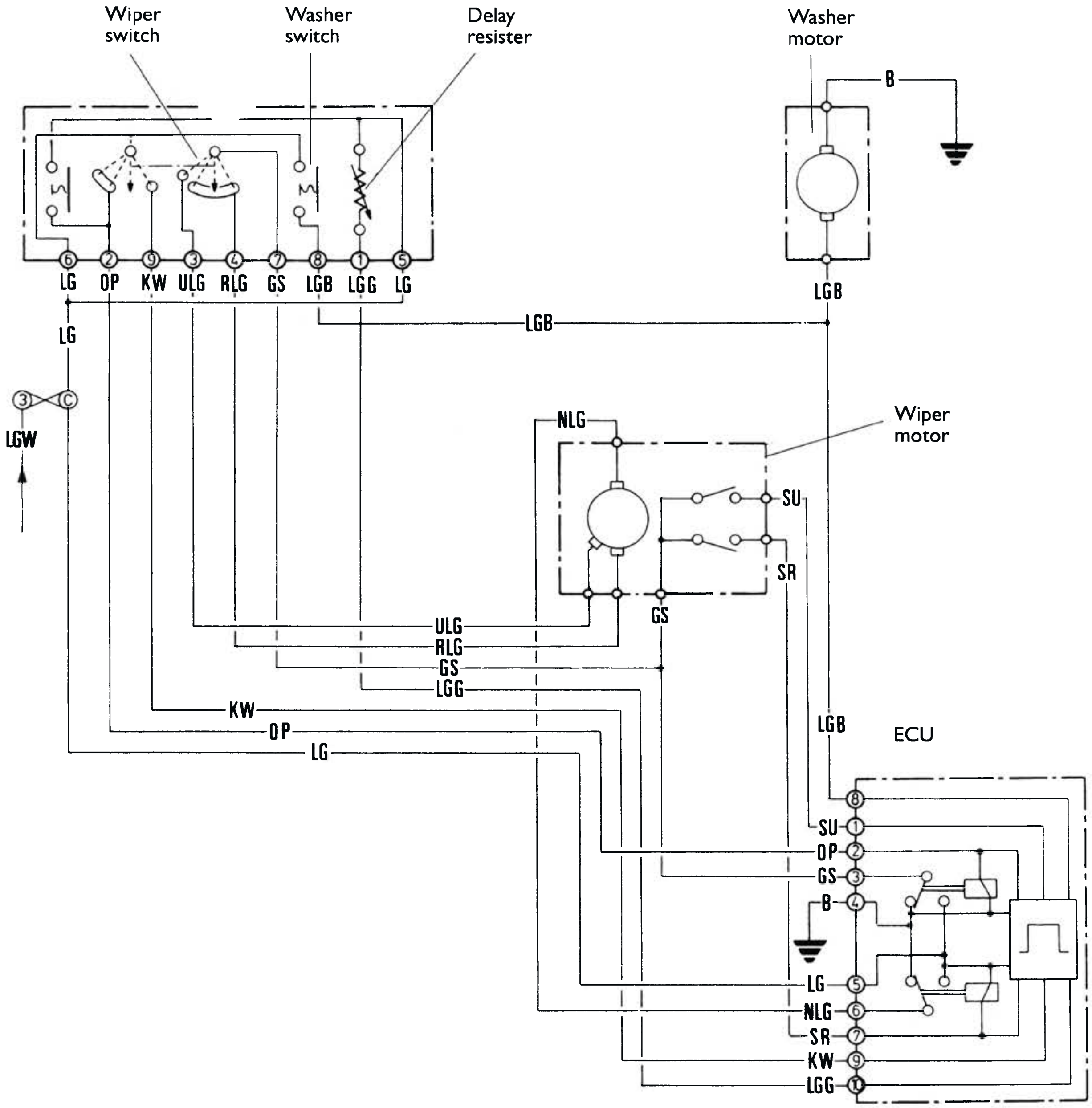


Figure 12.11 Programmed washer wipe and variable intermittent wipe circuit

12.1.7 Electronic control of windscreen wipers

Key fact

Further control of wipers other than just delay is possible with appropriate electronic control.

Further control of wipers other than just delay is possible with appropriate electronic control. Manufacturers have used programmed electronic control of the windscreen wipers for a number of years now. One system consists of a two-speed motor with two limit switches, one for the park position and one that operates at the top limit of the sweep. A column switch is utilized that has positions for wash/wipe, fast speed, slow speed, flick wipe and delay, and which has several settings. The heart of this system is the programmed wiper control unit. An innovative feature is that the wiper blades may be parked below the screen. This is achieved by utilizing the top limit switch to signal the ECU to reverse the motor for parking. The switch is normally closed and switches open circuit when the blades reach the 'A' post. Due to the design of the linkage, the arms move further when working in reverse and pull the blades off the screen. The normal park limit switch stops the motor, via the ECU in this position.

Many vehicles use a system with more enhanced facilities. This is regulated, by what may be known as, a central control unit (CCU), a multi-function unit (MFU) or a general electronic module (GEM) (Figure 12.12).

Using electronic control, a CCU allows the following facilities for the wipers:

- Front and rear wash/wipe.
- Intermittent wipe.
- Time delay set by the driver.
- Reverse gear selection rear wipe operation.
- Rear wash/wipe with 'dribble wipe' (an extra wipe several seconds after washing).
- Stall protection.

These units can often control other systems as well as the wipers, thus allowing reduced wiring bulk under the dash area. Electric windows, headlights and a heated rear window, to name just a few, are now often controlled by a central unit. A CCU allows the following facilities for the wipers (front and rear).

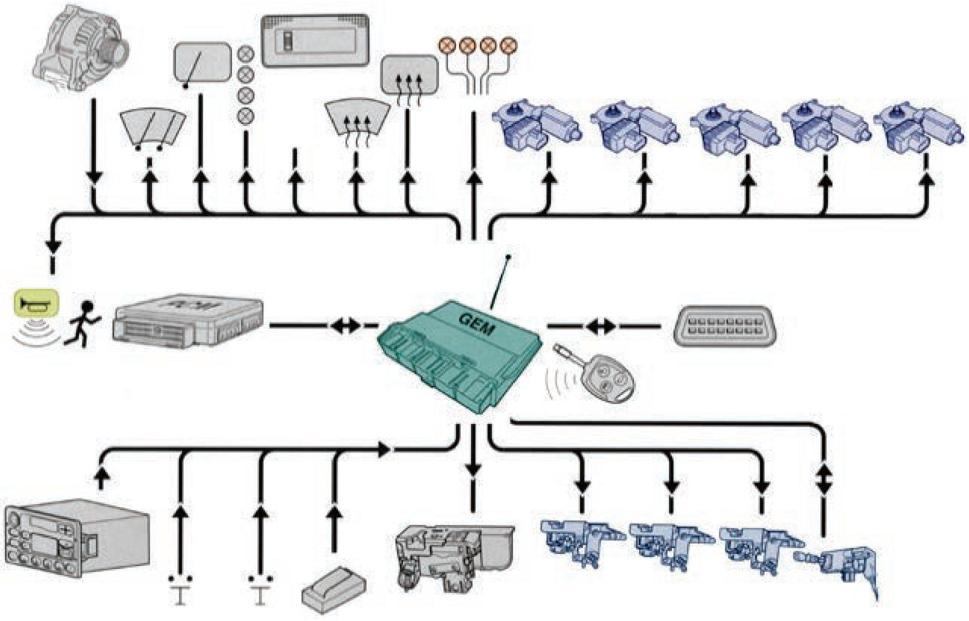


Figure 12.12 GEM and components (Source: Ford Motor Company)

Front wash/wipe

The CCU activates the wipers when the washer switch is pressed and keeps them going for a further six seconds when the switch is released.

Intermittent wipe

When the switch is moved to this position, the CCU operates the wipers for one sweep. When back in the rest position, the CCU waits for a set time and then operates another sweep and so on. This continues until the switch is moved to the off position. The time delay can be set by the driver – as one of five settings of a variable resistor. This changes the delay from about 3 s with a resistance of 500 Ω, to a delay of about 20 s with a resistance of 5400 Ω.

Rear wiper system

When the switch is operated, the CCU operates the rear wipers for three sweeps by counting the signal from the park switch. The wiper will then be activated once every six seconds until switched off by the driver.

Rear wash/wipe

When the rear washer switch is pressed, the CCU will operate the rear wiper and then continue its operation for three sweeps after the washer switch is released. If the rear wiper is not switched on the CCU will operate the blades for one more sweep after about 18 s. This is commonly known as the ‘dribble wipe’!

Rear wiper when reverse gear is selected

If the front wipers are switched on and reverse gear is selected the CCU will operate the rear wiper continuously. This will stop when either the front wipers are switched off or reverse gear is deselected.

Stall protection

When the rear wiper is operated, the CCU starts a timer. If no movement is detected within 15 s the power to the motor is removed. This is reset when the driver’s switch is moved to the off position.

12.1.8 Synchronized wipers

A direct drive system for windshield wipers has been developed by Bosch. The two drives of the dual motor wiper system do not need additional mechanical linkage and are therefore smaller – the volume of each unit is about half a litre.

Figure 12.13 Synchronized wiper motors (Source: Bosch Media)

The direct drive system needs up to 75% less space and is over a kilogram lighter than standard drive and linkage systems.

Key fact

Because the new drives require no linkage, installation space is freed up.

Each wiper has its own compact drive and is mounted directly on the drive shaft, which makes the new system easier to integrate into vehicles. Depending on the arrangement of the wiper arms, conventional wiper drives can be nearly as wide as the car body. Because the new drives require no linkage, installation space is freed up. Therefore, more room is available for other components, such as the air conditioning unit, head-up displays and other new comfort features, as well as for larger brake power boosters and pedestrian airbags to make vehicles safer (Figure 12.14).

An electronic control unit takes the place of the mechanical linkage. The control unit synchronizes the two drive units, with sensors in the wiper drive monitoring the position of the two wiper arms. The new system can be used for parallel and opposed-pattern wiper layout. Electronic position sensing and control enables the wipers to always sweep very close to the A-pillar.

Once the system has been mounted onto a vehicle, the distances between the wiper and the A-pillar on that specific vehicle can be easily programmed on the production line, eliminating assembly and bodywork tolerances. This allows the wipers to sweep a maximum possible area without running the risk of hitting the pillars. An energy and thermal management device protects the drives from overload without restricting wiper function. A software-based blockage recognition system detects obstructions on the windshield such as accumulations of snow and reduces the area swept, but only to the extent necessary. Current wiper systems react by stopping completely, leaving the driver unable to see the road ahead.

Key fact

Electronic position sensing and control enables the wipers to always sweep very close to the A-pillar.

Each drive unit consists of a mechatronic drive that can run backward and forward. Since the wiper drives no longer require vehicle-specific linkage, they can be of identical design for both sides. Specifications such as the sweep angle and rest position are individually programmed on the production line after the wiper system has been installed.

The systems can also be designed completely identically for right- and left-hand drive: the alignment is simply specified in the software. In the same way, identical drive units can be tailored to the windscreen shape, which varies depending on whether the vehicle is a sedan, coupe, convertible, etc. This considerably simplifies logistics and storage.

The electronic synchronized wiper system reduces the impact of headwind and rain intensity on the wiping frequency, and the size of the wipe pattern. In this way, the electronic system always provides the maximum field of view at a constant sweep rate.

12.1.9 Wiper blade pressure control

Key fact

Sensors are used to determine the air stream velocity and intensity of the rain.

A system of wiper pressure control, which can infinitely vary the pressure of the blade onto the screen depending on vehicle speed, has been developed. At high speeds the air stream can cause the blades to lift and judder. This seriously reduces the cleaning effectiveness. If the original pressure is set to compensate this, the pressure at rest could deform the arms and blades.

The pressure control system is shown in Figure 12.15. Sensors are used to determine the air stream velocity and intensity of the rain. An ECU evaluates the data from these sensors and passes an appropriate signal to the servo motor. When the blades are in the rest position, pressure is very low to avoid damage. The pressure rises with increasing vehicle speed and heavy rain.

Figure 12.14 Comparison of single and twin-motor wiper systems (Source: Bosch Media)

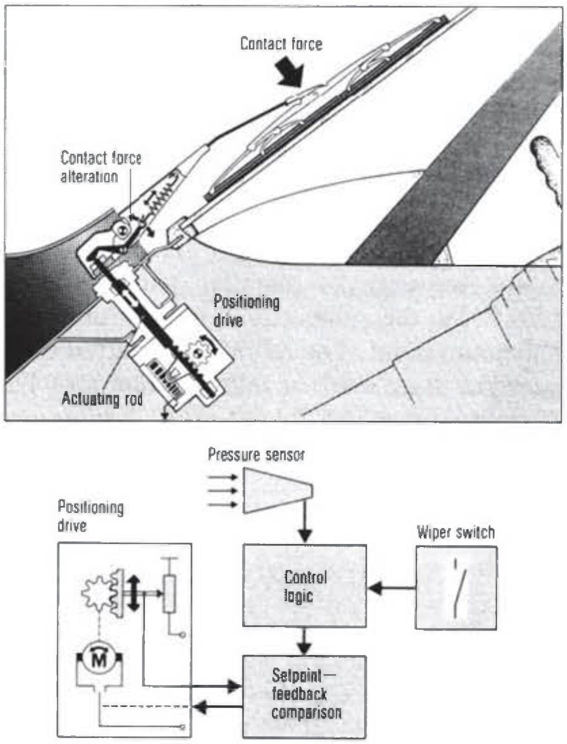


Figure 12.15 Wiper blade pressure control system

The system is able to respond very quickly such that, when overtaking, the deluge of spray is cleared by increased pressure and also, if the screen dries off, the pressure is reduced to prevent scraping.

12.1.10 Linear wiper systems

Car makers are constantly looking for ways to reduce the noise generated by wiper systems. The two main sources of noise are the wiper blade (particularly when it turns over at the end of each movement) and the wiper motor.

Key fact

An H-bridge uses four power devices that are connected to reverse the voltage across both terminals of a load. This is used to control the direction of a motor.

An interesting rear wiper module offering an original solution to these problems in the form of a specific, integrated electronics control system, has been developed. This system is designed around an H-bridge power stage, which has no relays. This eliminates all switching noise. The control algorithm provides pinpoint management of wiper speed; it slows the blades at the end of each cycle, thus cutting out turning noise.

Current wiper systems that are based on an alternative rotary movement cover a wipe-area of between 50 and 60% of the total surface area of the rear window. This limit is due to the height/width ratio and the curve of the window. The linear rear wiper concept ensures optimum visual comfort as it covers over 80% of the rear window surface; this is a visibility gain for the driver exceeding 60%.

This increase in the driver's field of vision enhances safety, especially during low-speed manoeuvres such as reversing or parking. The linear rear wiper concept is in keeping with the trend towards narrower, highly convex rear windows and can be fully integrated into vehicle design (Figure 12.16).

Figure 12.16 Linear rear wiper system (Source: Valeo)

Key fact

Indicator flashing rate must be between one and two per second with a relative 'on' time of between 30 and 57%.

12.2 Signalling circuits

12.2.1 Introduction

Direction indicators (turn signals) have a number of statutory requirements. The light produced must be amber (the exception being some USA vehicle where the red stoplight is used as the turn signal). The indicators may be grouped with other lamps. The flashing rate must be between one and two per second with a relative 'on' time of between 30 and 57%. If a fault develops, this must be apparent to the driver by the operation of a warning light on the dashboard. The fault can be indicated by a distinct change in frequency of operation, or the warning light remaining on. If one of the main bulbs fails then the remaining lights should continue to flash perceptibly.

Legislation exists as to the mounting position of the exterior lamps, such that the rear indicator lights must be within a set distance of the tail lights and within a set height. The wattage of the indicator light bulbs is normally 21 W at 6, 12 or 24 V as appropriate, except where LEDs are used.

Brake or stop lights fall under the heading of auxiliaries or 'signalling'. Circuits are examined later in this section.

12.2.2 Flasher units

Figure 12.17 shows the internal circuit of an electronic flasher unit. The operation of this unit is based around an integrated circuit. The type shown can operate at least four 21 W bulbs (front and rear) and two 5 W side repeaters when operating in hazard mode. This will continue for several hours if required. Flasher units are rated by the number of bulbs they are capable of operating. When towing a trailer or caravan the unit must be able to operate at a higher wattage. Most units use a relay for the actual switching as this is not susceptible to voltage spikes and also provides an audible signal.

The electronic circuit is constructed together with the relay, on a printed circuit board. Very few components are used as the integrated circuit is specially designed for use as an indicator timer. The integrated circuit itself has three main sections: the relay driver, an oscillator and a bulb failure circuit. A Zener diode is built in to the IC to ensure constant voltage such that the frequency of operation

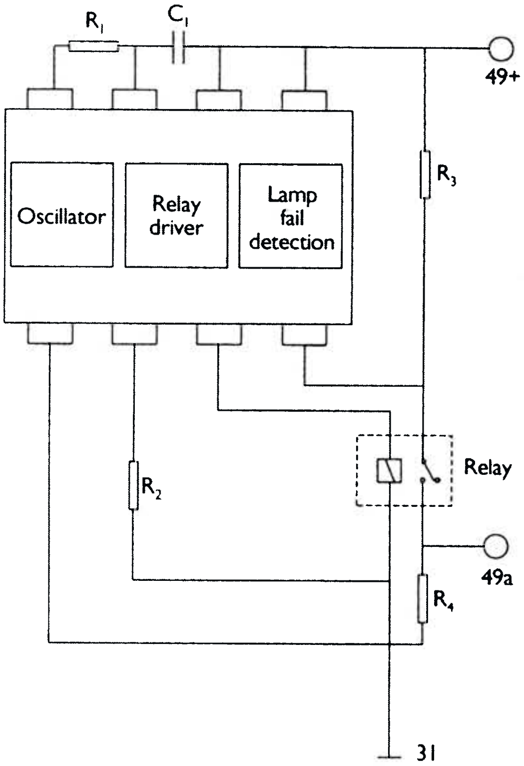


Figure 12.17 Circuit diagram of an electronic flasher unit

will remain constant in the range 10–15 V. The timer for the oscillator is controlled by R1 and C. The values are normally set to give an on–off ratio of 50% and an operating frequency of 1.5 Hz (90 per minute).

The on–off signals produced by the oscillator are passed to a driver circuit, which is a Darlington pair with a diode connected to protect it from back-EMF as the relay coil is switched on and off. Bulb failure is recognized when the volt drop across the low value resistor R2 falls. The bulb failure circuit causes the oscillator to double the speed of operation. Extra capacitors can be used for added protection against transient voltages and for interference suppression. Figure 12.18 shows typical ‘packaging’ for a flasher unit.

Key fact

The integrated circuit itself has three main sections: the relay driver, an oscillator and a bulb failure circuit.

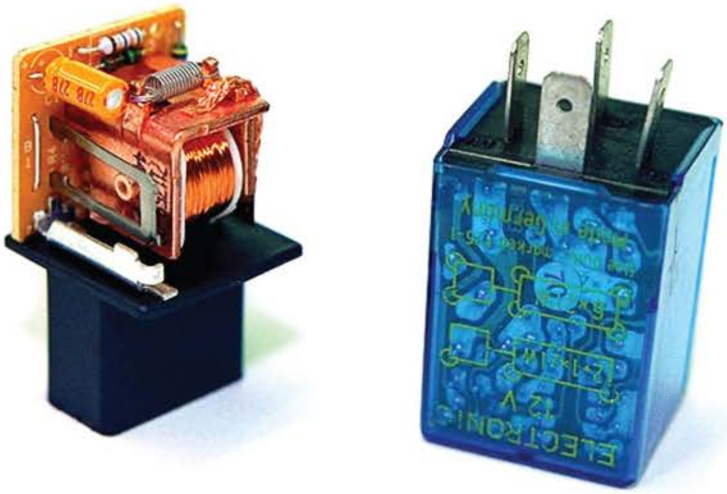


Figure 12.18 Electronic flasher unit

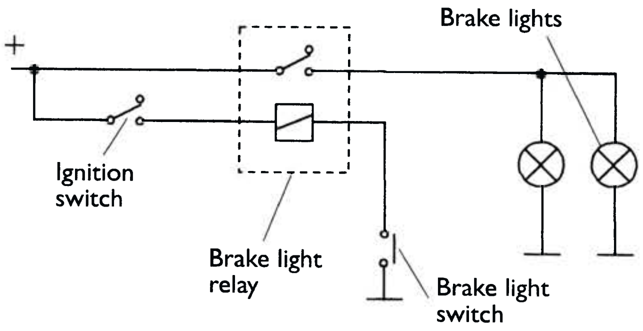


Figure 12.19 Typical brake light circuit

12.2.3 Brake lights

Figure 12.19 shows a typical brake light circuit. Most incorporate a relay to switch the lights, which is in turn operated by a spring-loaded switch on the brake pedal. Links from this circuit to cruise control may be found. This is to cause the cruise control to switch off as the brakes are operated.

12.2.4 Indicators and hazard circuit

The circuit diagram shown in Figure 12.20 shows the full layout of a standard indicator and hazard lights wiring system. Note how the hazard switch, when operated, disconnects the ignition supply from the flasher unit and replaces it with a constant supply. The hazard system will therefore operate at any time but the indicators will only work when the ignition is switched on. When the indicator switch is operated left or right, the front, rear and repeater bulbs are connected to the output terminal of the flasher unit, which then operates and causes the bulbs to flash.

When the hazard switch is operated, five sets of contacts are moved. Two sets connect left and right circuits to the output of the flasher unit. One set disconnects the ignition supply and another set connects the battery supply to the unit. The final set of contacts causes a hazard warning light to be operated. On this and most vehicles the hazard switch is illuminated when the sidelights are switched on.

Key fact

When the hazard switch is operated, five sets of contacts are moved.

When operating in hazard mode the indicator bulbs would draw 7.8 A (94 W/12 V). However, this current will peak much higher due to the cold resistance of the bulbs. In the circuit shown, the top fuse is direct from the battery and the other is ignition controlled.

With the ignition switched on, fuse 1 in the passenger compartment fuse box provides a feed to the hazard warning switch on the G wire. Provided the hazard warning switch is in the off position the feed crosses the switch and supplies the flasher unit on the LG/K wire. When the switch control is moved for a right turn, the switch makes contact when the LG/N wire from the flasher unit is connected to the G/W wire, allowing a supply to pass the right-hand front and rear indicator lights and then to earth on the B wire. When the switch control is moved for a left turn, the switch makes contact with the G/R wire, which allows the supply to pass to the left-hand front and rear indicator lights and then to earth on the B wire. The action of the flasher unit causes the circuit to ‘make and break’.

By pressing the hazard warning switch a battery supply on the N/O from fuse 3 (1.4, 2.0 and diesel models) or 4 (1.6 models) in the engine bay fusebox crosses

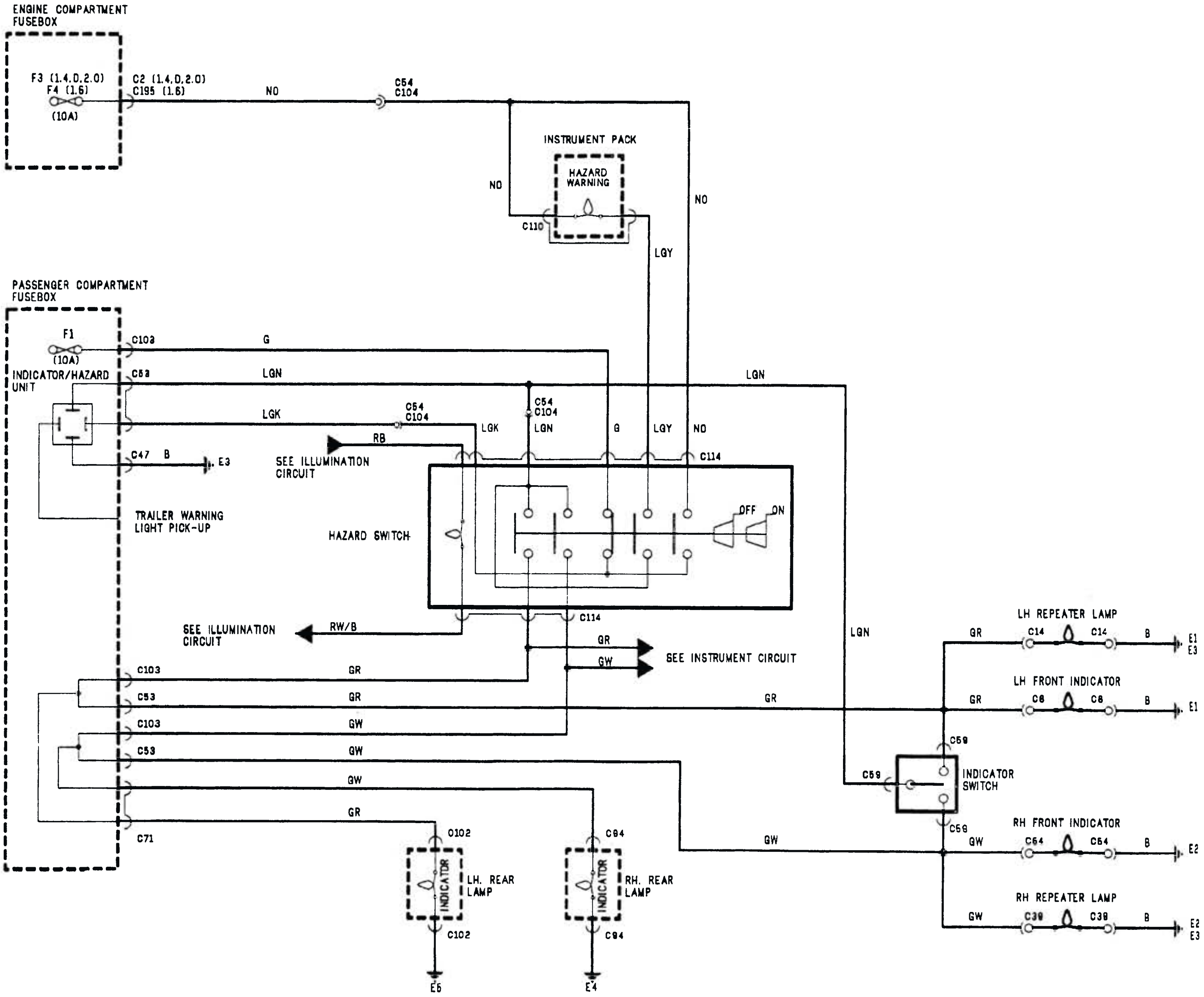


Figure 12.20 Indicator and hazard circuit

the switch and supplies the flasher unit on the LG/K wire. At the same time contacts are closed to connect the hazard warning light and the flasher unit to both the G/W and GIR wires, the right-hand and left-hand indicators and the warning light flash alternately.

12.3 Other auxiliary systems

12.3.1 Electric horns

Regulations in most countries state that the horn (or audible warning device) should produce a uniform sound. This consequently makes sirens and melody-type fanfare horns illegal! Most horns draw a large current, so are switched by a suitable relay.

The standard horn operates by simple electromagnetic switching. As current flow causes an armature that is attached to a tone disc to be attracted to a stop, a set

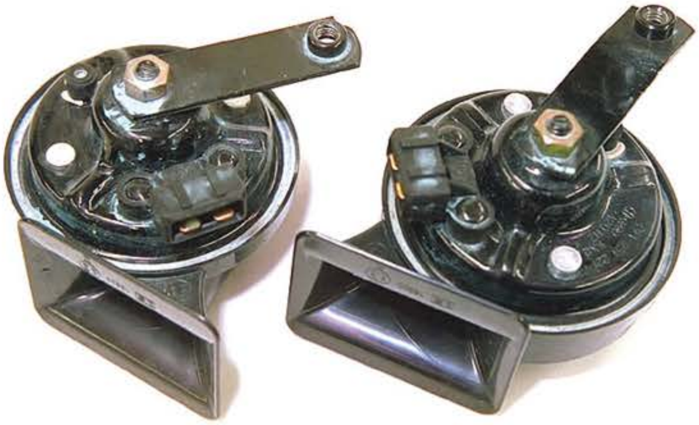


Figure 12.21 Horns

Key fact

The standard horn operates by simple electromagnetic switching.

of contacts is opened. This disconnects the current allowing the armature and disc to return under spring tension. The whole process keeps repeating when the horn switch is on. The frequency of movement and hence the fundamental tone is arranged to lie between 1.8 and 3.5 kHz. This gives good penetration through traffic noise. Twin horn systems, which have a high and low tone horn, are often used. This produces a more pleasing sound but is still very audible in both town and higher speed conditions. Figure 12.21 shows two typical horns and with the associated circuit is Figure 12.22.

12.3.2 Engine cooling fan motors

Key fact

When twin cooling fans and motors are fitted, they can be run in series or parallel.

Most engine cooling fan motors (radiator cooling) are simple permanent magnet types. Figure 12.23 shows a typical example. The fans used often have the blades placed asymmetrically (balanced but not in a regular pattern) to reduce noise when operating.

When twin cooling fans and motors are fitted, they can be run in series or parallel. This is often the case when air conditioning is used as the condenser is usually placed in front of the radiator and extra cooling air speed may be needed. A circuit for series or parallel operation of cooling fans is shown in Figure 12.24.

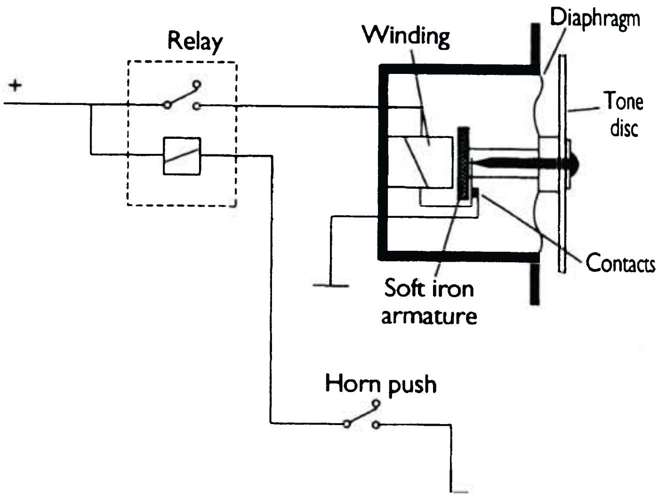


Figure 12.22 Horn and circuit

Figure 12.23 Cooling fan motor

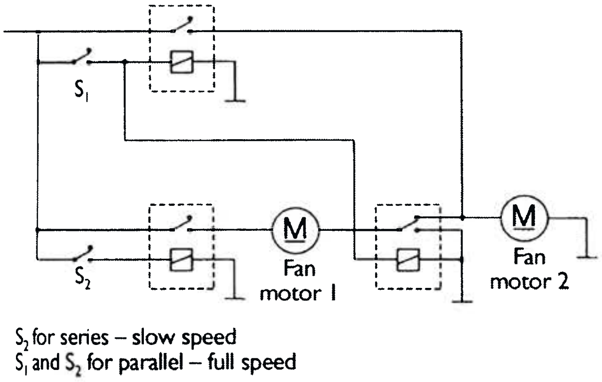


Figure 12.24 Circuit for series or parallel operation of cooling fans

12.3.3 Headlight wipers and washers

There are two ways in which headlights are cleaned, first by high pressure jets, and secondly by small wiper blades with low pressure water supply. The second method is, in fact, much the same as windscreen cleaning but on a smaller scale. The high pressure system tends to be favoured but can suffer in very cold conditions due to the fluid freezing. It is expected that the wash system should be capable of about 50 operations before refilling of the reservoir is necessary. Figure 12.25 shows the pressure wash technique.

Headlight cleaners are often combined with the windscreen washers. They operate each time the windscreen washers are activated, if the headlights are also switched on.

A retractable nozzle for headlight cleaners is often used. When the water pressure is pumped to the nozzle it pushes the nozzle from its retracted position, flush with the bodywork. When the washing is completed the jet is retracted back into the housing.

Figure 12.25 Headlight washers in action

12.3.4 Other circuits

Some minor vehicle electrical systems, which are not covered elsewhere, are shown in Figure 12.26. Cigar lighter, clock, rotating beacon and electric aerial are all circuits that could be used by many other systems.

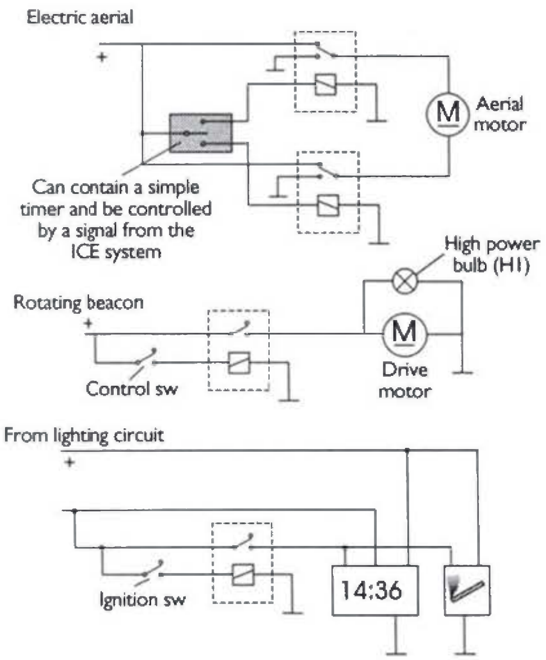


Figure 12.26 Electric aerial, rotating beacon, cigar lighter and clock circuit

12.3.5 Diagnosing auxiliary system faults

The process of checking an auxiliary system circuit is broadly as follows.

1. Hand and eye checks (loose wires, loose switches and other obvious faults) – all connections clean and tight.
2. Check battery (see Chapter 5) – must be 70% charged.
3. Check motor linkage/bulbs – visual check.
4. Fuse continuity – (do not trust your eyes) voltage at both sides with a meter or a test lamp.
5. If used does the relay click (if yes, jump to stage 8) – this means the relay has operated, but it is not necessarily making contact.
6. Supply to switch – battery volts.
7. Supply from the switch – battery volts.
8. Supplies to relay – battery volts.
9. Feed out of the relay – battery volts.
10. Voltage supply to the motor – within 0.5 V of the battery.
11. Earth circuit (continuity or voltage) – 0Ω or 0V.

Table 12.1 lists some common symptoms of an auxiliary system malfunction together with suggestions for the possible faults.

12.4 Advanced auxiliary systems technology

12.4.1 Wiper motor torque calculations

The torque required to overcome starting friction of each wiper blade, can be calculated as follows:

$$T = F\mu_{\max} f_s f_t l \left(\frac{\omega_a}{\omega_m}\right)\left(\frac{1}{e}\right)\left(\frac{R_h}{R_c}\right)$$

Table 12.1 Common auxiliary system symptoms and possible faults

Symptom	Possible fault
Horn not working or poor sound quality	Loose or broken wiring/connections/fuse Corrosion in horn connections Switch not making contact High resistance contact on switch or wiring Relay not working
Wipers not working or poor operation	Loose or broken wiring/connections/fuse Corrosion in wiper connections Switch not making contact High resistance contact on switch or wiring Relay/timer not working Motor brushes or slip ring connections worn Limit switch contacts open circuit or high resistance Blades and/or arm springs in poor condition
Washers not working or poor operation	Loose or broken wiring/connections/fuse Corrosion in washer motor connections Switch not making contact Pump motor poor or not working Blocked pipes or jets Incorrect fluid additive used
Indicators not working or incorrect operating speed	Bulb(s) blown Loose or broken wiring/connections/fuse Corrosion in horn connections Switch not making contact High resistance contact on switch or wiring Relay not working
Heater blower not working or poor operation	Loose or broken wiring/connections/fuse Switch not making contact Motor brushes worn Speed selection resistors open circuit

where: T = torque to move one wiper arm, F = force of one blade onto the screen, μ_{max} = max. dry coefficient of friction (e.g. 2.5), f_s = multiplier for joint friction (e.g. 1.15), f_t = tolerance factor (e.g. 1.12), l = wiper arm length, ω_a = max angular velocity of arm, ω_m = mean angular velocity of motor crank, e = efficiency of the motor gear unit (e.g. 0.8), R_h = motor winding resistance – hot, R_c = motor winding resistance – cold.

12.4.2 PM Motor – electronic speed control

The automotive industry uses PM motors because they are economical to produce and provide good performance. A simple current limiting resistor or a voltage regulator can vary the motor's speed. This simple method is often used for motors requiring variable speed control. However, to control the speed of a motor that draws 20 A at full speed and about 10 A at half speed is a problem.

At full speed, the overall motor control system's efficiency around 80%. If the speed is reduced to half the system's efficiency drops to 40%. This is because there would be a heat loss of 70 W in the series resistor and 14 W lost in the motor. A more efficient speed control system is therefore needed.

One way is to interrupt the motor's voltage at a variable duty cycle using a switching power supply. A system known as pulse width modulation (PWM) has been developed. An introduction to this technique follows.

Because the armature of the PM motor acts as a flywheel, the voltage interruption rate can be 1 kHz or slower, without causing the motor speed to pulsate. A problem at this or other audible frequencies is the noise generated from within the motor. At higher frequencies, 16 kHz for example, the audible noise is minimised. A further noise problem is significant EMR. This is generated by the fast switching speeds. This can be improved by slowing down the switching edge of the operating signal. A compromise has to be made between the edge speeds and power device heat loss.

When the EMR problems are safely contained, the stalled motor condition must be considered. The motor's copper windings have a positive temperature coefficient of 0.00393 °C. Therefore, a 0.25 Ω motor resistance value at 25 °C would be about 0.18 Ω at –40 °C. Using a typical 20 A motor as the load, the maximum stalled or locked rotor current can be calculated to be about 77 A as shown:

$$I_{\max} = \frac{E_{\max}}{R_{mtr}}$$

where: E_{max} *= maximum power supply voltage (14.4 V),* R_{mtr} *= minimum motor resistance (0.18 Ω).*

When the maximum motor current has been calculated, the specifications of the power transistor can be determined. In this case, the device needs an average current rating of at least 77 A. However, a further consideration for reliable power transistor operation is its worst case heat dissipation.

The worst case includes maximum values for the supply voltage, ambient temperature and motor current. A junction temperature of 150 °C for the power transistors is used as a maximum point. The following equation calculates the transistor's maximum allowable heat dissipation for use in an 85 °C environment using a 2.7 °C/W heatsink and a 1 °C/W junction to case power FET thermal resistance.

$$PD_{\max} = \frac{TJ_{\max} - TA_{\max}}{R\phi JC + R\phi CS + R\phi SA}$$

where: TJ_{max} *= maximum allowable junction temperature (150 °C),* TA_{max} *= maximum ambient temperature (85 °C),* $R\phi JC$ *= junction to case thermal resistance (1 °C/W),* $R\phi CS$ *= case to heatsink interface thermal resistance (0.1 °C/W),* $R\phi SA$ *= heatsink to ambient thermal resistance (2.7 °C/W).*

Using the figures given, results in a value of about 17.1 W. This is considerably better than using a dropping resistor but to achieve this several power transistors would have to be connected in parallel. Significant heat sinking is also necessary.

This technique may become popular because of its significant improvement in efficiency over conventional methods and the possibilities for greater control over the speed of a PM motor.

CHAPTER 13

Instrumentation

13.1 Gauges and sensors

13.1.1 Introduction

The topic of instrumentation has now reached such a level as to have become a subject in its own right. However, this chapter covers some of the basic principles of the science, with examples as to how it relates to automobile systems. By definition, an instrumentation system can be said to convert a ‘variable’, into a readable or usable display. For example, a fuel level instrument system will display, often by an analogue gauge, a representation of the fuel in the tank.

Definition

An instrumentation system can be said to convert a ‘variable’, into a readable or usable display.

Instrumentation is not always associated with a gauge or a read-out type display. In many cases the whole system can be used just to operate a warning light. However, the system must still work to certain standards, for example if a low outside temperature warning light did not illuminate at the correct time, a dangerous situation could develop.

This chapter will cover vehicle instrumentation systems in use and examine in more detail the issues involved in choosing or designing an instrumentation system. Chapter 2 contains many details associated with sensors, an integral part of an instrumentation system, and it may be appropriate to refer back for some information related to this chapter.

13.1.2 Sensors

In order to put some limit on the size of this section, only electrical sensors associated with vehicle use will be considered. Sensors are used in vehicle applications for many purposes; for example, the coolant temperature thermistor is used to provide data to the engine management system as well as to the driver via a display.

Figure 13.1 Typical instrument panel

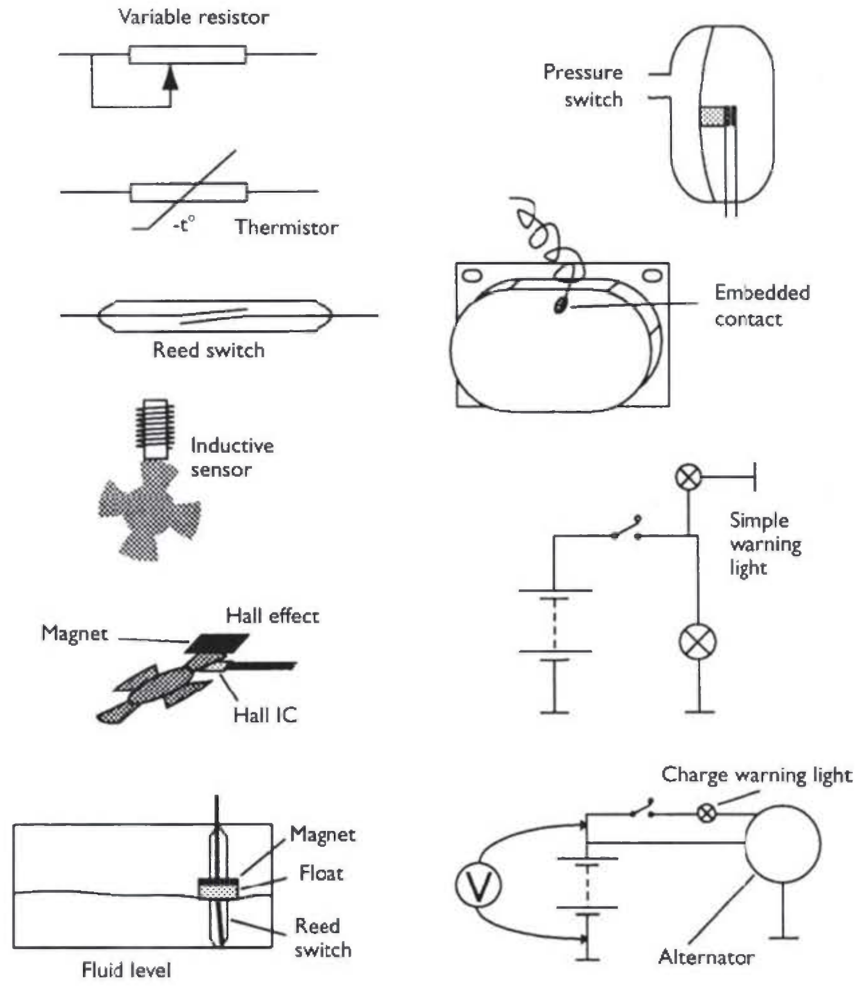


Figure 13.2 Sensors used for instrumentation

For the purpose of providing information to the driver, Table 13.1 gives a list of measurands (things that are measured) together with typical sensors, which is representative of today's vehicles. Figure 13.2 shows some of the sensors listed in Table 13.1.

Table 13.1 Measurements and sensors

Measurement required	Sensor example
Fuel level	Variable resistor
Temperatures	Thermistor
Bulb failure	Reed relay
Road speed	Inductive pulse generator
Engine speed	Hall effect
Fluid levels	Float and reed switch
Oil pressure	Diaphragm switch
Brake pad wear	Embedded contact wire
Lights in operation	Bulb and simple circuit
Battery charge rate	Bulb circuit / voltage monitor

Figure 13.3 Variable resistance fuel tank unit

13.1.3 Thermal-type gauges

Thermal gauges, which are ideal for fuel and engine temperature indication, have been in use for many years. This will continue because of their simple design and inherent 'thermal' damping. The gauge works by utilizing the heating effect of electricity and the benefit of the widely adopted bimetal strip. As a current flows through a simple heating coil wound on a bimetal strip, heat causes the strip to bend. The bimetal strip is connected to a pointer on a suitable scale. The amount of bend is proportional to the heat, which in turn is proportional to the current flowing. Providing the sensor can vary its resistance in proportion to the measurand (e.g. fuel level), the gauge will indicate a suitable representation providing it has been calibrated for the particular task. Figure 13.5 shows a representation of a typical thermal gauge and Figure 13.3 shows a typical sensor..

Key fact

Thermal gauges have inherent 'thermal' damping.

The inherent damping is due to the slow thermal effect on the bimetal strip. This causes the needle to move very slowly to its final position. It can be said to have a large time constant. This is a particular advantage for displaying fuel level, as the variable resistor in the tank will move, as the fuel moves, due to vehicle movement! If the gauge were able to react quickly it would be constantly moving.

Figure 13.4 Fuel gauge

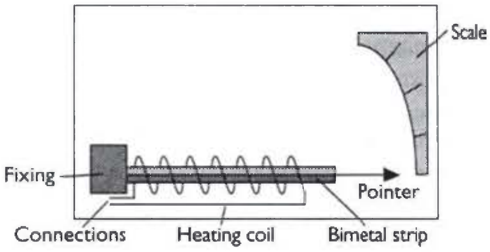


Figure 13.5 Bimetal strip operation in a thermal-type gauge

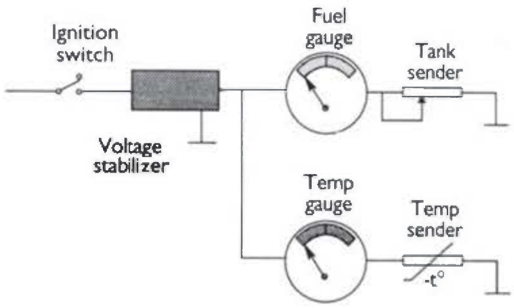


Figure 13.6 Traditional fuel and temperature gauge circuits

The movement of the fuel however is, in effect, averaged out and a relatively accurate display can be obtained. Some electronically driven thermal fuel gauges are damped even more by the control system.

Thermal-type gauges are used with a variable resistor and float in a fuel tank or with a thermistor in the engine water jacket. Figure 13.6 shows the circuit of these two together. The resistance of the fuel tank sender can be made non-linear to counteract any non-linear response of the gauge. The sender resistance is at a maximum when the tank is empty.

A constant voltage supply is required to prevent changes in the vehicle system voltage affecting the reading. This is because, if the system voltage increased, the current flowing would increase and hence the gauges would read higher. Most voltage stabilizers are simple Zener diode circuits, as shown in Figure 13.7.

Key fact

A constant voltage supply is required to prevent changes in the vehicle system voltage affecting the readings of thermal gauges.

13.1.4 Moving iron gauges

The moving iron gauge was in use earlier than the thermal type but is now gaining popularity for some applications. Figure 13.8 shows the circuit and principle of the moving iron gauge system. Two small electromagnets are used which act upon a small soft iron armature connected to a pointer. The armature will position itself between the cores of the electromagnets depending on the magnetic strength of each. The ratio of magnetism in each core is changed as the linear variable resistance sender changes and hence the needle is moved.

This type of gauge reacts very quickly (it has a small time constant) and is prone to swing about with movement of the vehicle. Some form of external damping can be used to improve this problem. Resistor R_1 is used to balance out the resistance of the tank sender. A good way to visualize the operation of the circuit is to note that when the tank is half full, the resistance of the sender will be the same as the resistance of R_1. This makes the circuit balanced and the gauge will read half full. The sender resistance is at a maximum when the tank is full.

Key fact

The ratio of magnetism in each core of a moving iron gauge is changed as the linear variable resistance sender changes and hence the needle is moved.

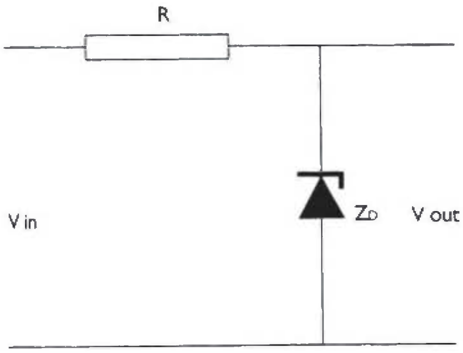


Figure 13.7 A voltage stabilizer

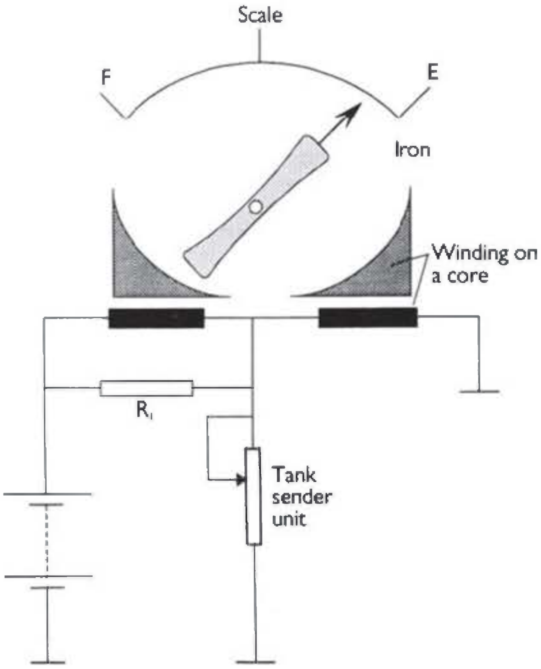


Figure 13.8 Circuit/principle of the moving iron gauge

13.1.5 Air-cored gauges

Air-cored gauges work on the same principle as a compass needle lining up with a magnetic field. The needle of the display is attached to a very small permanent magnet. Three coils of wire are used and each produces a magnetic field (Figure 13.9). The magnet will line up with the resultant of the three fields. The current flowing and the number of turns (ampere-turns) determine the strength of the magnetic flux produced by each coil. As the number of turns remains constant the current is the key factor.

Key fact

Air-cored gauges work on the same principle as a compass needle lining up with a magnetic field.

Figure 13.10 shows the principle of the air-cored gauge together with the circuit for use as a temperature indicator. The ballast resistor on the left is used to limit maximum current and the calibration resistor is used for calibration. The thermistor is the temperature sender. As the thermistor resistance is increased, the current in all three coils will change. Current through C will be increased but the current in coils A and B will decrease. The resultant magnetic fields are shown in Figure 13.10. This moves the magnetic armature accordingly.

Figure 13.9 Air cored gauge unit

The air-cored gauge has a number of advantages. It has almost instant response and, as the needle is held in a magnetic field, it will not move as the vehicle changes position. The gauge can be arranged to continue to register the last position even when switched off or, if a small ‘pull off’ magnet is used, it will return to its zero position. As a system voltage change would affect the current flowing in all three coils variations are cancelled out, negating the need for voltage stabilization. Note that the operation is similar to the moving iron gauge.

Key fact

The air-cored gauge has almost instant response and, as the needle is held in a magnetic field, it will not move when the vehicle changes position.

Figure 13.11 shows the system used on some vehicles for the temperature gauge. It is an air-cored device with fluid damping. The temperature gauge is fitted with a spiral pull off spring to make the gauge read ‘cold’ when the ignition is switched off. The fuel gauge is very similar but retains its position when the ignition is off.

When the system receives a supply from the ignition the resistance of the thermistor determines the current flowing through the coils. When engine coolant

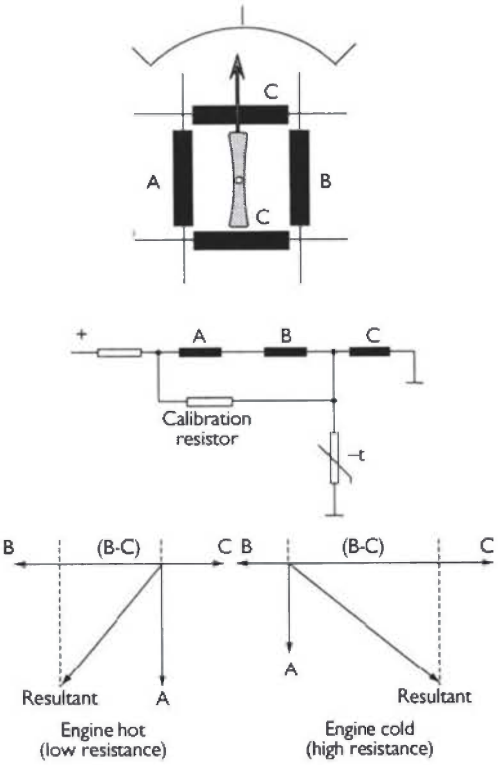


Figure 13.10 Principle of the air-cored gauge together with the circuit when used as a fuel level or temperature indicator and resultant magnetic fields

temperature is low, the resistance of the sender will be high. This will cause the voltage at point X to be higher than that at point Y. This will be above the Zener voltage and so the diode will conduct in its reverse direction. Current will flow through coil A and coil B directly but also a further path will exist through R and

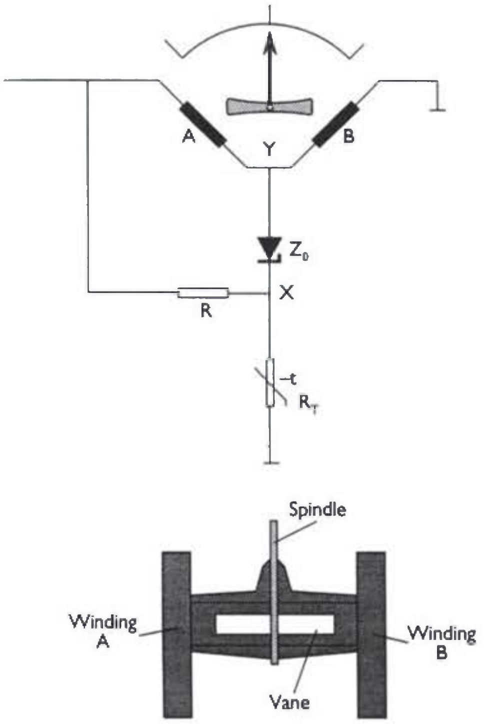


Figure 13.11 Air-cored gauge with fluid damping

the diode, effectively bypassing coil A. This will cause the magnetism of coil B to be greater than coil A, deflecting the magnet and pointer towards the cold side. As the resistance of the sender falls with increasing temperature, the voltage at X will fall, reducing the current through coil B, allowing the needle to rise.

At normal operating temperature, the voltage at X will be just under the Zener diode breakdown voltage. Current through each coil will now be the same and the gauge will read in the centre. If coolant temperature increases further, then current will flow through the diode in its forward direction, thus increasing the current through coil A, which will cause the needle to move to the hot side. Operation of the fuel gauge is similar but a resistor is used in place of the Zener diode. The diode is used to stabilize the gauge when reading ‘normal’ to reduce fluctuations due to thermostat operation.

13.1.6 Other types of gauges

A variation of any of the above types of gauge can be used to display other required outputs, such as voltage or oil pressure. Gauges to display road or engine speed, however, need to react very quickly to change. Many systems now use stepper motors for this purpose although some retain the conventional cable driven speedometers (Figure 13.12).

Key fact

Gauges to display road or engine speed need to react very quickly to change.

Figure 13.13 shows a block diagram of a speedometer, which uses an ammeter as the gauge. This system uses a quenched oscillator sensor that will produce a constant amplitude signal even at very low speed. The frequency of the signal is proportional to road speed. The sensor is driven from the gearbox or a final drive output. The electronic control or signal conditioning circuit consists firstly of a Schmitt trigger, which shapes the signal and suppresses any noise picked up in the wiring.

The monostable is used to produce uniform signals in proportion to those from the pulse generator. The moving coil gauge will read an average of the pulses. This average value is dependent on the frequency of the input signal, which in turn is dependent on vehicle speed. The odometer is driven by a stepper motor, which is driven by the output of a divider and a power amplifier. The divider is to calibrate the action of the stepper motor to the distance covered. The actual

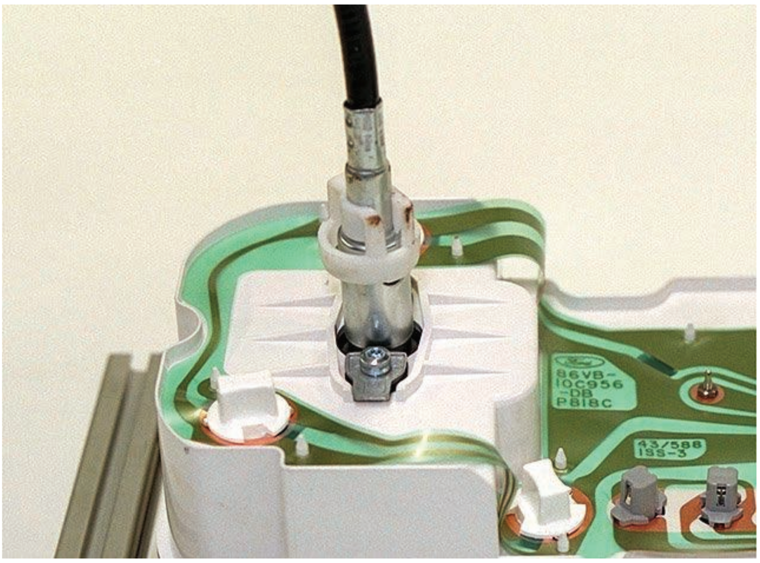


Figure 13.12 Cable driven speedometer

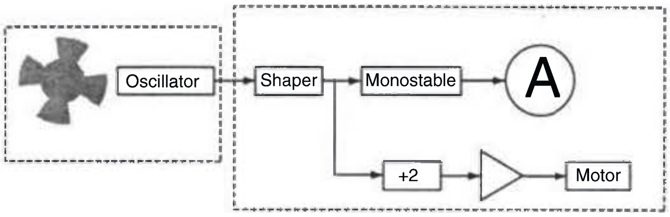


Figure 13.13 Block diagram of a speedometer system which uses a simple ammeter as the gauge

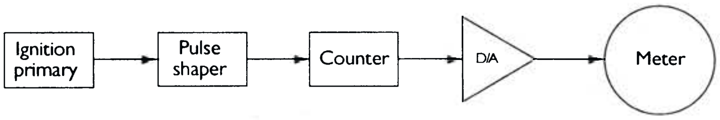


Figure 13.14 Block diagram of a tachometer which uses signals from the ignition coil

speedometer gauge can be calibrated to any vehicle by changing the time delay of the monostable (see Chapter 2).

A system for driving a tachometer is similar to the speedometer system. Pulses from the ignition primary circuit are often used to drive this gauge. Figure 13.14 shows the block diagram of a typical system.

Most modern speedometers and tachometers are now driven by road and engine speed sensors.

13.1.7 A digital instrumentation system

Figure 13.15 shows a typical digital instrumentation system. All signal conditioning and logic functions are carried out in the ECU. This will often form part of the dashboard assembly.

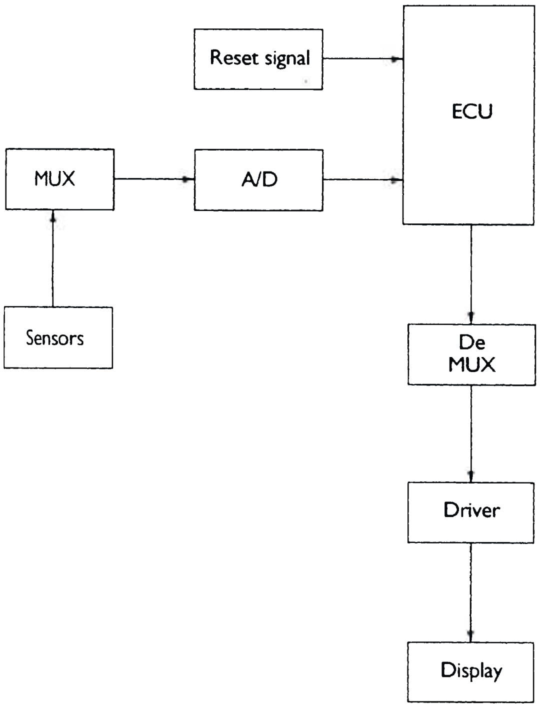


Figure 13.15 Digital instrumentation system

Standard sensors provide information to the ECU, which in turn will drive suitable displays. The ECU contains a ROM section, which allows it to be programmed to a specific vehicle. The gauges used are as described in the above sections. Some of the extra functions available with this system are described briefly as follows.

- Low fuel warning light – can be made to illuminate at a particular resistance reading from the fuel tank sender unit.
- High engine temperature warning light – can be made to operate at a set resistance of the thermistor.
- Steady reading of the temperature gauge – to prevent the gauge fluctuating as the cooling system thermostat operates the gauge can be made to read only at, say, five set figures. For example, if the input resistance varies from 240 to 200 Ω as the thermostat operates, the ECU will output just one reading, corresponding to 'normal' on the gauge. If the resistance is much higher or lower the gauge will read to one of the five higher or lower positions. This gives a low resolution but high readability for the driver.
- Oil pressure or other warning lights can be made to flash – this is more likely to catch the driver's attention.
- Service or inspection interval warning lights can be used – the warning lights are operated broadly as a function of time but, for example, the service interval is reduced if the engine experiences high speeds and/or high temperatures. Oil condition sensors are also used to help determine service intervals.
- Alternator warning light – works as normal but the same or an extra light can be made to operate if the output is reduced or if the drive belt slips. This is achieved by a wire from one phase of the alternator providing a pulsed signal, which is compared to a pulsed signal from the ignition. If the ratio of the pulses changed this would indicate a slipping belt.

As an example of how some of this system works consider the high temperature and low fuel warning lights as examples. Figure 13.16 shows a block diagram of just this part of the overall system.

The analogue to digital converter (ADC) is time division multiplexed to various sensors. The signals from the temperature and fuel level sensors will produce a certain digital representation of a numerical value when they reach say 180 (about 105 °C) and 200 (10 litres left), respectively. These figures (assigned to variables 'temp_input' and 'fuel_input') can then be compared with those

Key fact

The analogue to digital converter (ADC) is time division multiplexed to various sensors.

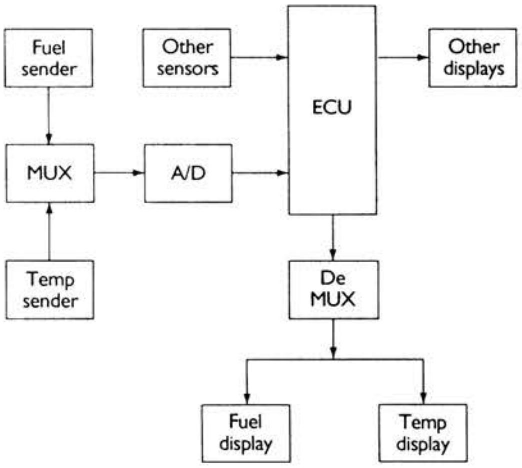


Figure 13.16 Block diagram of high temperature and low fuel warning lights

pre-programmed into memory, variables 'high_temp' and 'low_fuel'. The following simplified lines of computer program indicate the logical result.

IF temp_input > high_temp THEN high_temp_light = on

IF fuel-input > low_fuel THEN low_fuel_light = on

A whole program is built up which can be made suitable for any particular manufacturer's requirements.

13.2 Visual displays

13.2.1 Choosing the best display – readability

Key fact

The function of any visual display is to communicate information to the desired level of accuracy.

The function of any visual display is to communicate information to the desired level of accuracy. Most displays used in the vehicle must provide instant data but the accuracy is not always important. Analogue displays can provide almost instant feedback from one short glance. For example, if the needle of the temperature gauge is about in the middle then the driver can assume that the engine temperature is within suitable limits. A digital read-out of temperature such as 98 °C would not be as easy to interpret. This is a good example as to why even when digital processing and display techniques are used, the actual read-out will still be in analogue form. Figure 13.17 shows a display using analogue gauges.

Figure 13.18 shows an instrument display using digital representation. Numerical and other forms of display are, however, used for many applications. Some of these are as follows:

- Vehicle map.
- Trip computer.
- Clock.
- Radio displays.
- Route finding displays.
- General instruments.

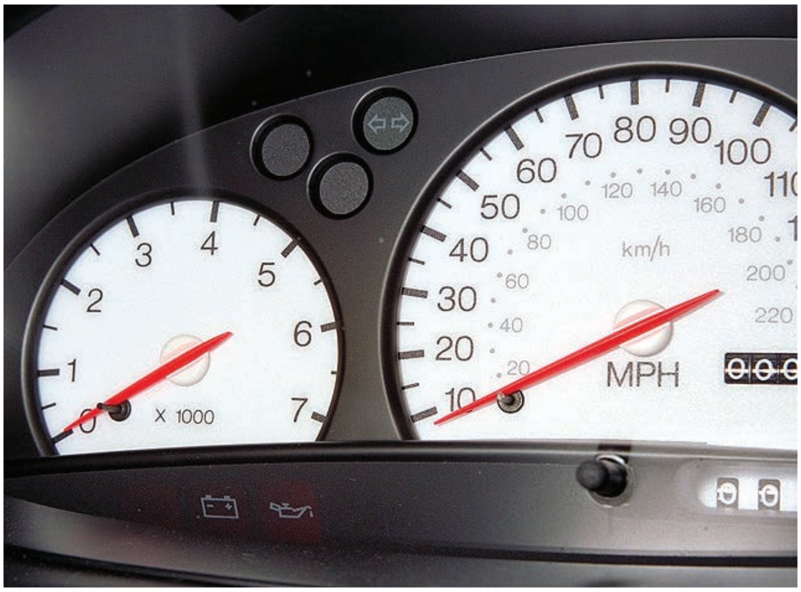


Figure 13.17 Analogue display

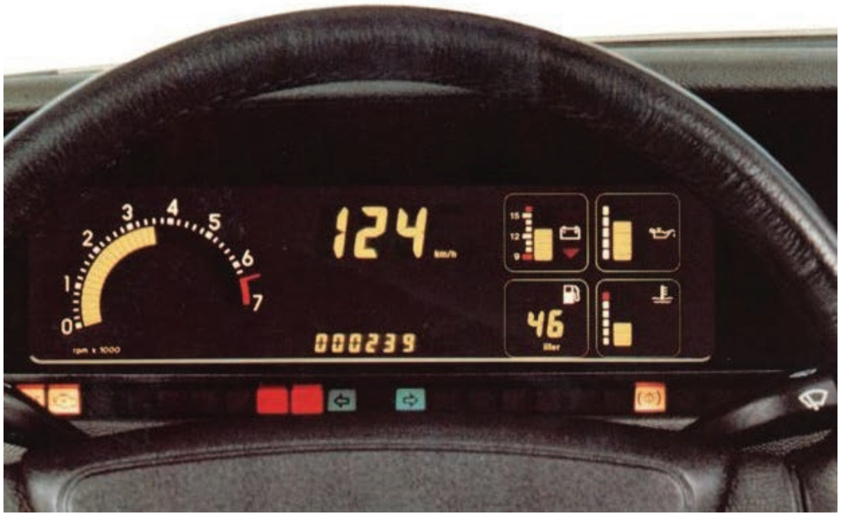


Figure 13.18 Digital display

These displays can be created in a number of ways; the following sections examine each of these in more detail. To drive individual segments or parts of a complete display, a technique called time division multiplexing is often used.

Definition

Time division multiplexing: A transmission method shared in time (rather than frequency). Signals from different sources share a single channel or bus in successive time slots.

13.2.2 Light-emitting diode displays

If the PN junction of a diode is manufactured from gallium arsenide phosphide (GaAsP), light will be emitted from the junction when a current is made to pass in the forward-biased direction. This is a light-emitting diode (LED) and will produce red, yellow or green light with slight changes in the manufacturing process. LEDs are used extensively as indicators on electronic equipment and in digital displays. They last for a very long time (50 000 hours) and draw only a small current.

LED displays are tending to be replaced for automobile use by the liquid crystal type display, which can be backlit to make it easier to read in the daylight. However, LEDs are still popular for many applications.

The actual display will normally consist of a number of LEDs arranged into a suitable pattern for the required output. This can range from the standard seven-segment display to show numbers, to a custom-designed speedometer display. A small number of LED displays are shown in Figure 13.19.

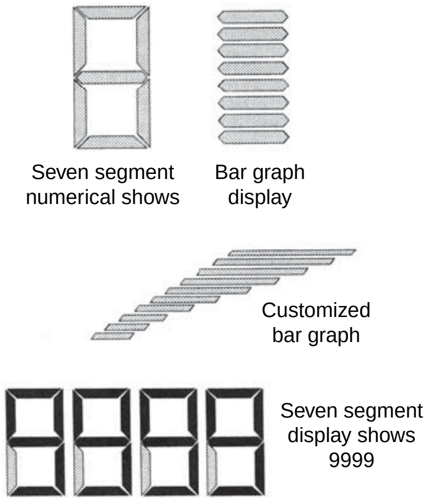


Figure 13.19 LED displays

13.2.3 Liquid crystal displays

Liquid crystals are substances that do not melt directly from a solid to the liquid phase, but first pass through a para-crystalline stage in which the molecules are partially ordered. In this stage, a liquid crystal is a cloudy or translucent fluid but still has some of the optical properties of a solid crystal.

The three main types of liquid crystals are smectic, nematic and cholesteric (twisted nematic), which are differentiated by the alignments of the rod-shaped molecules. Smectic liquid crystals have molecules parallel to one another, forming a layer, but within the layer no pattern exists. Nematic types have the rod-like molecules oriented parallel to one another but have no layer structure. The cholesteric types have parallel molecules, and the layers are arranged in a helical, or spiral, fashion.

Definition

Liquid crystal: A state of matter similar to a liquid, in which the molecules are arranged in one or two dimensions.

Mechanical stress, electric and magnetic fields, pressure and temperature can alter the molecular structure of liquid crystals. A liquid crystal also scatters light that shines on it. Because of these properties, liquid crystals are used to display letters and numbers on calculators, digital watches and automobile instrument displays. LCDs are also used for portable computer screens and even television screens. The LCD has many more areas of potential use and developments are on-going. In particular, this type of display is now good enough to reproduce pictures and text on computer screens and TVs.

One type of display uses the cholesteric type of liquid crystal. This display is achieved by only allowing polarized light to enter the liquid crystal which, as it passes through the crystal, is rotated by 90°. The light then passes through a second polarizer, which is set at 90° to the first. A mirror at the back of the arrangement reflects the light so that it returns through the polarizer, the crystal and the front polarizer again. The net result is that light is simply reflected, but only when the liquid crystal is in this one particular state.

When a voltage of about 10 V at 50 Hz is applied to the crystal, the electric field causes it to become disorganized and the light passing through it is no longer twisted by 90°. This means that the light polarized by the first polarizer will not pass through the second, and will therefore not be reflected. This will show as a dark area on the display.

These areas are constructed into suitable segments in much the same way as with LEDs to provide whatever type of display is required. The size of each individual area can be very small, such as to form one pixel of a TV or computer screen if appropriate. Figure 13.20 shows a representation of how this liquid crystal display works.

Key fact

LCDs use very low power but do require a source of light to operate.

LCDs use very low power but do require a source of light to operate. To be able to read the display in the dark some form of lighting for the display is required. Instead of using a reflecting mirror at the back of the display a source of light known as backlighting can be used. A condition known as DC electroluminescence is an ideal phenomenon. This uses a zinc-sulphide based compound, which is placed between two electrodes in much the same way as the liquid crystal, but it emits light when a voltage is applied. Figure 13.21 shows how this backlighting effect can be used to good effect for display purposes.

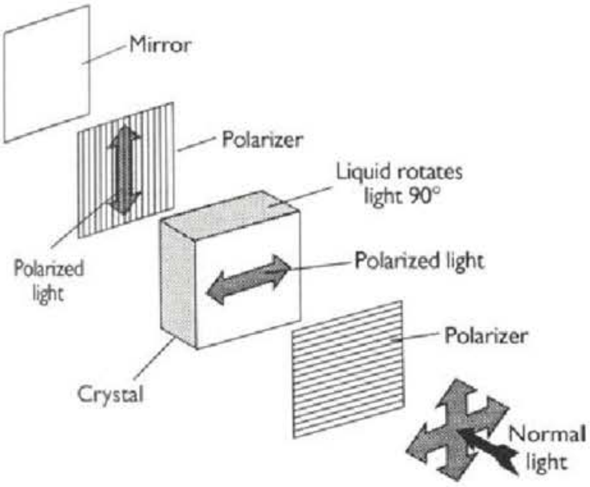


Figure 13.20 Principle of a liquid crystal display

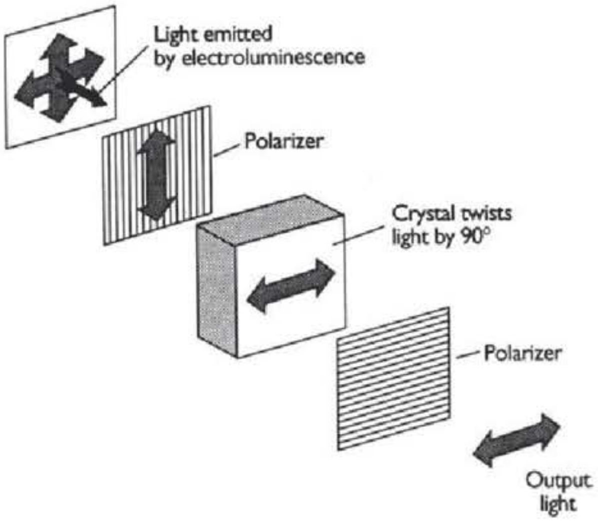


Figure 13.21 Backlighting effect can be used to good effect for display purposes

13.2.4 Vacuum fluorescent displays

A vacuum fluorescent display (VFD) works in much the same way as an old fashioned television tube and screen. It has some benefits for vehicle use because it produces a bright light (which is adjustable) and a wide choice of colours. Figure 13.22 shows that the VFD system consists of three main components. These are the filament, the grid and the screen with segments placed appropriately for the intended use of the display. The filament forms the cathode and the segments the anode of the main circuit. The control grid is used to control brightness as the voltage is altered.

When a current is passed through the tungsten filaments they become red hot (several hundred degrees centigrade) and emit electrons. The whole unit is made to contain a good vacuum so that the electrons are not affected by any outside influence. The segments are coated with a fluorescent substance and connected to a control wire. The segments are given a positive potential to attract the electrons. When electrons strike the segments they fluoresce, emitting a yellow-green or a blue-green light depending on the type of phosphor used to coat the segments. If the potential of the grid is changed, the number of electrons striking the segments can be changed, thus affecting the brightness. If no segments are connected to a supply (often only about 5 V), then all the electrons emitted are

Key fact

When a current is passed through the tungsten filaments they become red hot and emit electrons.

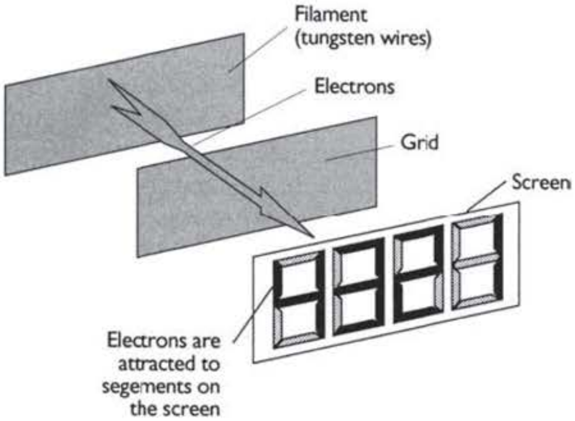


Figure 13.22 Vacuum florescent display

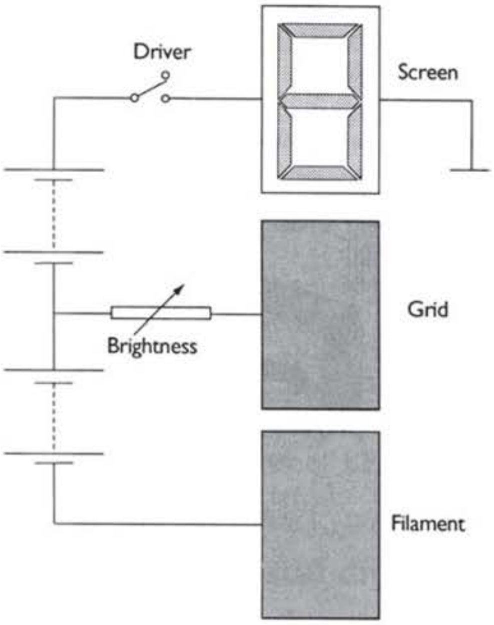


Figure 13.23 Circuit which could be used to control a VFD

stopped at the grid. The grid is also important in that it tends to organize the movement of electrons.

Figure 13.23 shows a circuit used to control a VFD. Note how the potential of the segments when activated is above that of the grid. The driver circuit for this system is much the same, in principle, as any other display, i.e. the electronic control will connect one or more of the appropriate segments to a supply to produce the desired output.

The glass front of the display can be coloured to improve the readability and aesthetic value. This type of display has many advantages but the main problem for automobile use is its susceptibility to shock and vibration. This can be overcome, however, with suitable mountings.

13.2.5 Head-up displays

One of the main problems to solve with any automobile instrument or monitoring display is that the driver has to look away from the road to see the information. Also, in other situations, the driver does not look at the display, and therefore could miss an important warning such as low oil pressure. Many techniques can be used such as warning beepers or placing the instruments almost in view, but one of the most innovative is the head-up display (HUD). This was originally developed by the aircraft industry for fighter pilots; aircraft designers had similar problems in displaying up to 100 different warning devices in an aircraft cockpit. Figure 13.24 shows the principle of a head-up display and Figure 13.25 and example in use. Information from a display projector is directed onto a partially reflecting mirror. The information projected would therefore have to be reversed for this system. Under normal circumstances the driver would be able to see the road through the mirror. The brightness of the display is adjusted to suit ambient lighting conditions. A great deal of data can be presented when this system is computer controlled.

A problem, however, is which information to provide in this way. The speedometer could form part of a lower level display and a low oil pressure

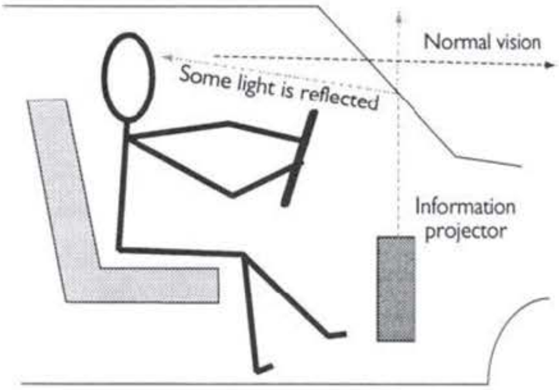


Figure 13.24 Head-up display operation

Figure 13.25 HUD example (Source: BMW)

could cause a flash right in front of the driver. A visual warning could also be displayed when forward-facing radar detects an impending collision. Current HUD systems are for straight-ahead vision, but liquid crystal rear view mirrors, used to dim and cut headlight glare automatically, can be used as an effective display screen for rear facing, blind spot detecting radar.

One interesting study is to determine exactly where the driver is looking at any point in time, which could be used to determine where the head-up display would be projected at any particular time. The technique involves tiny video cameras, coupled to a laser beam that reflects from the cornea of the driver's eye and can measure exactly where he or she is looking. Apart from its use in research, the eye motion detector is one of a series of tools used in biomechanical research that can directly monitor the physical well-being of the driver. Some of these tools could eventually be used actively to control the car or to wake up a driver who is at risk of falling asleep.

13.2.6 Electroluminescent instrument lighting

Electroluminescent backlighting is an interesting technology for the automotive industry because of its thin, uniform lighting characteristics. Electroluminescent (EL) lamps provide a range of exciting opportunities for instrument designers.

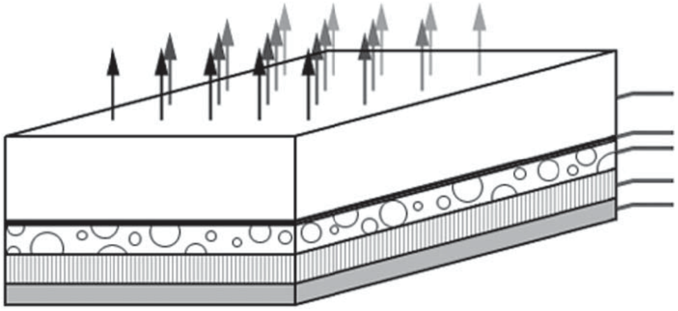


Figure 13.26 Construction of an EL lamp (Source: Durel)

Key fact

An EL lamp is similar to a capacitor; it consists of a dielectric layer and a light-emitting phosphor layer between two conductive plates.

An EL lamp is similar to a capacitor. It consists of a dielectric layer and a light-emitting phosphor layer between two conductive plates. The device needs to be protected from high voltages but the dielectric layer achieves this because it is an insulator. Alternating current (AC) is needed to operate an EL lamp. The AC generates an electric field across the phosphor and dielectric layers. The phosphor electrons are excited by the electric field which causes them to move to a higher energy orbit. When these electrons fall back to a lower orbit, energy is released in the form of light (Figure 13.26).

Polyethylene terephthalate (PET) is used as the base material for many EL lamps. The front electrode is made of indium tin oxide (ITO). The phosphor, dielectric and rear electrode are screen printed over the ITO side of the polyester, which results in a source of light that is thin and flat.

There are a number of benefits to EL lighting:

- Uniformity.
- Durability.
- Flexibility (thin and lightweight).
- Easy to make into different shapes.
- Low power consumption and low heat generation.
- Simple to design.

The other options for instrument lighting are bulbs, light emitting diodes and cold-cathode fluorescent lamps (vacuum fluorescent displays). EL lamps are often superior to these other types, particularly when instruments are considered as a complete system.

A wide range of colours can be created using the EL method. This is achieved by blending combinations of phosphors before screen printing. It is also possible to print selected areas with different phosphors, thus creating a multi-coloured lamp. Typical colours are blue-green, green, yellow-green, white, blue and orange-red.

Because EL lamps need AC to emit light, it is necessary to use an inverter. Typically, the signal used for EL operation is 60 to 150 V_{rms} at a frequency of 300 to 500 Hz. The current draw of the inverter and lit area is only about 1 to 2 mA/cm^2. EL lamps can operate for over 20 000 hours, which usually exceeds the life of the vehicle. For final assembly purposes the EL lamp is essentially a 2.5 mm-thick film that is sandwiched between a backplate and the graphic overlay.

13.2.7 Display techniques summary

Most of the discussion in previous sections has been related to the activation of an individual display device. The techniques used for – and the layout of –

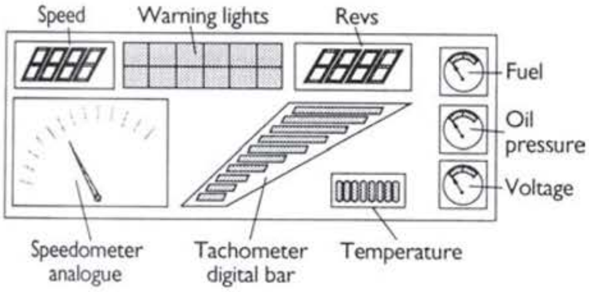


Figure 13.27 Displays which combine some of the devices discussed

dashboard or display panels are very important. To a great extent this again comes back to readability. When so many techniques are available to the designer it is tempting to use the most technologically advanced. This, however, is not always the best. It is prudent to ask the one simple question: what is the most appropriate display technique for this application? Figure 13.27 shows a display that combines some of the devices discussed previously.

Many of the decisions regarding the display need to be based on readability as well a driver preference. Personally, I find numerical display of vehicle speed or engine rpm irritating; others will, I am sure, disagree. Even the bar graph displays are not as good as simple analogue needles in my opinion.

The layout and the way that instruments are combined is an area in which much research has been carried out. This relates to the time it takes the driver to gain the information required when looking away from the road to glance at the instrument pack. Figures 13.28–13.30 show instrument panels and other readout displays. Note how compact they are so that the information can be absorbed without the driver having to scan each item in turn. The aesthetic look of the dashboard is also an important selling point for a vehicle. This could be at odds with the best readability on some occasions!

Key fact

The layout and the way that instruments are combined is an area in which much research has been carried out.

As a further example, the new Mercedes S-Class presents itself as an especially user-friendly information centre with innovative solutions for the instrument cluster and central display. In a variable, situation-related display, the driver is able to have only that information displayed which he anticipates and needs. Momentarily unimportant information fades into the background where it will not distract the driver. This new display system incorporates a dial-type gauge and various graphic representations in alternation.

When the night-vision system offered for the first time in the new S-Class is displayed in the instrument cluster, the speedometer appears as a bar at the

Figure 13.28 Analogue instrument panel

Figure 13.29 Mercedes S-Class information cluster classic display (Source: Bosch Media)

bottom of the screen. The driver especially benefits from this under poor visibility conditions.

13.2.8 Instrumentation system faults

Table 13.2 lists some common symptoms of an instrumentation system malfunction together with suggestions for the possible fault. The faults are very generic but will serve as a good reminder.

Figure 13.30 Mercedes S-Class information cluster night-vision display (Source: Bosch Media)

Table 13.2 Common symptoms and possible faults of an instrumentation system malfunction

Symptom	Possible fault
Fuel and temperature gauges both read high or low (thermal type gauges)	Voltage stabiliser
Gauges read full/hot or empty/cold all the time	Short/open circuit sensors Short or open circuit wiring
Individual instruments do not work	Loose or broken wiring/connections/fuse Sender units (sensor) faulty Gauge unit fault (not very common)

The process of checking a thermal gauge fuel or temperature instrument system is broadly as follows.

1. Hand and eye checks (loose wires, loose switches and other obvious faults) – all connections clean and tight.
2. Either fit a known good 200 resistor in place of the temperature sender – gauge should read full.
3. Or short fuel tank sender wire to earth – gauge should read full.
4. Check continuity of wire from gauge to sender – 0 to 0.5 Ω.
5. Check supply voltage to gauge (pulsed 0–12 V on old systems) – 10 V stabilized on most.
6. If all above tests are OK the gauge head is at fault.

13.3 Global Positioning System (GPS)

13.3.1 Introduction

From 1974 to 1979, a trial using six satellites allowed navigation in North America for just four hours per day. This trial was extended worldwide by using eleven satellites until 1982 at which time it was decided that the system would be extended to twenty four satellites, in six orbits, with four operating in each. There

Figure 13.31 Display in a Jaguar

Figure 13.32 If you look really hard you can see the satellites, honest…

are now some thirty one satellites in use. They are set at a height of about 21 000 km (13 000 miles), inclined 55 degrees to the equator and take approximately twelve hours to orbit the Earth. The orbits are designed so that there are always six satellites in view, from most places on the earth (Figure 13.32).

The system was developed by the American Department of Defence. Using an encrypted code allows a ground location to be positioned to within a few centimetres. The signal employed for civilian use is artificially reduced in quality so that positioning accuracy is in the region of 50 m. Some systems however now improve on this and can work down to about 15 m.

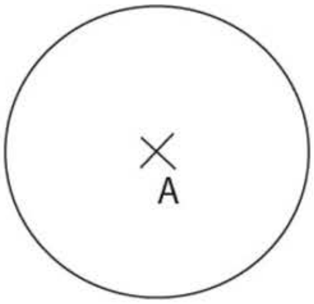


Figure 13.33 At a known distance from a fixed point 'A' you could be anywhere on a circle

13.3.2 Calculating position

Key fact

Position can be worked out using three coordinates and this is known as triangulation.

GPS satellites send out synchronized information fifty times a second. Orbit position, time and identification signals are transmitted. A modern GPS receiver will typically track all of the available satellites, but only a selection of them will be used to calculate position. The times taken for the signals to reach the vehicle are calculated and from this information the computer can determine the distance from each satellite. The current vehicle position can then be worked out using three coordinates. Imagine the three satellites forming a triangle (represented here as A, B, C), the position of a vehicle within that triangle can be determined if the distance from each fixed point (satellite) is known. This is called triangulation (Figures 13.33–13.35).

The GPS receiver gets a signal from each GPS satellite. The satellites transmit the exact time that the signals are sent. By subtracting the time the signal was transmitted from the time it was received, the GPS can tell how far it is from each satellite. The GPS receiver also knows the exact position in the sky of the satellites, at the moment they sent their signals. So, given the travel time of the GPS signals from three satellites and their exact position in the sky, the GPS receiver can determine position in three dimensions – east/west, north/south and altitude.

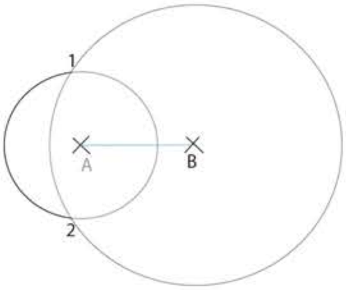


Figure 13.34 At a known distance from two fixed points 'A and B' you must be at position 1 or 2

To calculate the time the GPS signals took to arrive, the GPS receiver needs to know the time very accurately. The GPS satellites have atomic clocks that keep very precise time, but it is not feasible to equip a GPS receiver with such a device. However, if the GPS receiver uses the signal from a fourth satellite it can solve an equation that lets it determine the exact time, without needing an atomic clock.

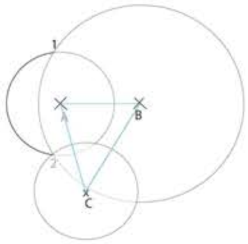


Figure 13.35 At a known distance from three fixed points 'A, B and C' then you must, in this case, be at position 2

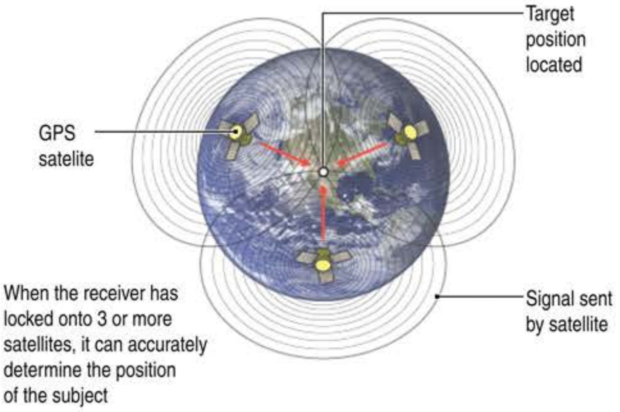


Figure 13.36 How GPS works

If the GPS receiver is only able to get signals from three satellites, position can still be calculated, but less accurately. If only three satellites are available, the GPS receiver can get an approximate position by making the assumption that you are at mean sea level. If you really are at sea level, the position will be reasonably accurate, but if you are driving in the mountains, the two dimensional fix could be several hundreds of metres out (Figures 13.36 and 13.37).

Figure 13.37 Four satellites determine vehicle position more accurately

13.3.3 Sensors

The magnetic field sensor in a GPS unit is a key component (Figure 13.38). It determines direction of travel in relation to the Earth's magnetic field. It also senses the changes in direction when driving round a corner or a bend. The two crossed measuring coils sense changes in the Earth's magnetic field because it has a different effect in each of them. The direction of the Earth's field can be calculated from the polarity and voltage produced by these two coils. The smaller excitation coil produces a signal that causes the ferrite core to oscillate. The direction of the Earth's magnetic field causes the signals from the measuring coils to change depending on the direction of the vehicle.

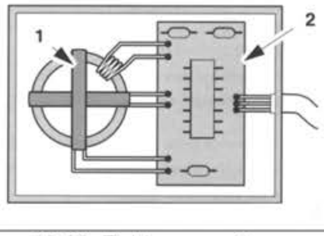


Figure 13.38 Field sensor: 1, crossed coils; 2, control circuit (Source: Ford Motor Company)

13.3.4 Data input and output

To use most satellite navigation systems, the destination address is entered using a joystick control, cursor keys or something similar. The systems usually 'predict' the possible destination as letters are entered so it is not usually necessary to enter the complete address. Once the destination is set the unit will calculate the journey. Options may be given for the shortest or quickest routes at this stage. Driving instructions, relating to the route to be followed, are given visually on the display and audibly through speakers.

13.3.5 Accuracy

Even though the satellite information only provides a positional accuracy of 30–50 m, using dead-reckoning, intelligent software can still get the driver to their destination with an accuracy of about 5 m in some cases. Dead-reckoning means that the vehicle position is determined from speed and turn signals. Different systems do vary however.

The computer can update the vehicle position from the GPS data, by using the possible positions on the stored digital map. This is because in many places on

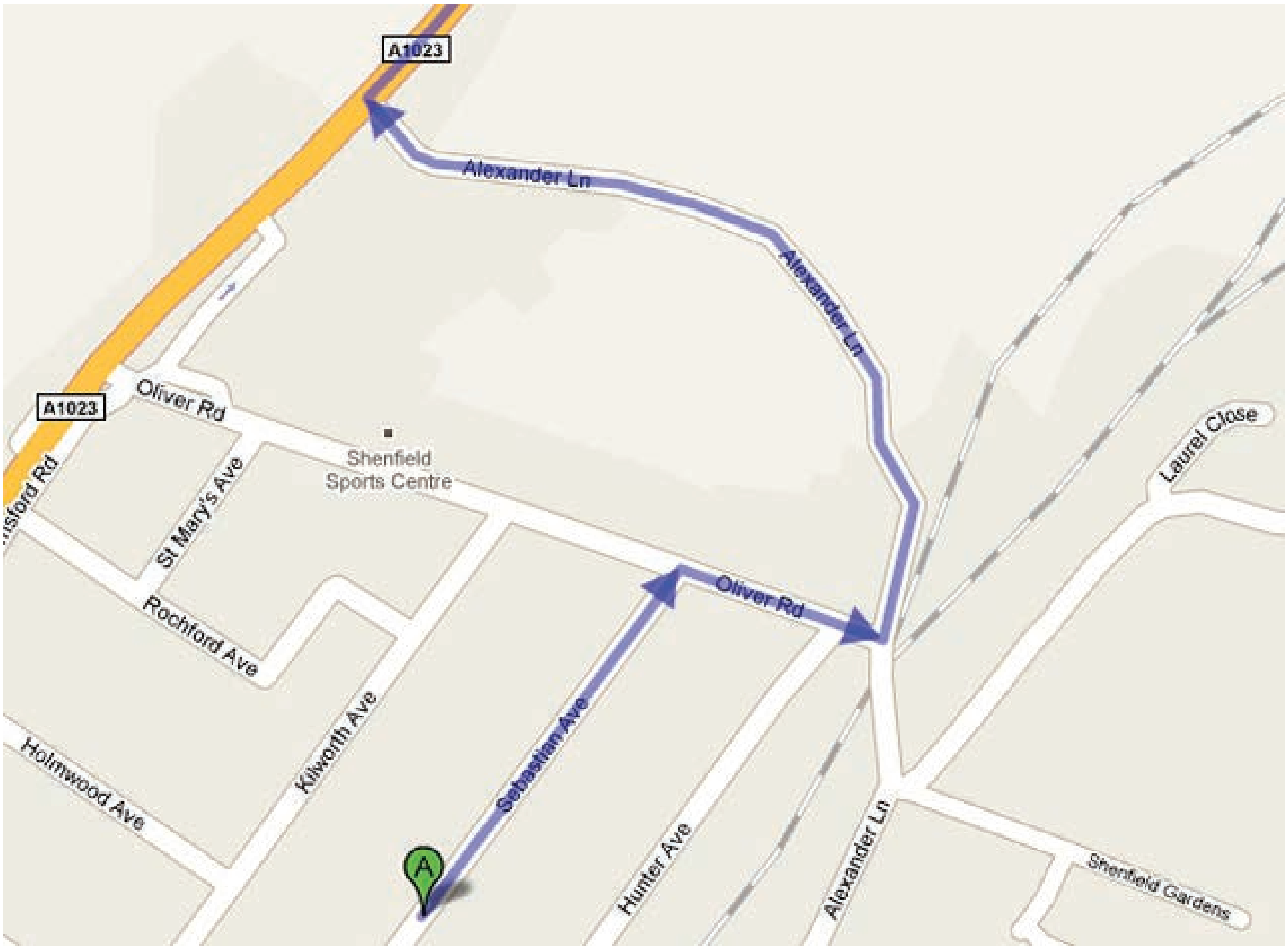


Figure 13.39 As the driver follows the instructions for the first right turn here, the system will 'know' the location to within a metre or so if steering angle is used as in input

Key fact

Dead-reckoning allows basic navigation when satellite signals are disrupted, in a long tunnel for example.

the map only one particular position is possible – it is assumed that short cuts across fields are not taken! Dead-reckoning even allows navigation when satellite signals are disrupted.

Vehicle global positioning systems use a combination of information from satellites and sensors to accurately determine the vehicle position on a digital map. A route can then be calculated to a given destination. Like all vehicle systems, GPS continues to develop and will do for some time yet as more features are added to the software. Already it is possible to 'ask' many systems for the nearest fuel station or restaurant for example.

13.4 Driver information

13.4.1 Vehicle condition monitoring

VCM or vehicle condition monitoring is a form of instrumentation. It has now become difficult to separate it from the more normal instrumentation system discussed in the first part of this chapter. The complete VCM system can

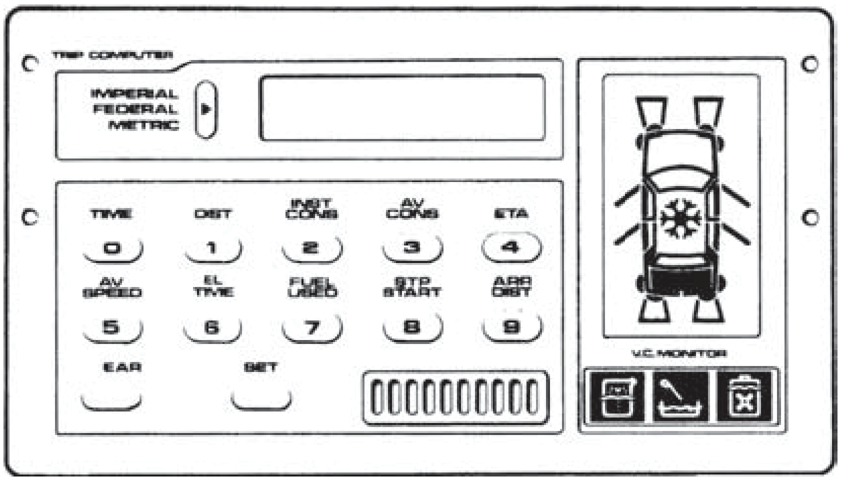


Figure 13.40 Earlier type trip computer display and vehicle 'map'

include driver information relating to the following list of systems that can be monitored.

- High engine temperature.
- Low fuel.
- Low brake fluid.
- Worn brake pads.
- Low coolant level.
- Low oil level.
- Low screen washer fluid.
- Low outside temperature.
- Bulb failure.
- Doors, bonnet/hood or boot open warning.

The circuit shown in Figure 13.42 can be used to operate bulb failure warning lights for whatever particular circuit it is monitoring. The simple principle is that the reed relay is only operated when the bulb being monitored is drawing current.

Figure 13.41 Simplified display on a BMW (MY 2009) showing time, temperature, range, mileage and trip distance

Definition

MY: Model year.

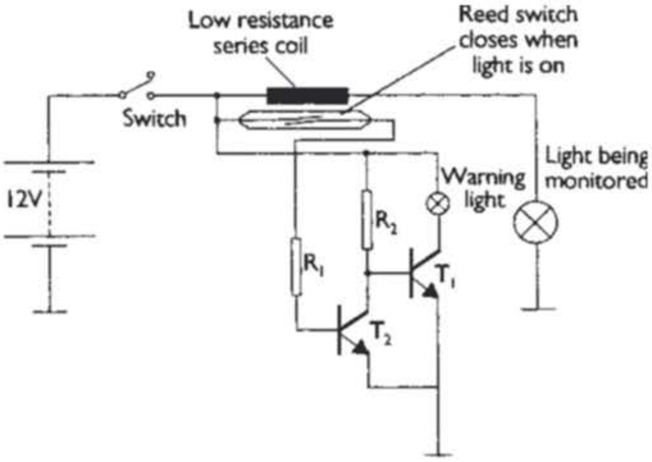


Figure 13.42 Bulb failure warning circuit

Key fact

Many of the circuits monitored use a dual resistance system so that the circuit itself is also checked.

The fluid and temperature level monitoring systems work in a similar way to the systems described earlier but in some cases the level of a fluid is monitored by a float and switch.

Oil level can be monitored by measuring the resistance of a heated wire on the end of the dipstick. A small current is passed through the wire to heat it. How much of the wire is covered by oil will determine its temperature and therefore its resistance.

Many of the circuits monitored use a dual resistance system so that the circuit itself is also checked. Figure 13.43 shows the equivalent circuit for this technique. In effect, it will produce one of three possible outputs:

- high resistance (switch open);
- low resistance (switch closed);
- out-of-range (short or open circuit).

The high or low resistance readings are used to indicate say, correct fluid level and low fluid level. A figure outside these limits would indicate a circuit fault of either a short or open circuit connection.

The display is often just a collection of LEDs or a back lit LCD. These are arranged into suitable patterns and shapes such as to represent the circuit or system being monitored. An open door will illuminate a symbol that looks like the door of the vehicle map (plan view of the car) is open. Low outside temperature or ice warning is often a large snowflake.

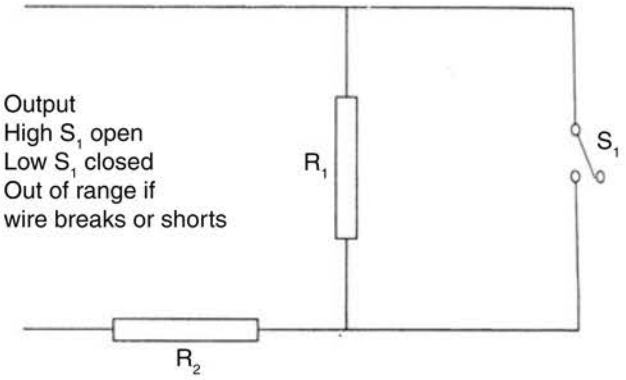


Figure 13.43 Equivalent circuit of a dual resistance self-testing system

Table 13.3 Inputs to a driver information system

Input	Source
Clock signal	Crystal oscillator
Vehicle speed	Speed sensor or instruments ECU
Fuel being used	Injector open time or flow meter
Fuel in the tank	Tank sender unit
Mode/Set/Clear	Data input by the driver

13.4.2 Trip computer

The trip computer was traditionally only fitted to top of the range vehicles. However, many features are now built in to standard cars. The functions available on many systems are:

- Time and date.
- Elapsed time or a stop watch.
- Estimated time of arrival.
- Average fuel consumption.
- Range on remaining fuel.
- Trip distance.

The above details can usually be displayed in imperial, US or metric units as required. In order to calculate the above outputs the inputs to the system shown in Table 13.3 are required.

Figure 13.44 shows a block diagram of a trip computer system. Note that several systems use the same inputs and that several systems 'communicate' with each other. This makes the overall wiring very bulky – if not complicated. This type of interaction and commonality between systems has been one of the reasons for the development of multiplexed wiring techniques (see Chapter 3).

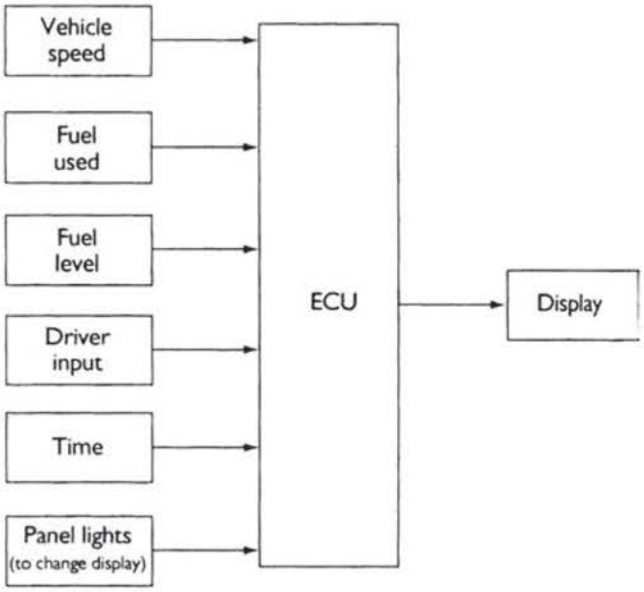


Figure 13.44 Layout of a typical trip computer system

13.5 Advanced instrumentation technology

13.5.1 Multiplexed displays

Key fact

In order to drive even a simple seven-segment display, at least eight wiring connections are required.

In order to drive even a simple seven-segment display, at least eight wiring connections are required. This would be one supply and seven earths (one for each segment). This does not include auxiliary lines required for other purposes, such as backlighting or brightness. To display three seven-segment units, up to about 30 wires and connections would be needed.

To reduce the wiring, time division multiplexing is used. This means that the individual display unit will only be lit during its own small time slot. From Figure 13.45 it can be seen that, if the bottom connection is made at the same time as the appropriate data is present on the seven input lines, only one seven-segment display will be activated. This is carried out for each in turn, thousands of times a second and the human eye does not perceive a flicker.

Key fact

The electronic control unit samples each input in turn in its own time slot, and outputs to the appropriate display again in a form suitable for the display device used.

The technique of multiplexing is taken a stage further by some systems, in that one digital controller carries out the whole of the data or signal processing. Figure 13.46 shows this in block diagram form. The technique is known as data sampling. The electronic control unit samples each input in turn in its own time slot, and outputs to the appropriate display again in a form suitable for the display device used. The electronics will contain a number of A/D and D/A converters and these will also be multiplexed where possible.

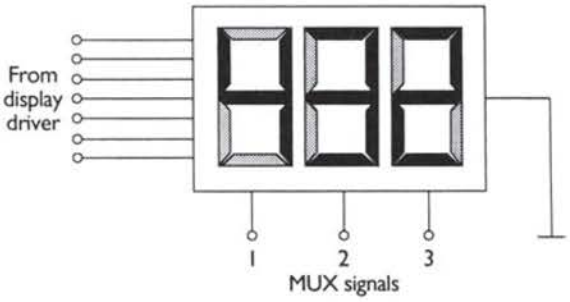


Figure 13.45 Time divisions multiplexing is used so the individual display unit will only be lit during its own small time slot

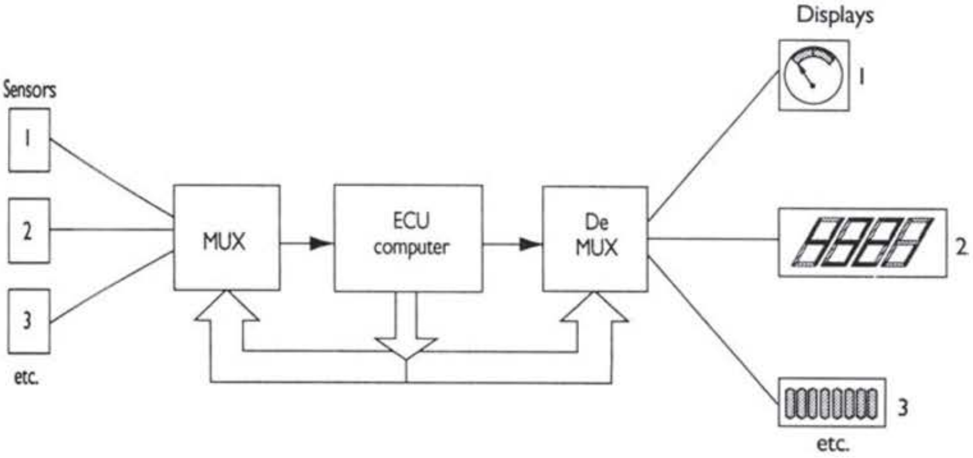


Figure 13.46 Block diagram showing how multiplexing is taken a stage further by some systems

13.5.2 Quantization

When analogue signals are converted to digital, a process called quantization takes place. This could be described as digital encoding. Digital encoding breaks down all data into elementary binary digits (bits), which enable it to be processed, stored, transmitted and decoded as required by computer technologies.

Definition

Quantization: The process of transforming a continuous signal into one of finite steps or levels.

The value of an analogue signal changes smoothly between zero and a maximum. This infinitely varying quantity is converted to a series of discrete values of 0 or 1 by a process known as quantization. The range of values from zero to the maximum possible is divided into a discrete number of steps or quantization levels. The number of steps possible depends on the bit size of the word the digital processors can deal with. For an 8-bit word, the range can be divided into 256 steps (2^8), i.e. from 00000000_2 to 11111111_2. These digital 'samples' should always be taken at more than twice the frequency of the analogue signal to ensure accurate reproduction.

Key fact

11111111_2: The small '2' here means the number is base 2 (binary).

Quantization introduces an error into the process, as each value is 'rounded' to the nearest quantization level. The greater the number of quantization levels the more accurate the process will be, but obviously, increased accuracy involves more bits being used to define the increased number of levels.

13.5.3 Holography

A holographic image is a three-dimensional representation of the original subject. It can be created by splitting a laser beam into object and reference beams. These beams produce an interference pattern, which can be stored on a plate or projected on to a special screen. Research is currently on-going towards using holography to improve night driving safety. Information from infrared cameras can be processed, and then an enhanced holographic image can be projected onto a vehicle windscreen. See also section 4.9.2 for more information about a vision enhancement method.

13.5.4 Telemetry

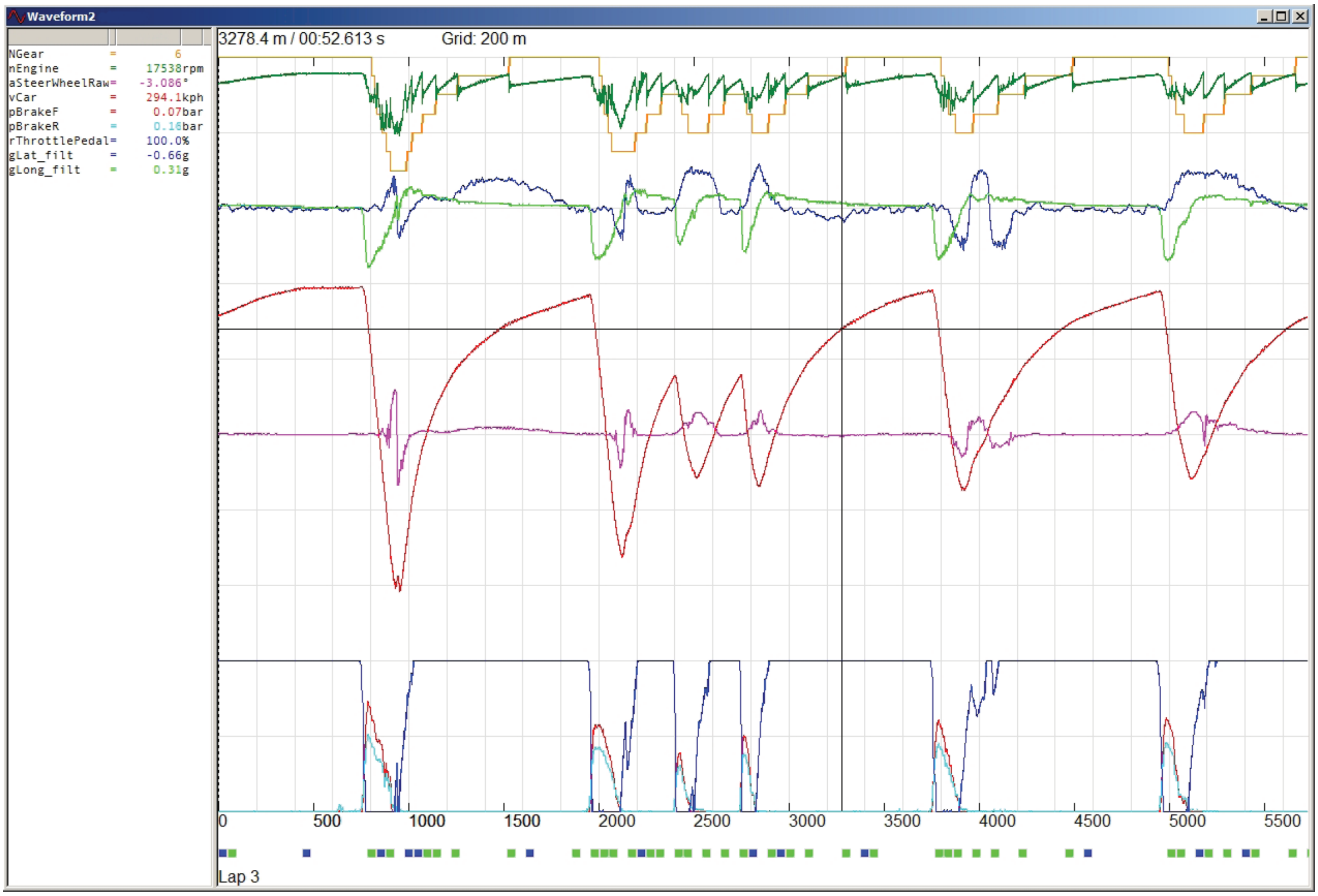
Telemetry is a technology that allows remote measurement and transmission of information. The Greek root of the word is that *tele* means remote, and *metron* means measure. Telecommand is, in a way, a response to telemetry, as it means sending a command or instruction.

Definition

Telemetry: remote measurement and transmission of information.
Telecommand: sending a command or instruction.

Telemetry is used extensively in motorsport and in particular F1, because it allows engineers to collect and analyse a huge amount of data during a race. The data can be interpreted and used to ensure that the car and driver are performing at their optimum. F1 systems have advanced such that even the potential lap time of the car can be calculated. Examples of operating data collected from an F1 car:

- acceleration (G force) in all 3 axis;
- temperature readings (brakes, tyres, engine, transmission, etc.);
- wheel speed;
- suspension displacement;
- hydraulic pressure;
- tyre pressures;
- track position.



```
NGear          =         6
nEngine        =   17538rpm
aSteerWheelRaw=    -3.086°
vCar           =   294.1kph
pBrakeF        =    0.07bar
pBrakeR        =    0.16bar
rThrottlePedal=     100.0%
gLat_filt      =    -0.66g
gLong_filt     =     0.31g
```

Figure 13.47 ATLAS screenshot showing operating data (Source: McLaren Electronic Systems)

Driver inputs are also recorded so that the team can assess performance and, in the case of an accident, the FIA can determine or rule out driver error as a possible cause. Examples of driver inputs:

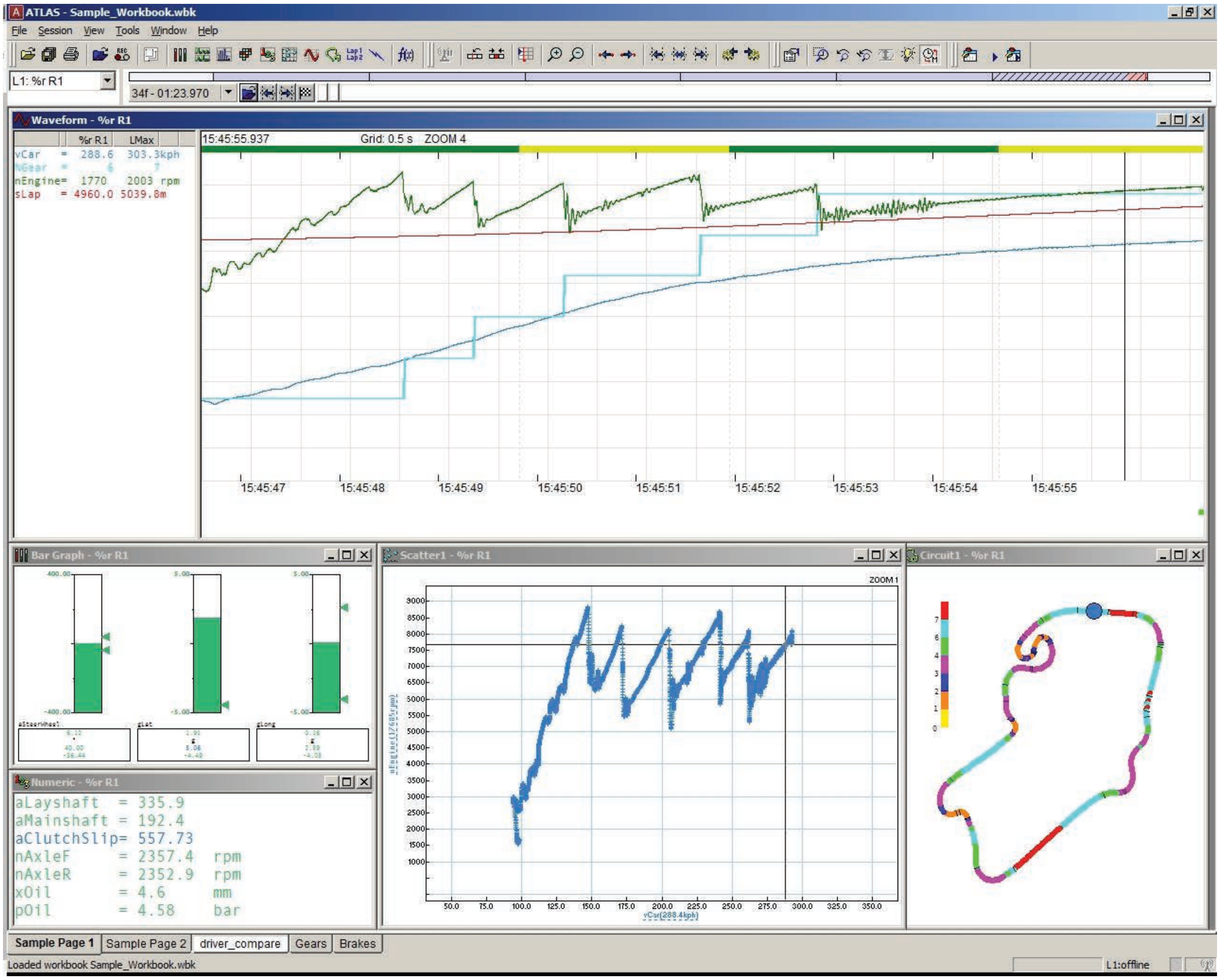
- brake pedal movement;
- throttle pedal movement;
- steering angle;
- gear position.

Two-way telemetry (telemetry and telecommand) is possible and was originally developed by three different companies. The system started as a way to send a message to the electronic systems allowing the race engineers to update the car in real time, for example, changing engine mapping. However, the FIA banned two-way telemetry from F1 in 2003.

McLaren Electronic Systems have developed a system called the Advanced Telemetry Linked Acquisition System (ATLAS). This system displays graphs of each of the cars' systems in real time as the car travels round the track. Since the standard ECU (SECU) was introduced in 2008 almost all Formula One teams use the ATLAS data analysis software.

An F1 car can use two types of telemetry:

- real time information, which is sent in small packets. Hundreds of channels can be transmitted including track position, sensor readings and much more;



```
          %r R1   LMax
vCar   =  288.6  303.3kph
nGear  =      6        7
nEngine=  1770   2003 rpm
sLap   = 4960.0 5039.8m
```

```
aLayshaft   = 335.9
aMainshaft  = 192.4
aClutchSlip= 557.73
nAxleF      = 2357.4  rpm
nAxleR      = 2352.9  rpm
xOil        = 4.6     mm
pOil        = 4.58    bar
```

```
Loaded workbook Sample_Workbook.wbk
```

Figure 13.48 ATLAS screenshot showing the track and a range of data and displays (Source: McLaren Electronic Systems)

- a data burst, which is sent as the car passes the pits.

With ATLAS, this data burst is to make sure the engineers have all the data in case there were any gaps due to poor coverage. Some other systems have to use the data burst to gather more detailed information. The telemetry is transmitted by a small aerial located in the car sidepod (or on the chassis centreline) to a large aerial often fitted to the truck.

Additional data can be downloaded from the car or a test bed by connecting directly to a computer. This is referred to as cable data.

ATLAS uses a single computer to process and distribute the data. A bank of computers is used to allow many engineers to analyse the data. The software displays the information on screens in a way that can be quickly and easily interpreted by the engineers. During a race, readings such as engine temperature and hydraulic pressure are examined in detail to make sure that a major failure is not imminent. If any readings go above or below what is normally expected, the engineers can radio the driver and, for example, ask them to use less engine revs or brake earlier to try and prevent failure.

Key fact

Telemetry is transmitted by a small aerial located on the racing car.

Figure 13.49 Telemetry aerial in use by Team Lotus

See: http://www.mclarenelectronics.com/Products/All/sw_atlas.asp for more information (McLaren 2011).

13.5.5 Telematics

13.5.5.1 Introduction

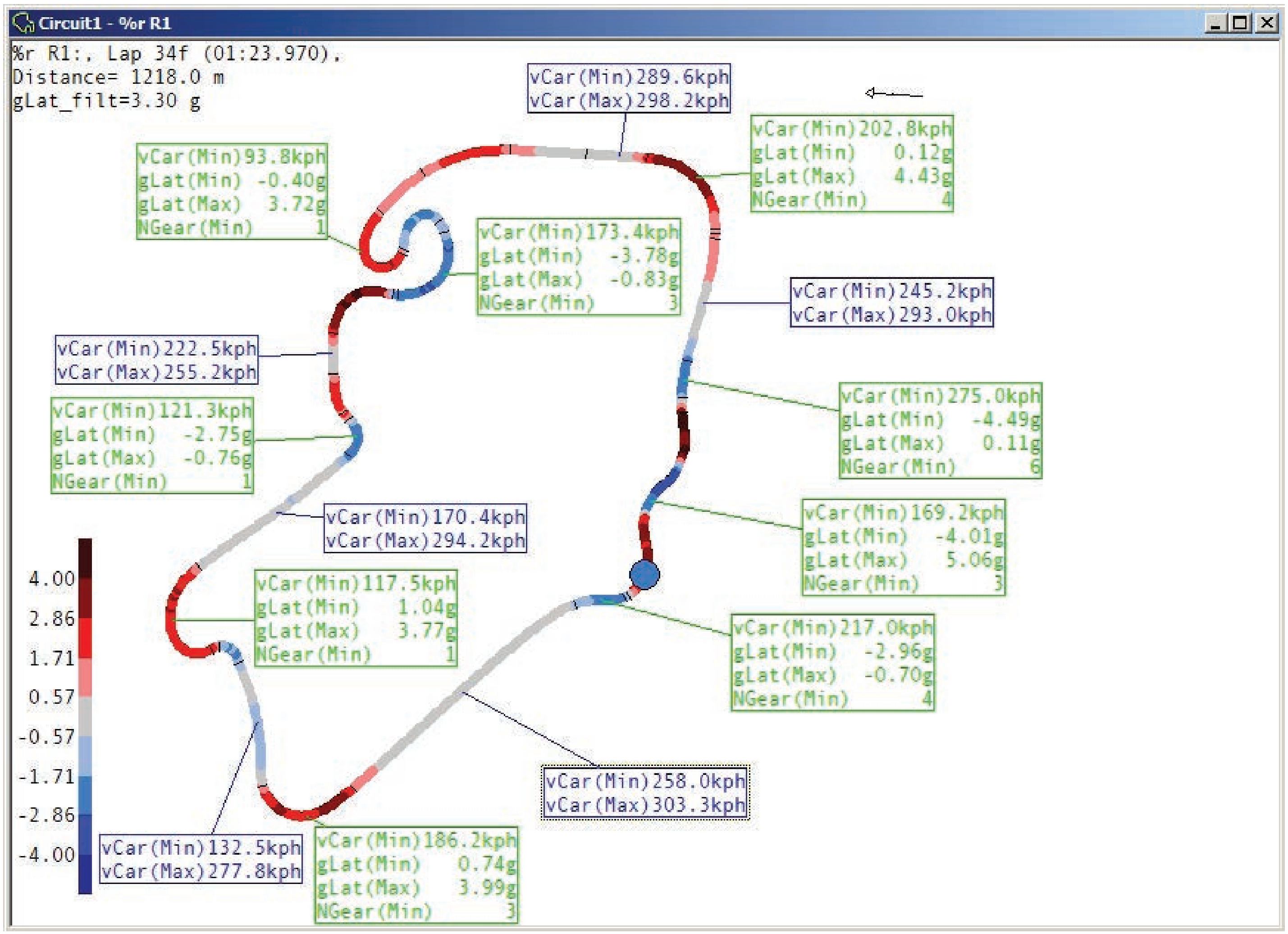
Definition

Telematics: A contraction of telecommunications and informatics.

The word telematics is a contraction of ‘telecommunications’ and ‘informatics’ (or ‘information and communication technology’). Telematics communication is two way, in other words a suitable equipped vehicle can receive and send information. The interface for incoming and outgoing data is a mobile/cellular phone or, more often now, the same technology built in to the vehicle.

The most common use of vehicle telematics is traffic management such as to improve mobility on the available road network. However, there are many other uses and as other technologies converge, more are being developed. Some examples of features and facilities follow:

- Traffic management and (with GPS) associated route planning.
- Vehicle and/or trailer tracking.
- Cold store freight; temperature information for example.



```
%r R1:, Lap 34f (01:23.970),
Distance= 1218.0 m
gLat_filt=3.30 g
```

Figure 13.50 ATLAS screenshot showing the track and live data (Source: McLaren Electronic Systems)

- Fleet management.
- Mobile data.
- Mobile television.
- Emergency/Accident reporting.
- Pay as you drive (PAYD) insurance.
- Remote locking and other remote functions.
- Message services; sending destination information to a car for example.
- Driving monitor; identification of hard braking or cornering too fast for example (Korolov 2011).
- Stolen vehicle recovery.

Some of these points will now be examined further.

Figure 13.51 SOS button in the roof of a car for use in the event of an accident

13.5.5.2 Traffic telematics

Telematics in this context is all about improving mobility on the road network. As well as informing the driver about traffic on the intended route, the system is able to recalculate the route to miss traffic jams (Figure 13.52). Emergency or breakdown calls can be transmitted by some systems. Breakdown information can also be sent so that diagnostic information can be accessed before the

Figure 13.52 Navigation screen showing the route and additional traffic information

breakdown truck even leaves the garage. Crash sensor information can be used to report accidents automatically.

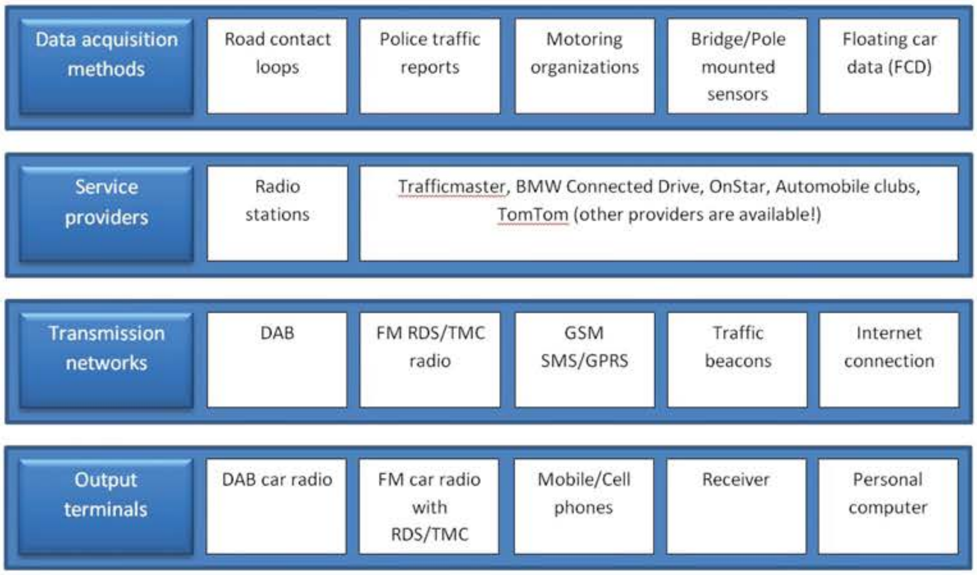
Traffic information is collected, collated and transmitted in a number of different ways as outlined by Figure 13.53. The usefulness of traffic telematics is determined by the quality of information collected. Traffic reports are received either through an FM receiver with RDS decoder, or from the cellular network using SMS.

An interesting way of collecting data is through what is known as a floating car – this is described, unsurprisingly, as floating car data (FCD). As a car move through traffic it periodically transmits to a central point, information on its position, direction of travel, distance travelled and speed. Statistical evaluation of this data generates up to date reports on the traffic situation in that area. This is, at the moment, a voluntary scheme.

Beside the data noted above, there is a whole range of other operating and switching data available in digital form on the bus systems of modern vehicles.

Key fact

Emergency or breakdown calls can be transmitted by some systems.



Data acquisition methods	Road contact loops	Police traffic reports	Motoring organizations	Bridge/Pole mounted sensors	Floating car data (FCD)
Service providers	Radio stations	Trafficmaster, BMW Connected Drive, OnStar, Automobile clubs, TomTom (other providers are available!)			
Transmission networks	DAB	FM RDS/TMC radio	GSM SMS/GPRS	Traffic beacons	Internet connection
Output terminals	DAB car radio	FM car radio with RDS/TMC	Mobile/Cell phones	Receiver	Personal computer

Figure 13.53 Traffic telematics for information acquisition and transmission

(Huber, Lädke et al. 1999) Being available in digital form, they can be registered on the vehicle without undue complexity and used for the process of obtaining traffic and environmental information. These data are referred to as extended floating car data (XFCD). Data from the following sources are of particular interest:

- the windscreen wipers or rain sensor;
- the external thermometer and the air-conditioning system;
- the vehicle's light system (brake and fog lights);
- the hazard warning flashers etc.;
- the sensors for the systems controlling the vehicle dynamics;
- driver assistance systems.

XFCD system may be able to determine local traffic and hazard situations, for example:

- approaching start of congested area;
- heavy rain, aquaplaning or sheet ice;
- poor visibility and fog.

At the time of writing these were still systems under development but they have so far produced so very promising results. (Ayala, Lin et al. 2010)

13.5.5.3 Mobile data

Some cars now allow direct communication so that messages can be sent to the car from a special website or from a smartphone application.

Figure 13.54 shows a smartphone application that can be used to send route information to a car and also to remotely lock or unlock the doors. SMS is the system used to communicate with the vehicle. Figure 13.55 shows an example of the data that can be accessed on what is effectively an Internet connected car.

13.5.5.4 Vehicle tracking

A vehicle tracking system uses an on-vehicle device and computer software at a base location. This allows the owner or authorised person to track the vehicle's location. This data is stored and analysed. Most vehicle tracking systems use GPS for locating the vehicle. Vehicle location and other information can be viewed on electronic maps via the Internet or by using suitable software.

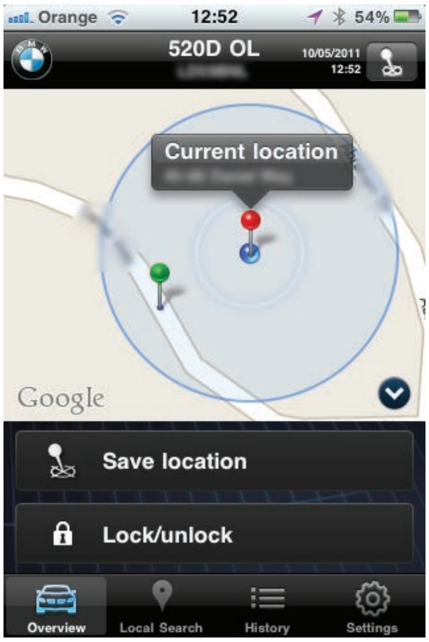


Figure 13.54 Screenshot of the 'MyBMW' iPhone application

Key fact

A smartphone application can be used to send route information to a car and also to remotely lock or unlock the doors (and more on some systems).

Figure 13.55 Online information relating to the vehicle's location accessed via the GPRS network

Key fact

Vehicle Tracking devices are described as 'active' or 'passive'.

Vehicle Tracking devices are described as 'active' or 'passive'. Passive devices store GPS location and other data and when the vehicle returns to base the information is downloaded to a computer for evaluation. Active devices collect the same data but transmit it in real-time via a cellular or satellite network.

Information that can be collected and analysed could include:

- GPS location information
- Fuel amount
- Engine temperature
- Altitude
- Door open/close
- Tyre pressures
- Ignition on/off
- Lights operation
- Battery status
- GSM area code/cell code decoded
- Cumulative idling
- Engine rpm
- Throttle position
- And more…

Vehicle tracking systems are often used by fleet operators for management functions such as:

- Fleet tracking
- Routing
- Dispatch
- On-board information and security.

Passenger transport and bus companies use the technology for a number of purposes:

- Monitoring of schedule.
- Triggering changes of destination sign displays.
- Triggering pre-recorded announcements for passengers.

CHAPTER 14

Heating ventilation and air conditioning

14.1 Conventional heating and ventilation

14.1.1 Introduction

The earliest electrical heating I have come across was a pair of gloves with heating elements woven into the material (c1920). These were then connected to the vehicle electrical system and worked like little electric fires. The thought of what happened in the case of a short circuit is a little worrying!

The development of interior vehicle heating has been an incremental process and will continue to be so – the introduction of air conditioning across mort new cars being the largest more recent step. The comfort we now take for granted had some very cold beginnings, but the technology in this area of the vehicle electrical system is still evolving. Systems range from basic hot/cold air blowers to complex automatic temperature and climate control systems.

Any heating ventilation and air conditioning (HVAC) system has a simple set of requirements, which are met to varying standards. These can be summarized as follows.

- Adjustable temperature in the vehicle cabin.
- Heat must be available as soon as possible.
- Distribute heat to various parts of the vehicle.

Figure 14.1 Ford Model T – The heating system was somewhat limited, but plenty of fresh air was provided…

Figure 14.2 HVAC controls (top right)

- Ventilate with fresh air with minimum noise.
- Facilitate the demisting of all windows.
- Ease of control operation.

The above list, whilst by no means definitive, gives an indication of what is required from a heating and ventilation system. As usual, the more complex the system the more the requirements are fulfilled.

Some solutions to the above requirements are discussed below, starting with simple ventilation and leading on to full automatic temperature control. Interestingly, temperature is not a hot and cold issue! Figure 14.3 shows a representation of how the perceived comfortable temperature in the vehicle compares with the outside temperature.

Key fact

The perceived comfortable temperature in the vehicle varies as the outside temperature changes.

14.1.2 Ventilation

To allow fresh air from outside the vehicle to be circulated inside the cabin, a pressure difference must be created. This is achieved by using a plenum chamber. A plenum chamber by definition holds a gas (in this case air), at a

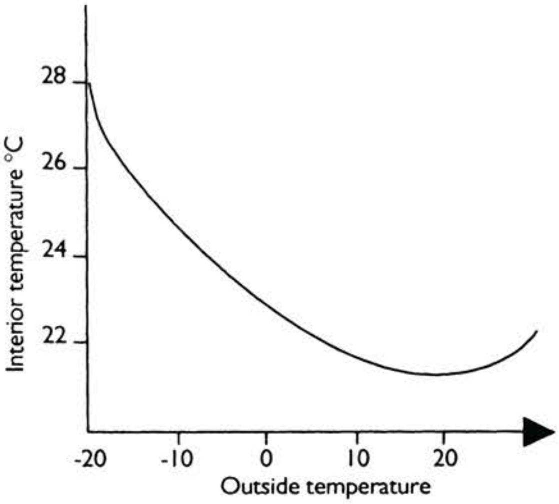


Figure 14.3 Representation of comfortable temperature

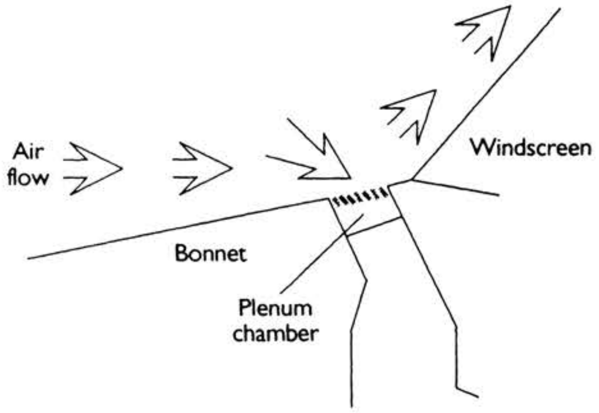


Figure 14.4 Plenum chamber effect

pressure higher than the ambient pressure. The plenum chamber on a vehicle is usually situated just below the windscreen, behind the bonnet hood. When the vehicle is moving, the air flow over the vehicle will cause a higher pressure in this area. Figure 14.4 shows an illustration of the plenum chamber effect. Suitable flaps and drains are utilized to prevent water entering the car through this opening.

Definition

A plenum chamber holds a gas at a pressure higher than the ambient pressure.

By means of distribution trunking, control flaps and suitable 'nozzles', the air can be directed as required (Figure 14.5). This system is enhanced with the addition of a variable speed blower motor. Figure 14.6 shows a typical (simplified) heating and ventilation system layout.

When extra air is forced into a vehicle cabin, the interior pressure would increase if no outlet is available. Most passenger cars have the outlet grills on each side of the vehicle above or near the rear quarter panels or doors.

14.1.3 Heating system – water-cooled engine

Heat from the engine is utilized to increase the temperature of the car interior. This is achieved by use of a heat exchanger, called the heater matrix. Due to the action of the thermostat in the engine cooling system the water temperature remains broadly constant. This allows for the air being passed over the heater

Figure 14.5 Air intake vents above the plenum chamber

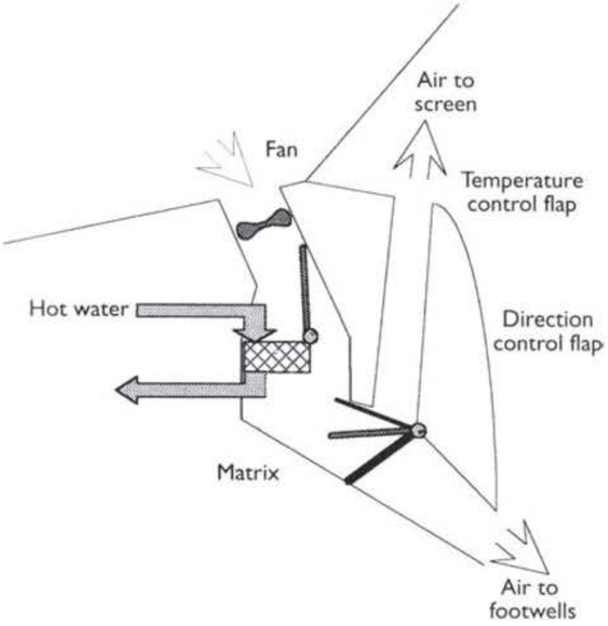


Figure 14.6 Heating and ventilation system basic layout

Key fact

Due to the action of the thermostat in the engine cooling system the water temperature remains broadly constant.

matrix to be heated by a set amount depending on the outside air temperature and the rate of air flow. A source of hot air is therefore available for heating the vehicle interior. However, some form of control is required over how much heat (if any), is required. The method used on most modern vehicles is the blending technique. This is simply a control flap, which determines how much of the air being passed into the vehicle is directed over the heater matrix. The main drawback of this system is the change in air flow with vehicle speed. Some systems use a valve to control the hot coolant flowing to the heater matrix.

By a suitable arrangement of flaps it is possible to direct air of the chosen temperature to selected areas of the vehicle interior. In general, basic systems allow the warm air to be adjusted between the inside of the windscreen and the driver and air passenger foot wells. Most vehicles also have small vents directing warm air at the drivers and front passenger's side windows. Fresh cool air outlets with directional nozzles are also fitted.

One final facility, which is available on many vehicles, is the choice between fresh or recirculated air. The main reason for this is to decrease the time it takes to demist or defrost the vehicle windows and to heat the car interior more quickly. The other reason is that, for example, in heavy congested traffic, the outside air may not be very clean.

14.1.4 Heater blower motors

The motors used to increase air flow are simple permanent magnet two-brush motors. The blower fan is often the centrifugal type and in many cases, the blades are positioned asymmetrically to reduce resonant noise. Figure 14.7 shows a typical motor and fan arrangement. Varying the voltage supplied controls motor speed. This is achieved by using dropping resistors. The speed in some cases is made 'infinitely' variable by the use of a variable resistor. In most cases the motor is controlled to three or four set speeds.

Figure 14.7 HVAC blower motor mounted in a housing

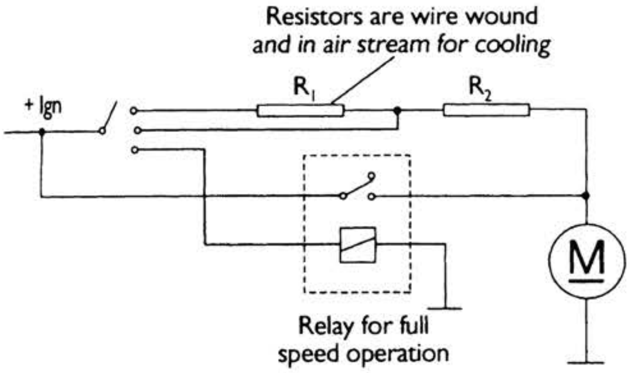


Figure 14.8 Circuit diagram of a three-speed control system

Figure 14.9 Blower motor and fan

Figure 14.8 shows a circuit diagram typical of a three-speed control system. The resistors are usually wire wound and are placed in the air stream to prevent overheating. These resistors will have low values in the region of 1 Ω or less. Most motors are now controlled by an electronic speed regulator.

14.1.5 Electronic heating control

Most vehicles that have electronic control of the heating system also include air conditioning, which is covered in the next section. However, a short description at this stage will help to lead into the more complex systems. Figure 14.10 shows a block diagram representing an electronically controlled vehicle-heating system.

This system requires control of the blower motor, blend flap, direction flaps and the fresh or recirculated air flap. The technique involves one or a number of temperature sensors suitably positioned in the vehicle interior to provide information for the ECU. The ECU responds to information received from these sensors and sets the controls to their optimum positions. The whole arrangement is, in fact, a simple closed loop feedback system with the air temperature closing the loop. The ECU has to compare the position of the temperature control switch with the information that is supplied by the sensors and either cool or heat the car interior as required.

Key fact

The ECU responds to information received from sensors and sets the controls to their optimum positions.

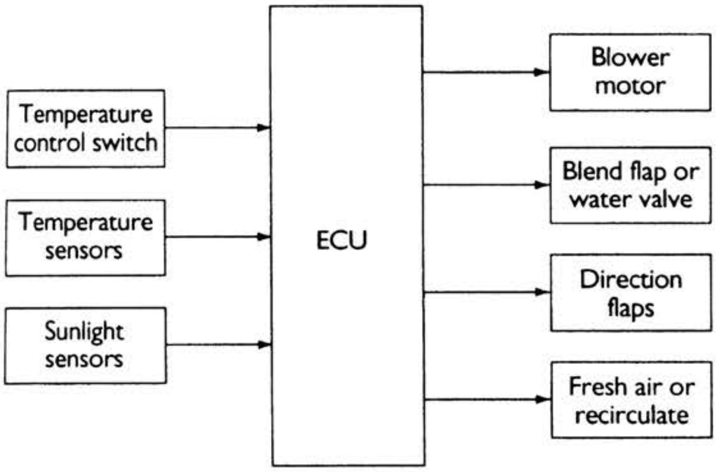


Figure 14.10 An electronically controlled vehicle-heating system

14.2 Air conditioning

14.2.1 Introduction

A vehicle fitted with air conditioning allows the temperature of the cabin to be controlled to the ideal or most comfortable value determined by the ambient conditions. The system as a whole still utilizes the standard heating and ventilation components, but with the important addition of an evaporator, which both cools and dehumidifies the air.

Air conditioning can be manually controlled or, as is now often the case, combined with some form of electronic control. The system as a whole can be thought of as a type of refrigerator or heat exchanger. Heat is removed from the car interior and dispersed to the outside air.

Key fact

An AC system can be thought of as a type of refrigerator or heat exchanger.

14.2.2 Principle of refrigeration

To understand the principle of refrigeration the following terms and definitions will be useful.

- Heat is a form of energy.
- Temperature means the degree of heat of an object.
- Heat will only flow from a higher to a lower temperature.
- Heat quantity is measured in ‘calories’ (more often kcal).
- 1 kcal heat quantity, changes the temperature of 1 kg of liquid water by 1 °C.
- Change of state is a term used to describe the changing of a solid to a liquid, a liquid to a gas, a gas to a liquid or a liquid to a solid.
- Evaporation is used to describe the change of state from a liquid to a gas.
- Condensation is used to describe the change of state from gas to liquid.
- Latent heat describes the energy required to evaporate a liquid without changing its temperature (breaking of molecular bonds) or the amount of heat given off when a gas condenses back into a liquid without changing temperature (making of molecular bonds).

Key fact

Latent heat in the change of state of a refrigerant is the key to air conditioning.

Latent heat in the change of state of a refrigerant is the key to air conditioning. A simple example of this is that if you put a liquid such as methylated spirits on your hand, it feels cold. This is because it evaporates and the change of state (liquid to gas) uses heat from your body. This is why the process is often thought of as ‘un-heating’ rather than cooling.

Figure 14.11 Blowing on a fluid makes it evaporate and feel cold

The refrigerant used in many air conditioning systems is known as R134a. This substance changes state from liquid to gas at 26.3 °C. R134a is hydrofluorocarbon (HFC) rather than chlorofluorocarbon (CFC) based, due to the problems with atmospheric ozone depletion associated with CFC-based refrigerants. Note that this type of refrigerant is not compatible with older systems.

A key to understanding refrigeration is to remember that a low-pressure refrigerant will have low temperature, and a high-pressure refrigerant will have a high temperature.

Figure 14.12 shows the basic principle of an air conditioning or refrigeration system. The three main components are the evaporator, condenser and pump (compressor). The evaporator is situated in the car; the condenser outside the car, usually in the air stream. The compressor is driven by the engine.

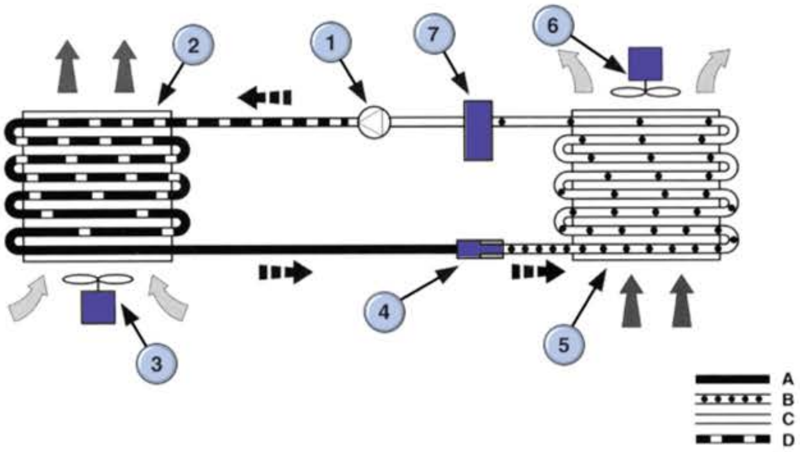


Figure 14.12 AC system layout: 1-Compressor, 2-Condenser, 3-Auxiliary fan (depending on model), 4-Fixed orifice tube, 5-Evaporator, 6-Heater/air conditioning blower, 7-Suction accumulator/drier, A-High-pressure warm liquid, B-Low-pressure cool liquid, C-Low pressure, gaseous and cool, D-High pressure, gaseous and hot (Source: Ford Motor Company)

As the pump operates it will cause the pressure on its intake side to fall, which will allow the refrigerant in the evaporator to evaporate and draw heat from the vehicle interior. The high pressure or output of the pump is connected to the condenser. The pressure causes the refrigerant to condense (in the condenser); thus giving off heat outside the vehicle as it changes state.

Several further components are needed for efficient operation; these are explained over the next few sections. Figure 14.13 shows some typical components of an air conditioning system.

14.2.3 Air conditioning overview

The operation of the system is a continuous cycle. The compressor pumps low pressure, heat-laden vapour from the evaporator, compresses it and pumps it

Figure 14.13 Heating ventilation and air conditioning (HVAC) components (Source: Delphi Media)

Key fact

The compressor pumps low pressure, heat laden vapour from the evaporator, compresses it and pumps it as a super-heated vapour under high pressure to the condenser.

as a super-heated vapour under high pressure to the condenser (Figure 14.14). The temperature of the refrigerant at this stage is much higher than the outside air temperature, hence it gives up its heat via the fins on the condenser as it changes state back to a liquid.

This high-pressure liquid is then passed to the receiver–drier where any vapour which has not yet turned back to a liquid is stored, and a desiccant bag removes any moisture (water) that is contaminating the refrigerant. The high-pressure liquid is now passed through the thermostatic expansion valve and is converted back to a low-pressure liquid as it passes through a restriction in the valve into the evaporator. This valve is the element of the system that controls the refrigerant flow and hence the amount of cooling provided. As the liquid changes state to a gas in the evaporator, it takes up heat from its surroundings, thus cooling or 'unheating' the air that is forced over the fins. The low-pressure vapour leaves the evaporator returning to the pump, thus completing the cycle. The cycle is represented in Figure 14.12.

If the temperature of the refrigerant increases, beyond a set limit, condenser cooling fans can be switched in to supplement the ram air effect.

A safety switch is fitted in the high-pressure side of most systems. It is often known as a high–low pressure switch, as it will switch off the compressor if the pressure is too high due to a component fault, or if the pressure is too low due to a leakage, thus protecting the compressor (Figure 14.15).

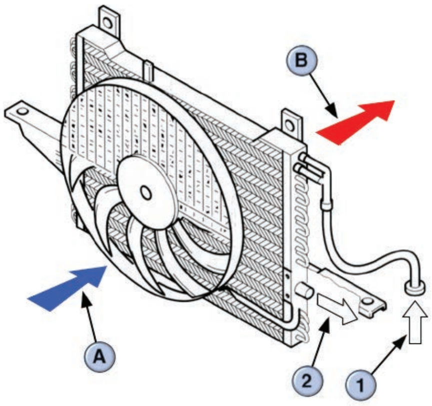


Figure 14.14 The condenser is very similar to the cooling system radiator: A-Cooled air, B-Heated air, 1-Gaseous refrigerant, 2-Liquid refrigerant

Figure 14.15 Pressure switch

14.2.4 Air conditioning system and components

The layout of a typical AC system is shown in Figure 14.16. This system uses a thermostatic expansion valve to control flow. Other types use an orifice system.

The compressor shown in Figure 14.18 is belt-driven from the engine crankshaft and it acts as a pump-circulating refrigerant through the system. The compressor shown is a piston and reed valve type. As the refrigerant is drawn into the cylinder due to the action of the piston, the outlet valve is closed due to the pressure. When the piston begins its compression stroke the inlet reed valve closes and the outlet opens. This compressor is controlled by an electromagnetic clutch, which may be either under manual control or electronic control depending on the type of system.

Figure 14.21 shows the condenser fitted in front of the vehicle radiator. It is very similar in construction to the radiator and fulfils a similar role. The heat is conducted through the aluminium pipes and fins to the surrounding air and then, by a process of convection, is dispersed by the air movement. The air movement is caused by the ram effect, which is supplemented by fans as required.

Figures 14.22 and 14.23 show a receiver–drier assembly. It is connected in the high-pressure line between the condenser and the thermostatic expansion valve. This component has four features.

- A reservoir to hold refrigerant until a greater flow is required.
- A filter to prevent contaminants circulating through the system.
- Vapour is retained in this unit until it finally converts back to a liquid.
- A drying agent removes any moisture from the system. The substance used in R134a systems is Zeolite. Some manufacturers recommend that this unit should be replaced if the system has been open to the atmosphere.

A sight glass is fitted to some receiver–driers to give an indication of refrigerant condition and system operation. The refrigerant generally appears clear if all is in order.

Key fact

The compressor is belt-driven from the engine and it circulates refrigerant through the system.

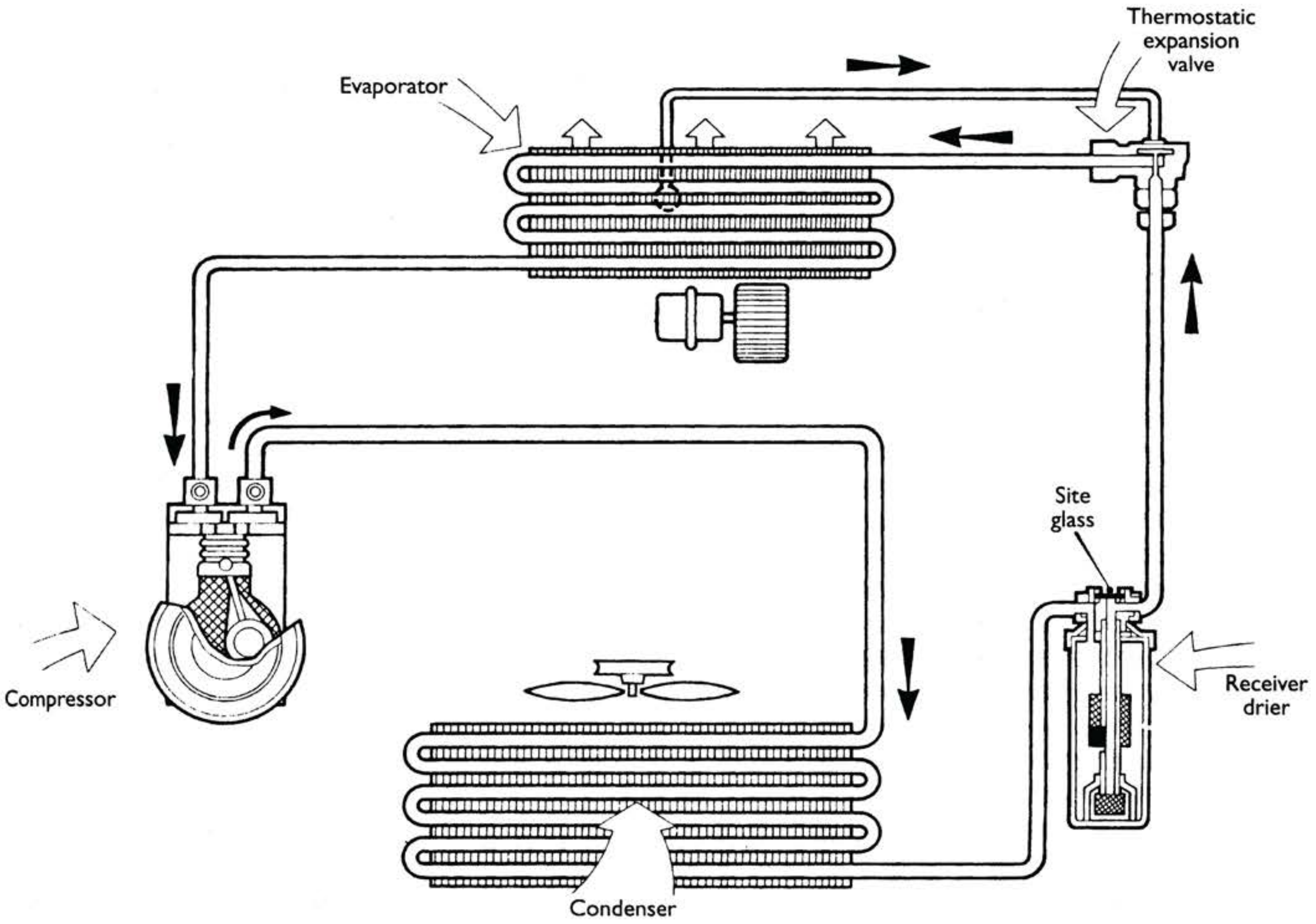


Figure 14.16 Air condition system layout

Figure 14.17 Thermostatic expansion valve

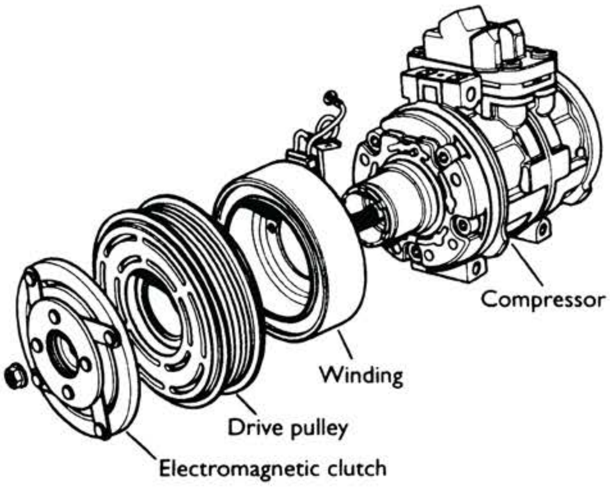


Figure 14.18 Air conditioning compressor

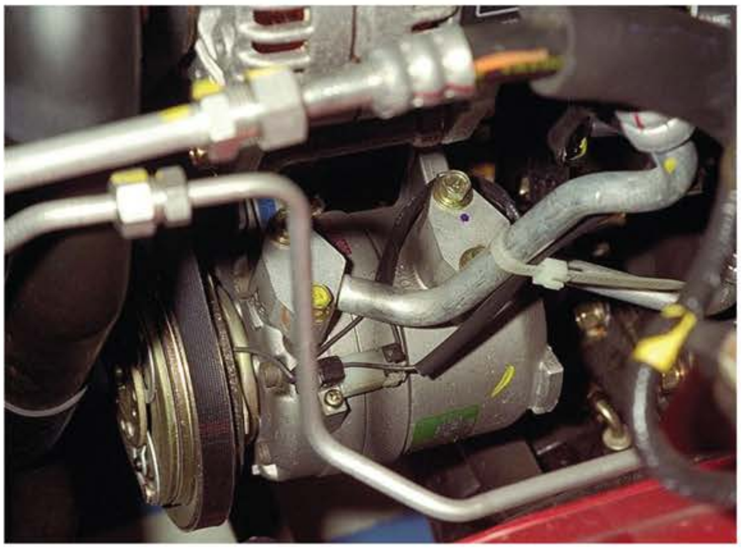


Figure 14.19 Compressor mounted on the engine

Figure 14.20 AC compressor clutch

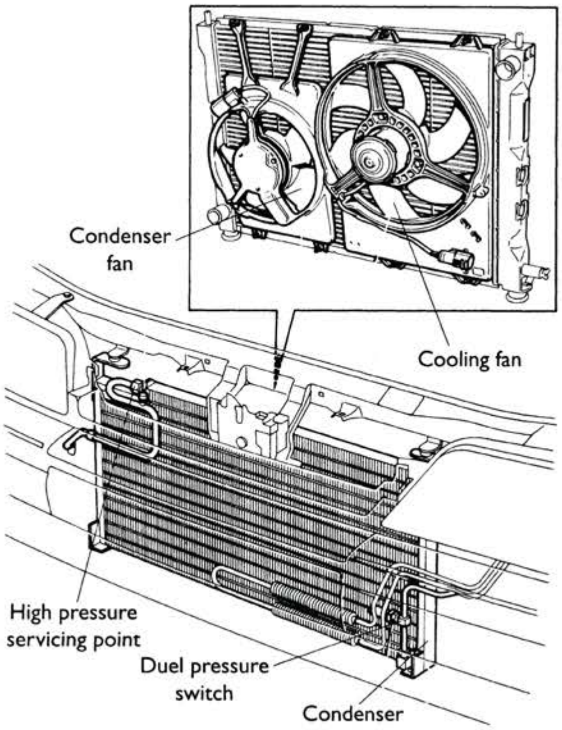


Figure 14.21 Air conditioning condenser

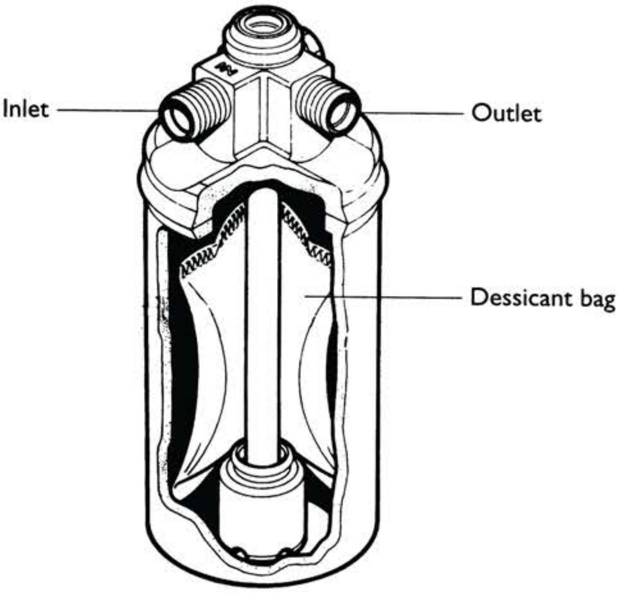


Figure 14.22 Cut-away receiver drier

A thermostatic expansion valve is shown as part of Figure 14.16 (and as Figure 14.17) together with the evaporator assembly. It has two functions to fulfil:

- Control the flow of refrigerant as demanded by the system.
- Reduce refrigerant pressure in the evaporator. The thermostatic expansion valve is a simple spring-controlled ball valve, which has a diaphragm attached to a spring. A temperature-sensitive gas such as carbon dioxide acts upon the diaphragm.

The gas is in a closed system including a capillary tube and a sensing bulb. This sensing bulb is secured on the evaporator. If the temperature of the evaporator

Figure 14.23 Receiver-drier example

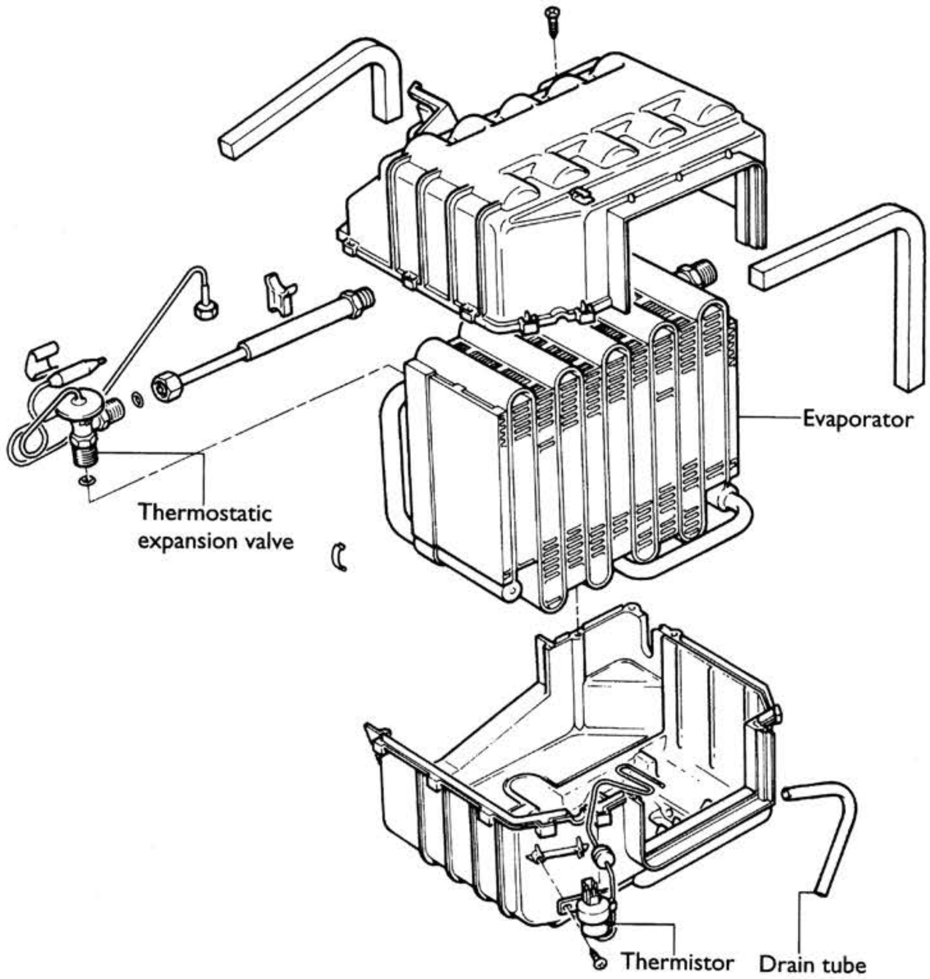


Figure 14.24 Evaporator and thermostatic expansion valve

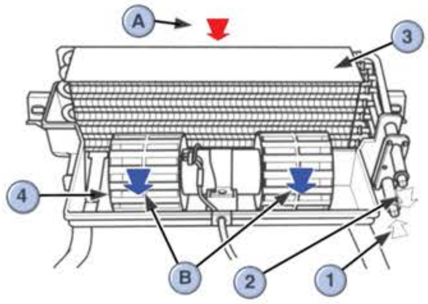


Figure 14.25 Evaporator housing and blower motor: A-Hot air, B-Cooled air, 1-Liquid refrigerant, 2-Gaseous refrigerant, 3-Evaporator, 4-Booster fan

Key fact

The evaporator assembly is similar in construction to the condenser, consisting of fins to maximize heat transfer.

rises, the gas in the bulb expands and acts on the diaphragm such as to open the ball valve and allow a greater flow of refrigerant. If the evaporator were to become too cold, then the gas in the bulb will contract and the ball valve will close. In this way, the flow of refrigerant is controlled and the temperature of the evaporator is held fairly constant under varying air flow conditions.

The evaporator assembly is similar in construction to the condenser, consisting of fins to maximize heat transfer. It is mounted in the car under the dash panel, forming part of the overall heating and ventilation system (Figures 14.24 and 14.25). The refrigerant changes state in the evaporator from a liquid to a vapour. As well as cooling the air passed over it, the evaporator also removes moisture from the air. This is because the moisture in the air will condense on the fins and can be drained away. The action is much like breathing on a cold pane of glass. A thermistor is fitted to the evaporator on some systems to monitor temperature. The compressor is cycled if the temperature falls below about 3 or 4 °C to prevent the chance of water freezing on the evaporator, which would restrict air flow.

The electrical circuit is shown in Figure 14.26. The following points are worthy of note.

- A connection exists between the air conditioning ECU and the engine management ECU. The reasons for this are so that the compressor can be switched off under very hard acceleration and to enable better control of engine idle.

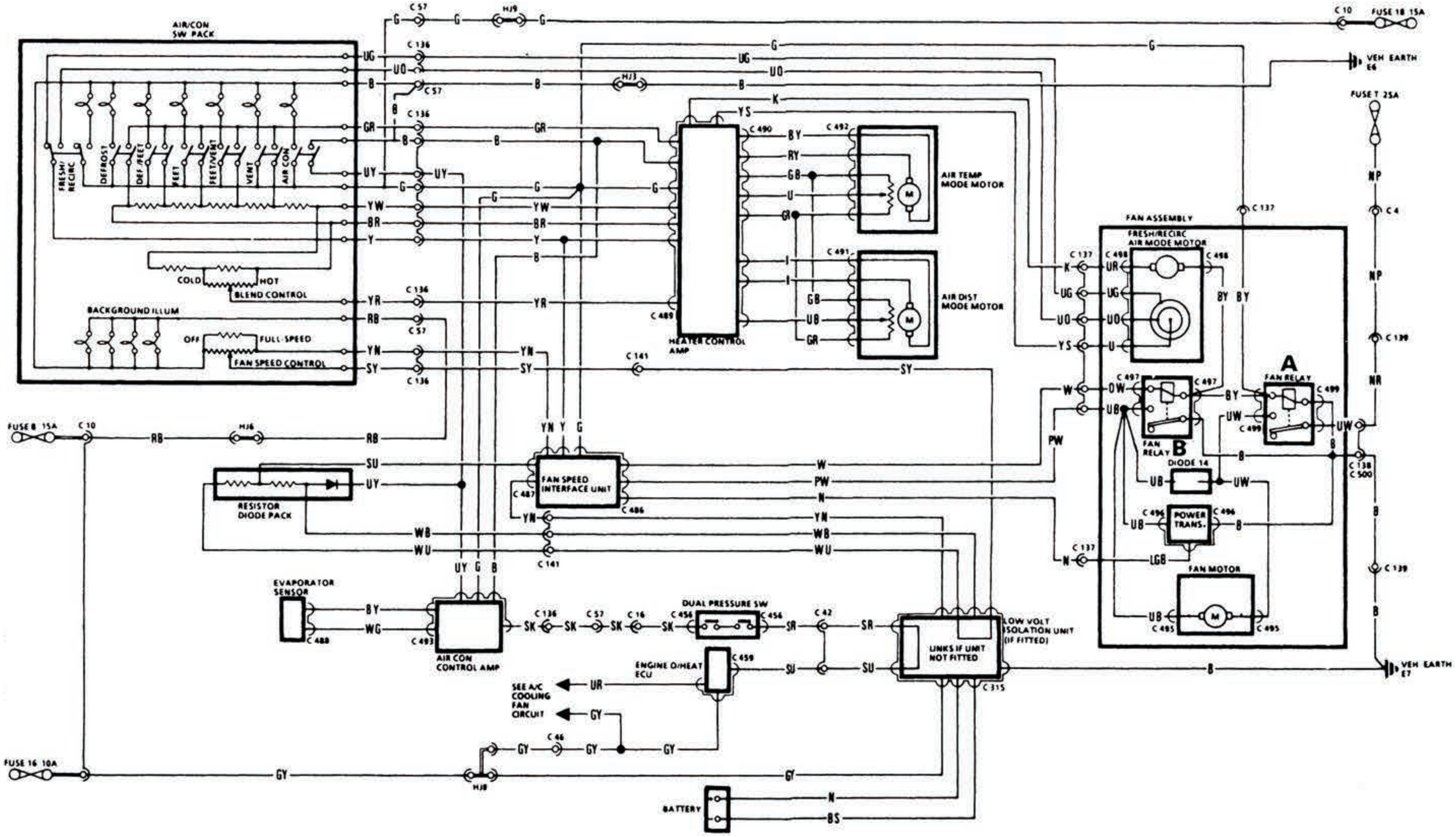


Figure 14.26 Air conditioning electrical unit

- Twin cooling fans are used to cool the condenser. These can be run at two speeds using relays to connect them in series for slow operation or in parallel for full speed.
- A number of safety features are included such as the high/low-pressure switches.

14.2.5 Automatic temperature control

Full temperature control systems provide a comfortable interior temperature in line with the passenger controlled input. The electronic control unit has full control of fan speed, air distribution, air temperature, fresh or recirculated air and the air conditioning pump. Some of these systems are also able to control automatic demist or defrost. A single button can also be used to set the system to full defrost or demist.

A number of sensors are used to provide input to the ECU:

- An ambient temperature sensor mounted outside the vehicle will allow compensation for extreme temperature variations. This device is usually a thermistor.
- A solar light sensor can be mounted on the fascia panel. This device is a photodiode and allows a measurement of direct sunlight from which the ECU can determine whether to increase the air to the face vents.
- The in-car temperature sensors are simple thermistors but, to allow for an accurate reading, a small motor and fan can be used to take a sample of interior air and direct it over the sensing elements.
- A coolant temperature sensor is used to monitor the temperature of the coolant supplied to the heater matrix. This sensor is used to prevent operation of the system until coolant temperature is high enough to heat the vehicle interior.
- Driver input control switches.

The ECU takes information from all of the above sources and will set the system in the most appropriate manner as determined by the software. Control of the flaps can be either by solenoid-controlled vacuum actuators or by small motors. The main blower motor is controlled by a heavy duty power transistor and is constantly variable. These systems are able to provide a comfortable interior temperature when exterior conditions range from 10 to 35 °C even in extreme sunlight.

Key fact

The climate control ECU takes information from all of the above sources and will set the system in the most appropriate manner as determined by the software.

14.2.6 Electrically driven air conditioning

Driving an air conditioning pump electrically can provide the following benefits:

- Sealed motor and pump assembly.
- Smaller and less complex compressor.
- Flexible positioning (no drive belt).
- Full cooling capacity at any engine speed.
- Greater control is possible.

The motor output power necessary to drive an electric automotive air conditioning (AC) system depends on the cooling capacity of the system, its efficiency and the boundary conditions (temperatures) it is operating against. All of these quantities are variable under normal vehicle operation. The use of a ‘brushless’ motor has been considered. The following ‘standard’ rating conditions are useful in assessing maximum power levels.

Stop-and-go driving in heavy traffic

Under these test conditions, high compressor discharge pressures will tend to overload the motor. To prevent this problem, fresh air must be restricted at idle to reduce evaporator load and, if possible, the condenser fan should operate at over-speed conditions. The motor must be operated at lower speeds during idle to prevent overload and, consequently, will not reach its maximum power requirement.

Hot soak followed by pull down

This test is established by placing the vehicle in a hot sunny environment until the cabin temperature rises to about 65 °C. The vehicle is then operated at about 50 km/h with maximum AC and fan speed control settings. An electric AC system operating at about half of its maximum speed offers pull down performance equivalent to a conventional AC system. If operated at maximum during the pull down test, a significant reduction in the time taken to reach acceptable cabin temperatures could be achieved.

Cruising with full fresh air intake

This operating condition requires the AC system to maintain comfortable cabin temperatures while processing significant quantities of outside air.

This establishes a maximum capacity level, which in turn sets the size of the motor and its drive electronics. For conventional AC systems, a 3.75 kW motor is a reasonable estimate for this condition. About 70% of the total load is used to condition the fresh outside air. Reducing or eliminating fresh air load at highway speeds has a direct influence on the size of the electric drive system.

A series of computer simulations was conducted to explore ways of reducing the motor power requirement. Using a two-stage cycle with 25% fresh air results in a 1.5 kW load on the motor. A conventional cycle using a high-efficiency compressor coupled with a 20% fresh air limitation also results in a 1.5 kW load. A 1.5 kW motor is a realistic option for automotive air conditioning.

The combined motor and electronics cost significantly affects the feasibility of electric automotive AC systems. Cost increases significantly as the required motor power increases. Development of more efficient AC systems is on-going. Assuming these developments are successful, a 1.5 kW electrically driven AC system will be possible and will be able to provide performance equal to or better than today's systems.

In this application, brushless DC motor systems are expected to achieve efficiencies of 85–90% when designed specifically for sealed automotive AC applications. This translates into a maximum electrical demand from the vehicle power supply system of 1.7 kW when the AC electrical drive operates under maximum cooling conditions.

14.3 Other heating systems

14.3.1 Seat heating

The concept of seat heating is very simple. A heating element is placed in the seat, together with an on–off or temperature control switch. However, the design of these heaters is more complex than first appears.

The heater must meet the following criteria.

Figure 14.27 Seat containing heating element (Source: Scandmec)

- The heater must only supply the heat loss experienced by the person's body.
- Heat to be supplied only at the major contact points.
- Leather and fabric seats require different systems due to their different thermal properties.
- Heating elements must fit the design of the seat.
- The elements must pass the same rigorous tests as the seat, such as squirm, jounce and bump tests.

In order for the passengers (including the driver) to be comfortable, rigorous tests have been carried out to find the optimum heat settings and the best position for the heating elements. Many tests are carried out on new designs, using a manikin with sensors attached, to measure the temperature and heat flow. Figure 14.27 shows a seat containing heating elements.

The cable used for most heating elements is known as a Sine Cable and consists of multi-strand alloyed copper. This cable may be coated with tin or insulated as the application demands. The heating element is laminated and bonded between layers of polyurethane foam.

The traditional method of control is a simple thermostat switch. Recent developments, however, tend to favour electronic control combined with a thermistor. Electronic systems that include push button switches, potentiometers, a timer function, short and open circuit detection are now common. This is in addition to accurate control of the chosen temperature setting. These seat heaters will heat up to provide an initial sensation in 1 minute and to full regulated temperature in 3 minutes.

14.3.2 Screen heating

Heating of the rear screen involves a very simple circuit as shown in Figure 14.28. The heating elements consist of a thin metallic strip bonded to the glass. When a current is passed through the elements, heat is generated and the window will defrost or demist. This circuit can draw high current, 10–15 A being typical. Because of this, the circuit often contains a timer relay to prevent the heater being left on too long. The timer will switch off after 10–15 minutes. The elements are usually positioned to defrost the main area of the screen and the rest position of the rear wiper blade if fitted.

Front windscreen heating is available on some vehicles. This of course presents more problems than the rear screen, as vision must not be obscured. The technology, drawn from the aircraft industry, involves very thin wires cast into the glass (Figure 14.29). As with the heated rear window, this device can consume a large current and is usually operated by a timer relay.

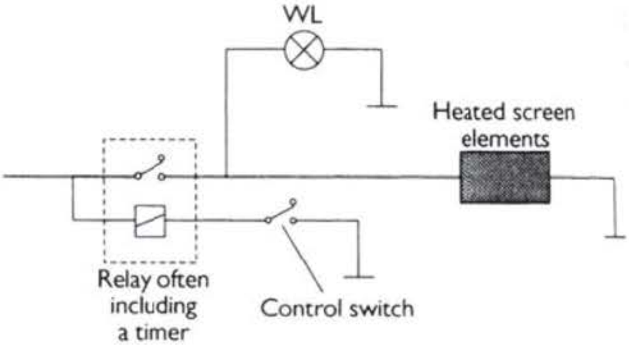


Figure 14.28 Screen heating circuit

Figure 14.29 Screen elements in a heated windscreen

14.3.3 Heating development

Two interesting developments show how heating systems can actually be quite cool! These are:

- Pollution sensor.
- Photo-catalyst.

The pollution sensor provides improved cabin air quality for enhanced comfort of vehicle passengers. The sensor detects the principal noxious atmospheric pollutants (carbon monoxide and nitrogen dioxides) present in the environment. It fits into the air inlets of the heating and AC system. By electronic control, the sensor automatically activates the air recirculation mode, sealing off the cabin from harmful air pollution. The flaps can be closed in as little as 1.8 seconds. This results in a 20% decrease in pollution concentration and a 40% decrease in passengers noticing the odours.

Using a photo-catalyst means that air in the cabin can be improved because pollutant gases are eliminated. These gases can be destroyed by the UV action of a photo-catalyst. Volatile organic components, nitrogen and sulphur oxides are the main culprits. Bacteria can also be eliminated because they are killed on the filter. The photo-catalyst, which is made of titanium oxide (TiO_2), is self-regenerating to ensure long life. The system is 70% efficient on toluene after 6 minutes in recirculation mode. It can also be made to start automatically in association with the pollution sensor.

14.3.4 Air conditioning system faults

Table 14.1 lists some common symptoms of an air conditioning system malfunction together with suggestions for the possible fault. The faults are generic but will serve as a good reminder. It is assumed an appropriate pressure gauge set has been connected.

The process of checking an air conditioning system is broadly as follows.

1. Hand and eye checks (loose wires, loose switches and other obvious faults) – all connections clean and tight.
2. Check system pressures.
3. Check discharge temperature.

Safety first

Do not work on the refrigerant side of air conditioning systems unless you have been trained and have access to suitable equipment.

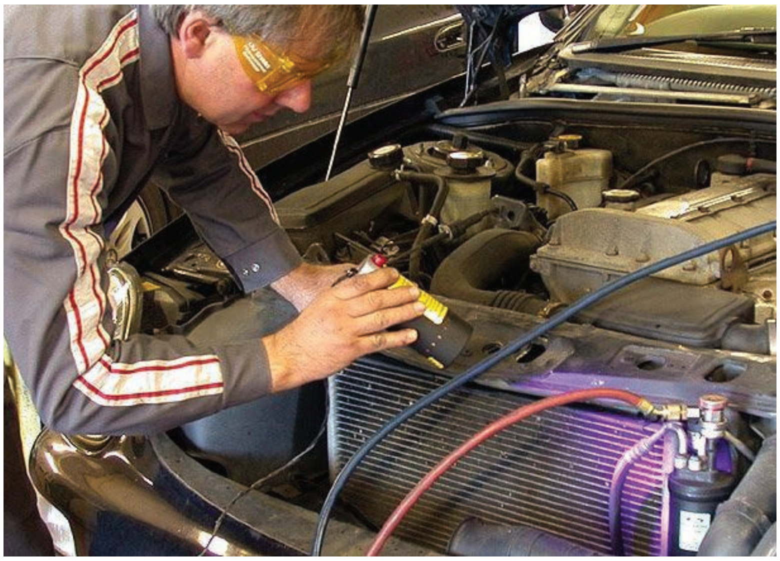


Figure 14.30 After adding a dye, leaks can be detected using a UV light

Table 14.1 Symptoms and faults of an air conditioning system

Symptom	Possible fault
After stopping the compressor, pressure falls quickly to about 195 kPa and then falls gradually	Air in the system or, if no bubbles are seen in the sight glass as the condenser is cooled with water, excessive refrigerant may be the fault.
Discharge pressure low	Fault with the compressor or, if bubbles are seen, low refrigerant.
Discharge temperature is lower than normal	Frozen evaporator.
Suction pressure too high	High pressure valve fault, excessive refrigerant or expansion valve open too long.
Suction and discharge pressure too high	Excessive refrigerant in the system or condenser not working due to fan fault or clogged fins.
Suction and discharge pressure too low	Clogged or kinked pipes.
Refrigerant loss	Oily marks (from the lubricant in the refrigerant) near joints or seals indicate leaks.

4. Inspect receiver–drier sight glass.
5. Refer to Table 14.1.

14.4 Advanced temperature control technology

14.4.1 Heat transfer

This section outlines three key terms associated with heat transfer and a typical heat loss calculation.

Convection: This is heat energy transfer that involves the movement of a fluid (gas or liquid). Fluid in contact with the source of heat expands and tends to rise within the bulk of the fluid. Cooler fluid sinks to take its place, setting up convection currents.

Conduction: Flow of heat energy through a material without the movement of any part of the material itself is conduction. Heat energy is present in all materials in the form of the kinetic energy of their vibrating molecules, and may be conducted from one molecule to the next in the form of this mechanical vibration. In the case of metals, which are particularly good conductors of heat, the free electrons within the material carry heat around very quickly.

Radiation: In physics, radiation is the emission of radiant energy as particles or waves; for example, heat, light, alpha particles and beta particles.

When designing a heating or air conditioning system, calculations can be used to determine the heating or cooling effect required. The following is the main heat current equation and can be used, for example, to help determine the heat loss through the windows.

$$\frac{\Delta Q}{\Delta t} = -kA\frac{\Delta T}{\Delta x} = -\frac{\Delta T}{{}^{\Delta x}\!/_{kA}}$$

where: ΔQ *= heat energy,* ΔT *= temperature,* Δx *= thickness/distance of material,* Δt *= time,* k *= thermal conductivity of the material (*$W\,m^{-1}\,K^{-1}$*),* A *= cross sectional area.*

$\frac{\Delta Q}{\Delta t}$ can be thought of as 'heat current'.

14.4.2 Types of heat and temperature

Heat: Heat is a form of energy that is transferred by a difference in temperature. It always moves from a higher to a lower temperature.

Latent heat: Heat that is absorbed without causing a rise in temperature is described as latent heat. For example, 'latent heat of vaporization' refers to the amount of heat required to convert a liquid to vapour at a particular temperature.

Sensible heat: Heat energy that causes a rise or fall in the temperature of a gas, liquid or solid when added or removed from that material is described as being sensible! Sensible heat changes the temperature by changing the speed at which the molecules move.

Temperature: Temperature is the degree or intensity of heat of a body and the condition that determines whether or not it will transfer heat to, or receive heat from, another body.

14.4.3 Armature reaction

Most heater motors, like many other motors, are unidirectional due to the positioning of the brushes. When a motor is running it also acts as a generator producing a back emf. The brushes of a motor (or generator) must be placed around the commutator in such a way that, as the armature rotates and the brushes effectively short one commutator segment to the next, no emf must be present in that associated armature winding.

If an emf is present, then current will flow when the short is made. This creates sparks at the brushes and is known as armature reaction. To overcome this problem, the brushes are moved from the geometric neutral axis to the magnetic neutral axis of the motor fields. This is because, as armature current flows, the magnetism created around the armature windings interacts with

Key fact

When a motor is running it also acts as a generator producing a back emf.

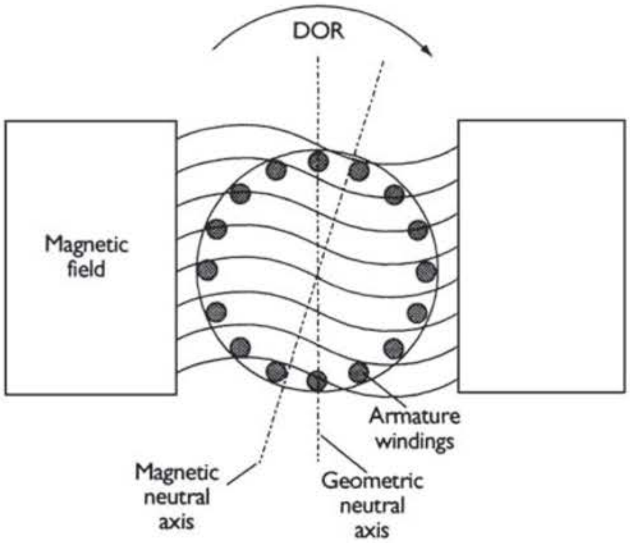


Figure 14.31 Field wrap which causes armature reaction

the main magnetic field causing it to warp. Figure 14.31 shows this field warp diagrammatically. This phenomenon was also used in some very early generators as a way of controlling output.

14.4.4 Refrigerant developments

Key fact

The ideal refrigerant has good thermodynamic properties, is unreactive chemically, and safe.

The ideal refrigerant has good thermodynamic properties, is unreactive chemically, and safe. The desired thermodynamic properties are a boiling point somewhat below the target temperature, a high heat of vaporization, a moderate density in liquid form, a relatively high density in gaseous form, and a high critical temperature. Since boiling point and gas density are affected by pressure, refrigerants may be made more suitable for a particular application by choice of operating pressure. These properties are ideally met by the chlorofluorocarbons. Corrosion properties are a matter of materials compatibility with the mechanical components: compressor, piping, evaporator, and condenser. Safety considerations include toxicity and flammability.

Until concerns about depletion of the ozone layer arose in the 1980s, the most widely used refrigerants were the halomethanes R-12 and R-22, with R-12 being more common in automotive air conditioning. R-134a and certain blends have tended to replace the earlier compounds.

Following the ban on chlorofluorocarbons (CFCs) and hydro-chlorofluorocarbons (HCFCs), substances used as substitute refrigerants such as fluorocarbons (FCs) and hydrofluorocarbons (HFCs) have also come under criticism. They are currently subject to prohibition discussions on account of their harmful effect on the climate.

Natural refrigerants such as ammonia, carbon dioxide and non-halogenated hydrocarbons preserve the ozone layer and have no (ammonia) or only a low (carbon dioxide, hydrocarbons) global warming potential (GWP).

Definition

GWP = 100 year warming potential of one kilogram of a gas relative to one kilogram of CO_2.

Emissions from automotive air-conditioning are a growing concern because of their impact on climate change. From 2011 on, the European Union will phase out refrigerants with a GWP of more than 150 in automotive air conditioning. This will ban potent greenhouse gases such as the refrigerant HFC (R-134a), which has a GWP of 1410. One of the most promising alternatives is the

Figure 14.32 Most cars have a label stating the type of refrigerant used

natural refrigerant CO_2 (R-744). Carbon dioxide is non-flammable, non-ozone depleting, has a global warming potential of 1, but is toxic and potentially lethal in concentrations above 5% by volume.

GM has announced that it will start using hydrofluoro-olefin (HFO-1234yf) in all of its brands by 2013. This new refrigerant has a GWP rating of 4.

An interesting possibility for use as a refrigerant is air. With suitable compression and expansion technology, air can be a practical refrigerant, free of the possibility of environmental contamination or damage, and almost completely harmless to plants and animals. It is however not as efficient as other substances.

Key fact

An interesting possibility for use as a refrigerant is air.

Figure 14.33 Refrigerant

This page intentionally left blank

CHAPTER 15

Chassis electrical

15.1 Anti-lock brakes

15.1.1 Introduction

The reason for the development of anti-lock brakes (ABS) is very simple. Under braking conditions, if one or more of the vehicle wheels locks (begins to skid), there are a number of consequences.

- Braking distance increases.
- Steering control is lost.
- Abnormal tyre wear.

The obvious result is that an accident is far more likely to occur. The maximum deceleration of a vehicle is achieved when maximum energy conversion is taking place in the brake system. This is the conversion of kinetic energy to heat energy at the discs and brake drums. The potential for this conversion process between a skidding tyre, even on a dry road, is far less. A good driver can pump the brakes on and off to prevent locking but electronic control can achieve even better results.

Definition

Maximum deceleration of a vehicle is achieved when maximum energy conversion is taking place in the brake system.

ABS is becoming now common even on lower price vehicles, which should be a significant contribution to safety. It is important to remember, however, that for normal use, the system is not intended to allow faster driving and shorter braking distances. It should be viewed as operating in an emergency only. Figure 15.1 shows how ABS can help to maintain steering control even under very heavy braking conditions.

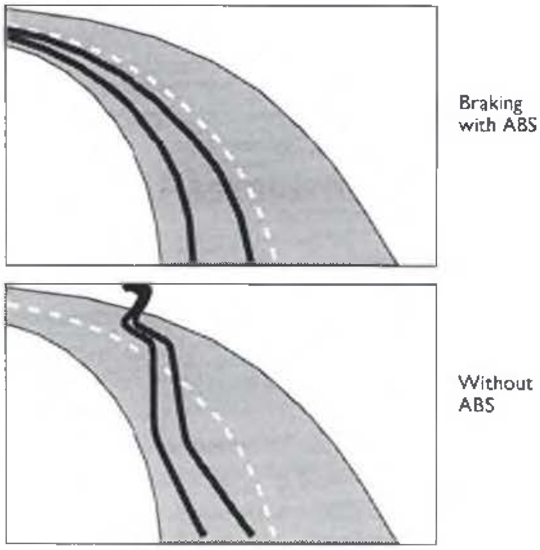


Figure 15.1 ABS can help maintain steering control

15.1.2 Requirements of ABS

A good way of considering the operation of a complicated system is to ask: ‘what must the system be able to do?’ In other words, what are the requirements? These can be considered for ABS under the following headings.

Fail-safe system

In the event of the ABS system failing the conventional brakes must still operate to their full potential. In addition, a warning must be given to the driver. This is normally in the form of a simple warning light.

Manoeuvrability must be maintained

Good steering and road holding must continue when the ABS system is operating. This is arguably the key issue, as being able to swerve around a hazard whilst still braking hard is often the best course of action.

Immediate response must be available

Even over a short distance the system must react such as to make use of the best grip on the road. The response must be appropriate whether the driver applies the brakes gently or slams them on hard.

Operational influences

Normal driving and manoeuvring should produce no reaction on the brake pedal. The stability and steering must be retained under all road conditions. The system must also adapt to braking hysteresis when the brakes are applied, released and then re-applied.

Even if the wheels on one side are on dry tarmac and the other side on ice, the yaw (rotation about the vertical axis of the vehicle) of the vehicle must be kept to a minimum and only increase slowly in order to allow the driver to compensate.

Controlled wheels

In its basic form, at least one wheel on each side of the vehicle should be controlled on a separate circuit. It is now general for all four wheels to be separately controlled on passenger vehicles.

Speed range of operation

The system must operate under all speed conditions down to walking pace. At this very slow speed even when the wheels lock the vehicle will come to rest very quickly. If the wheels did not lock then, in theory, the vehicle would never stop!

Other operating conditions

The system must be able to recognize aquaplaning and react accordingly. It must also still operate on an uneven road surface. The one area still not perfected is braking from slow speed on snow. The ABS will actually increase stopping distance in snow but steering will be maintained. This is considered to be a suitable trade-off.

A number of different types of anti-lock brake systems are in use, but all try to achieve the requirements as set out above.

15.1.3 General system description

As with other systems, ABS can be considered as a central control unit with a series of inputs and outputs. An ABS system is represented by the closed loop system block diagram shown in Figure 15.2. The most important of the inputs

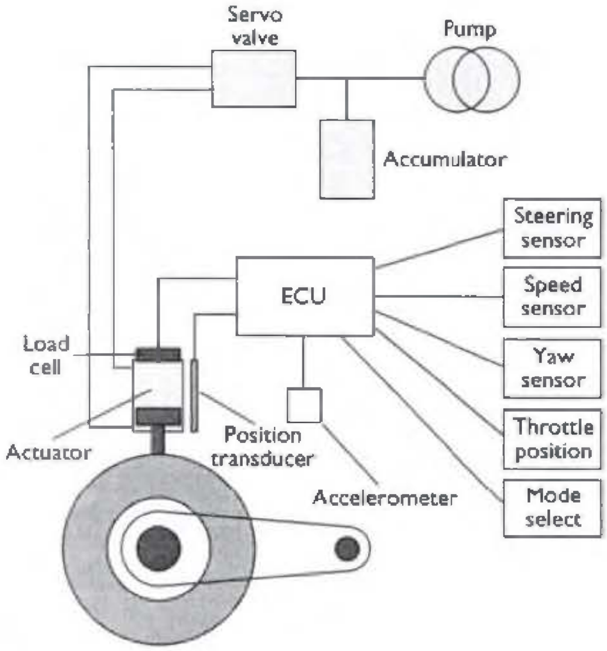


Figure 15.2 Anti-lock brake system

are the wheel speed sensors, and the main output is some form of brake system pressure control.

The task of the control unit is to compare signals from each wheel sensor to measure the acceleration or deceleration of an individual wheel. From these data and pre-programmed look-up tables, brake pressure to one or more of the wheels can be regulated. Brake pressure can be reduced, held constant or allowed to increase. The maximum pressure is determined by the driver's pressure on the brake pedal.

A number of variables are sensed, used or controlled by this system:

Pedal pressure

Determined by the driver.

Brake pressure

Under normal braking this is proportional to pedal pressure but under control of the ABS it can be reduced, held or allowed to increase.

Controlled variable

This is the actual result of changes in brake pressure, in other words the wheel speed, which then allows acceleration, deceleration and slip to be determined.

Road/vehicle conditions

Disturbances such as vehicle load, state of the road, tyre condition and brake system condition.

From the wheel speed sensors the ECU calculates the following:

Vehicle reference speed

Determined from the combination of two diagonal wheel sensor signals. After the start of braking the ECU uses this value as its reference.

Wheel acceleration or deceleration

This is a live measurement that is constantly changing.

Key fact

Driven and non-driven wheels on the vehicle must be treated in different ways as they behave differently when braking.

Brake slip

Although this cannot be measured directly, a value can be calculated from the vehicle reference speed. This figure is then used to determine when/if ABS should take control of the brake pressure.

Vehicle deceleration

During brake pressure control, the ECU uses the vehicle reference speed as the starting point and decreases it in a linear manner. The rate of decrease is determined by the evaluation of all signals received from the wheel sensors.

Driven and non-driven wheels on the vehicle must be treated in different ways as they behave differently when braking.

A logical combination of wheel deceleration/acceleration and slip is used as the controlled variable. The actual strategy used for ABS control varies with the operating conditions.

15.1.4 Components

There are a few variations between manufacturers involving a number of different components. For the majority of systems, however, there are three main components.

- Wheel speed sensors.
- Electronic control unit.
- Hydraulic modulator.

Wheel speed sensors

Most of these devices are simple inductance sensors and work in conjunction with a toothed wheel. They consist of a permanent magnet and a soft iron rod around which is wound a coil of wire. As the toothed wheel rotates, the changes in inductance of the magnetic circuit generate a signal; the frequency and voltage of which are proportional to wheel speed. The frequency is the signal used by the electronic control unit. The coil resistance is of the order of 1 k. Coaxial cable is used to prevent interference affecting the signal. Some systems now use ‘Hall effect’ sensors, as described in Chapter 2.

Electronic control unit

The function of the ECU (Figure 15.3 shows part of an ECU) is to take in information from the wheel sensors and calculate the best course of action for the hydraulic modulator. The heart of a modern ECU consists of two microprocessors such as the Motorola 68HC11, which run the same program independently of each other. This ensures greater security against any fault, which could adversely affect braking performance because the operation of each processor should be identical. If a fault is detected, the ABS disconnects itself and operates a warning light. Both processors have non-volatile memory into which fault codes can be written for later service and diagnostic access. The ECU also has suitable input signal processing stages and output or driver stages for actuator control.

The ECU performs a self-test after the ignition is switched on. A failure will result in disconnection of the system. The following list forms the self-test procedure.

- Current supply.
- Exterior and interior interfaces.
- Transmission of data.
- Communication between the two microprocessors.

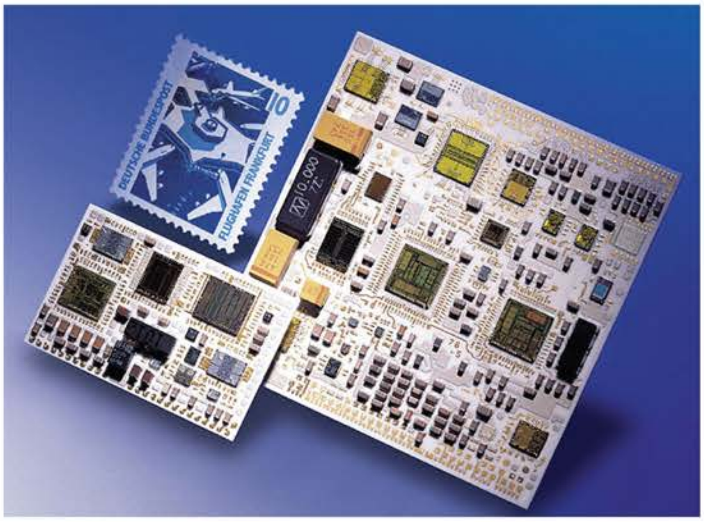


Figure 15.3 A microprocessor as used in an ECU

- Operation of valves and relays.
- Operation of fault memory control.
- Reading and writing functions of the internal memory.

All this takes about 300 ms.

Hydraulic modulator

The key to the operation of a hydraulic modulator, as shown in Figure 15.4, is to consider three operating positions:

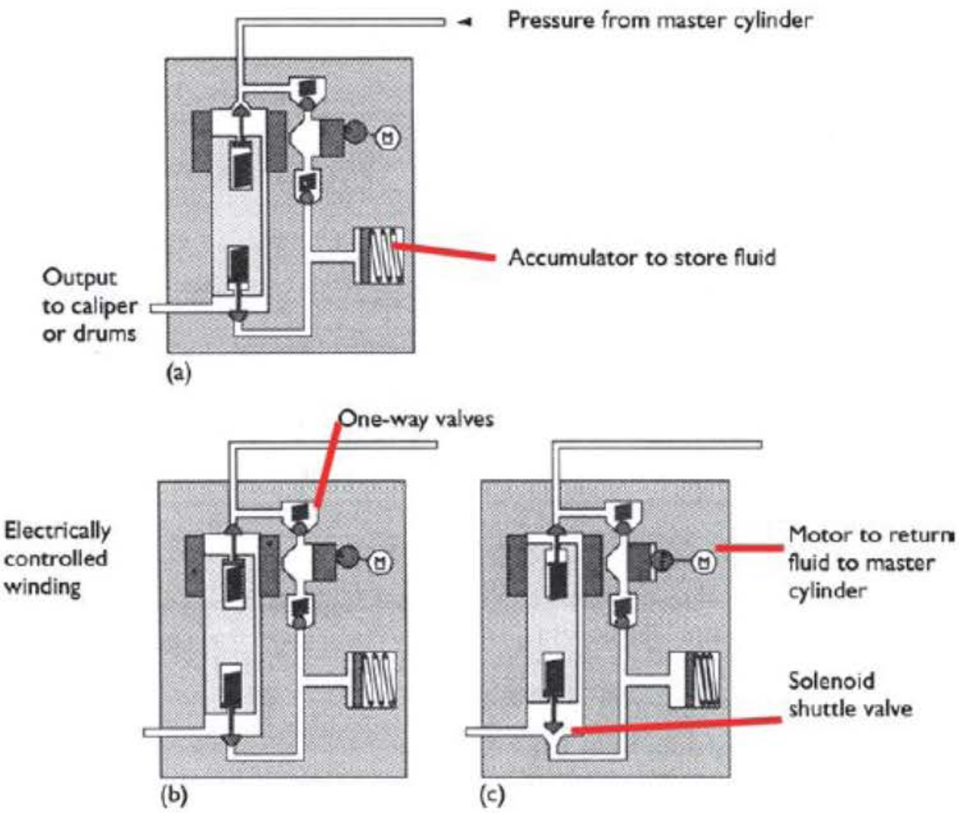


Figure 15.4 ABS Hydraulic modulator; (a) Normal pressure build-up; (b) holding phase; (c) Reducing

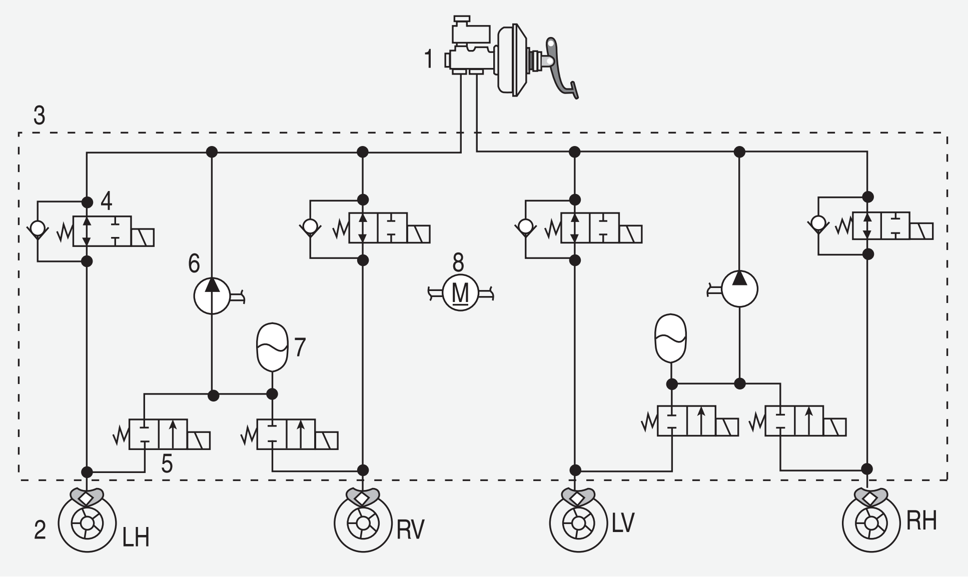


Figure 15.5 Hydraulic system for Bosch ABS-8 system
1 Master cylinder, 2 Wheel-brake cylinder, 3 Hydraulic modulator, Inlet valves, 5 Outlet valves, 6 Return pump, 7 Accumulator, 8 Pump motor. V front, H rear, R right, L left.

- Pressure build-up – brake line open to the master cylinder.
- Pressure reducing – brake line open to the accumulator.
- Pressure holding – brake line closed.

The valves are controlled by electrical solenoids, which have a low inductance so they react very quickly. The motor only runs when ABS is activated.

Figures 15.5 and 15.6 show the Bosch ABS-8 modulator and system. Note how the diagonal split brake circuits are separated.

Bosch began the manufacture of the current ABS 8 generation in 2001, which has been modified and improved several times since. In its currently most

Figure 15.6 Bosch ABS version 8 modulator and ECU

compact version, it weighs just 1.4 kg, with just 14 highly integrated components in the control unit and a 256 kB memory. Generation 8 is modular in design, which allows the various degrees of complexity of the brake control system, ABS, TCS, and ESP, to be manufactured in very similar ways.

15.1.5 Anti-lock brake system control

The control of ABS can be summarized under a number of headings as given below.

Brake pressure control commencement

The start of ABS engagement is known as 'first control cycle smoothing'. This smoothing stage is necessary in order not to react to minor disturbances such as an uneven road surface, which can cause changes in the wheel sensor signals. The threshold of engagement is critical as, if it were too soon, it would be distracting to the driver and cause unnecessary component wear; too late and steering/stability could be lost on the first control cycle.

Even road surface regulation

Under these ideal circumstances adhesion is almost constant. ABS works at its best under these conditions, regulation frequency is relatively low with small changes in brake pressure.

Vehicle yaw (twist about the vertical axis, swerving moment)

When braking on a road surface with different adhesion under the left and right wheels, the vehicle will yaw or start to spin. The driver can control this with the steering if time is available. This can be achieved if when the front wheel with poor adhesion becomes unstable, the pressure to the other front wheel is reduced. This acts to reduce the vehicle yaw, which is particularly important, when the vehicle is cornering.

Axle vibration

Wheel speed instability occurs frequently and at random on rough roads. Due to this instability, brake pressure tends to be reduced more than it is increased, during ABS operation. This could lead to loss of braking under certain conditions. Adaptation to the conditions is therefore necessary to overcome this problem. An increase in brake pressure is made easier during hard re-acceleration of the wheel after an unstable instant. With modern soft suspension systems the axle may be subject to vibration. This can cause superimposed signals on the wheel speed sensors. The indicated accelerations can be the same as for actual unstable braking conditions. A slight delay in the reaction of the ABS due to the delay in signal smoothing – the time taken to move control valves and a time lag in the brake lines – helps to reduce the effect of axle vibration. The regular frequency of the vibrations can be recognized by the ECU. A constant brake pressure is introduced when axle vibrations are recognized.

Key fact

The three main operating phases of ABS are: pressure reducing, holding and build up.

15.1.6 Control strategy

The strategy of the anti-lock brake system can be summarized as follows.

- Rapid brake pressure reduction during wheel speed instability so the wheel will re-accelerate fast without too much pressure reduction, which will avoid under braking.

- Rapid rise in brake pressure during and after a re-acceleration to a value just less than the instability pressure.
- Discrete increase in brake pressure in the event of increased adhesion.
- Sensitivity suited to the prevalent conditions.
- Anti-lock braking must not be initiated during axle vibration.

The application of these five main requirements leads to the need for compromise between them. Optimum programming and prototype testing can reduce the level of compromise but some disadvantages have to be accepted. The best example of this is braking on uneven ground in deep snow, as deceleration is less effective unless the wheels are locked up. In this example, priority is given to stability rather than stopping distance, as directional control is favoured in these circumstances.

15.1.7 Honda anti-lock brakes

A Honda anti-lock brake system is based on the plunger principle. Figure 15.7 shows the schematic diagram. When anti-lock is not operating, the chamber labelled W is connected to the reservoir via the outlet valve. The chamber is held at atmospheric pressure because the inlet valve blocks the line from the pressure accumulator. During braking, pressure is created in the master cylinder and fluid flows from chamber Z into chamber X, moving the piston and increasing pressure in chamber Y.

If a wheel threatens to lock, the outlet valve closes, pressure in chamber W rises and prevents further movement of the piston thus holding the pressure. If the risk of lock-up continues the inlet valve opens and allows fluid to flow from the accumulator into chamber W. This pressure moves the piston back, thus reducing the pressure to the wheel cylinder. When the risk of lockup has gone, the inlet valve closes and the hold phase is restored.

The Honda system is a relatively simple ABS and has just two control channels. The front wheel which has the higher coefficient of friction determines the brake pressure for both front wheels. The result is that one front wheel may lock during extreme braking. The rear wheel with the lower coefficient of friction determines the rear brake pressure.

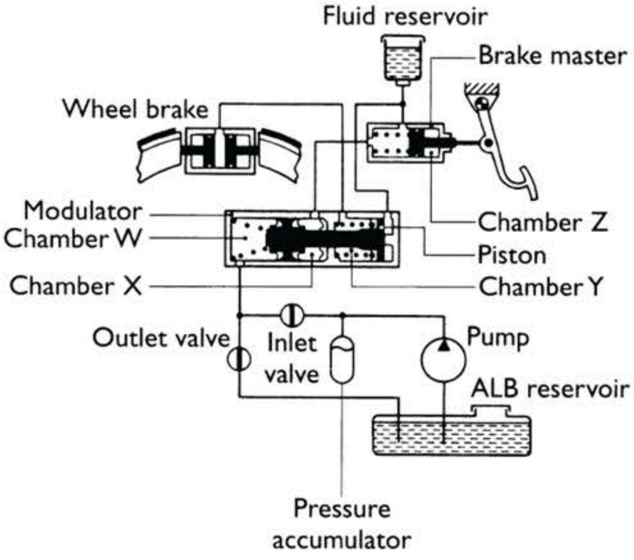


Figure 15.7 Honda ABS

15.2 Traction and stability control

15.2.1 Introduction

The steerability of a vehicle is not only lost when the wheels lock up on braking; the same effect arises if the wheels spin when driving off under severe acceleration. Electronic traction control has been developed as a supplement to ABS. This control system prevents the wheels from spinning when moving off or when accelerating sharply while on the move. In this way, an individual wheel, which is spinning is braked in a controlled manner. If both or all of the wheels are spinning, the drive torque is reduced by means of an engine control function. Traction control has become known as ASR, TCS or TCR.

Key fact

The steerability of a vehicle is not only lost when the wheels lock up on braking; the same effect arises if the wheels spin when driving off under severe acceleration.

Traction control is not normally available as an independent system, but in combination with ABS. This is because many of the components required are the same as for the ABS. Traction control only requires a change in logic control in the ECU and a few extra control elements such as control of the throttle. Figure 15.8 shows a block diagram of a traction control system. Note the links with ABS and the engine control system.

Traction control will intervene to achieve the following:

- Maintain stability.
- Reduction of yawing moment reactions.
- Provide optimum propulsion at all speeds.
- Reduce driver workload.

The following list of advantages can be claimed for a good traction control system.

- Improved tractive force.
- Better safety and stability on poor surfaces.
- Less driver stress.
- Longer tyre life.
- No wheel spin on turning and cornering.

An automatic control system can intervene in many cases more quickly and precisely than the driver of the vehicle. This allows stability to be maintained at time when the driver might not have been able to cope with the situation. Figure 15.9 shows an ABS and traction control modulator, complete with an ECU.

15.2.2 Control functions

Control of tractive force can be by a number of methods. Figure 15.10 shows a comparison of three techniques used to prevent wheel spin, throttle, ignition and brake control.

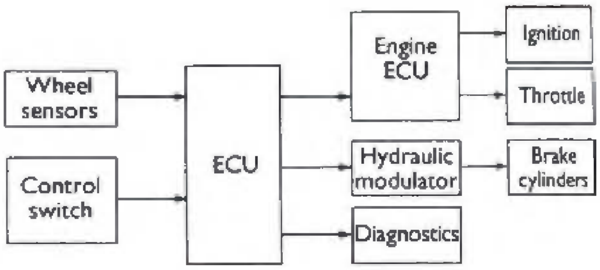


Figure 15.8 Traction control system

Figure 15.9 ABS and traction control ECU on the modulator

Throttle control

This can be via an actuator, which can move the throttle cable, or if the vehicle employs a drive-by-wire accelerator, then control will be in conjunction with the engine management ECU. This throttle control will be independent of the driver's throttle pedal position. This method alone is relatively slow to control engine torque.

Ignition control

If ignition is retarded, the engine torque can be reduced by up to 50% in a very short space of time.

The timing is adjusted by a set ramp from the ignition map value.

Braking effect

If the spinning wheel is restricted by brake pressure, the reduction in torque at the affected wheel is very fast. Maximum brake pressure is not used, to ensure passenger comfort is maintained.

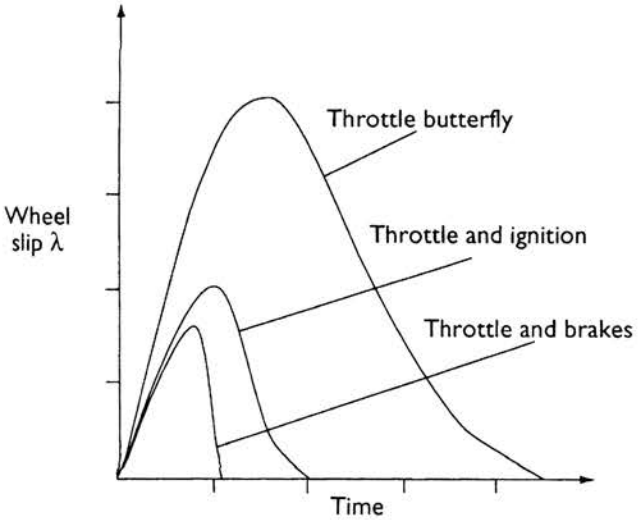


Figure 15.10 Comparison of three techniques used to prevent wheel spin: throttle, ignition and brake control

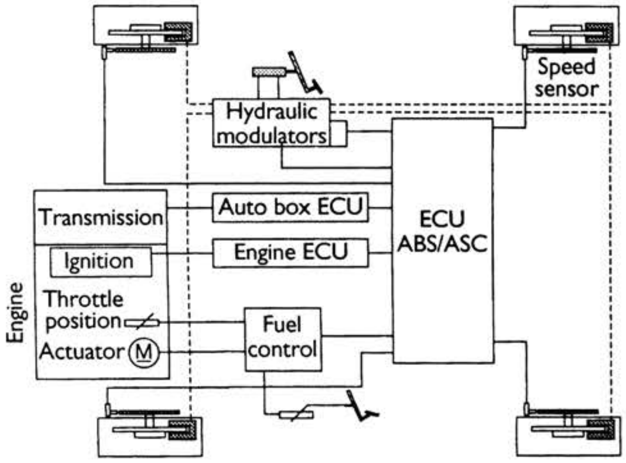


Figure 15.11 Layout of a traction control system which includes links with other vehicle control systems

15.2.3 System operation

The layout of a traction control system, which includes links with other vehicle control systems, is shown in Figure 15.11. The description that follows is for a vehicle with an electronic (drive-by-wire) accelerator.

A simple sensor determines the position of the accelerator and, taking into account other variables such as engine temperature and speed for example, the throttle is set at the optimum position by a servo motor. When accelerating, the increase in engine torque leads to an increase in driving torque at the wheels. In order for optimum acceleration, the maximum possible driving torque must be transferred to the road. If driving torque exceeds that which can be transferred, wheel slip will occur, at least at one wheel. The result of this is that the vehicle becomes unstable.

When wheel spin is detected the throttle position and ignition timing are adjusted but the best results are gained when the brakes are applied to the spinning wheel. This not only prevents the wheel from spinning but acts such as to provide a limited slip differential action. This is particularly good when on a road with varying braking force coefficients. When the brakes are applied, a valve in the hydraulic modulator assembly moves over to allow traction control operation. This allows pressure from the pump to be applied to the brakes on the offending wheel. The valves – in the same way as with ABS – can provide pressure build-up, pressure hold and pressure reduction. This all takes place without the driver touching the brake pedal.

The summary of this is that the braking force must be applied to the slipping wheel, such as to equalize the combined braking coefficient for each driving wheel.

Key fact

When wheel spin is detected the throttle position and ignition timing are adjusted but the best results are gained when the brakes are applied to the spinning wheel.

15.2.4 Electronic Stability Program (ESP)

Vehicle stability systems are designed to take over from the driver in extreme conditions to reduce the chances or severity of an accident. Manufacturers describe this in a number of ways, for example, electronic stability control (ESC), dynamic stability control (DSC), and a few other variations on a theme such as vehicle dynamic control (VDC).

Key fact

Vehicle stability systems are designed to take over from the driver in extreme conditions to reduce the chances or severity of an accident.

Figure 15.12 The first generation of the Bosch Electronic Stability Program (Source: Bosch Media)

Bosch ESP Bosch was the first supplier worldwide in 1995 to start the series production of the Electronic Stability Program (ESP) – so we will stick with this name (Figure 15.12). The system of the Generation 5.0, as it was called at that time, consisted of eleven components as shown here. More modern systems have a reduced number of components and some are integrated together (Figure 15.13).

The latest ESP system comprises the following components:

- Hydraulic unit with attached electronic control unit.
- Wheel speed sensors.
- Steering angle sensor.
- Yaw rate sensor with integrated acceleration sensor.

This ESP active safety system monitors vehicle stability 25 times per second. The terms active and passive are simple but important terms in the world of automotive safety.

Definition

Active and Passive Safety: Active safety refers to technology assisting in the prevention of a crash and passive safety refers to components that protect occupants during a crash.

However, the terms tend to blur, because stability control (for example) actively intervenes to prevent a crash. So, here is my latest version: Active safety refers to relevant technologies up to the point of impact. Passive safety refers to systems that operate from the point of impact.

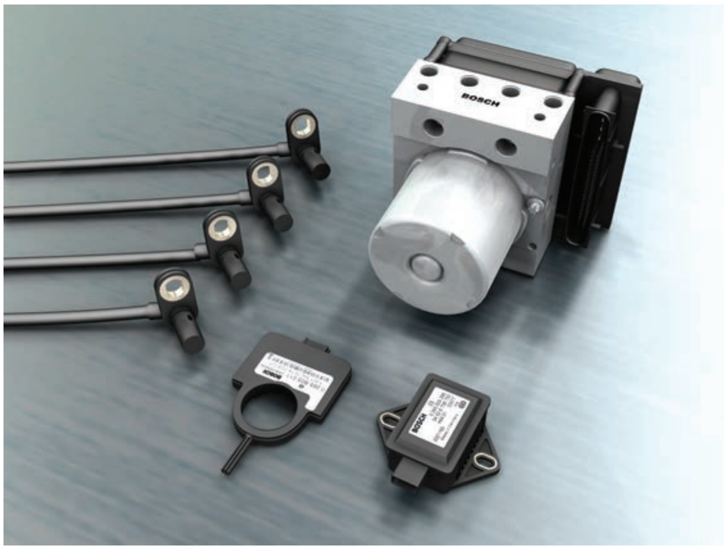


Figure 15.13 Bosch Electronic Stability Program (generation 8) (Source: Bosch Media)

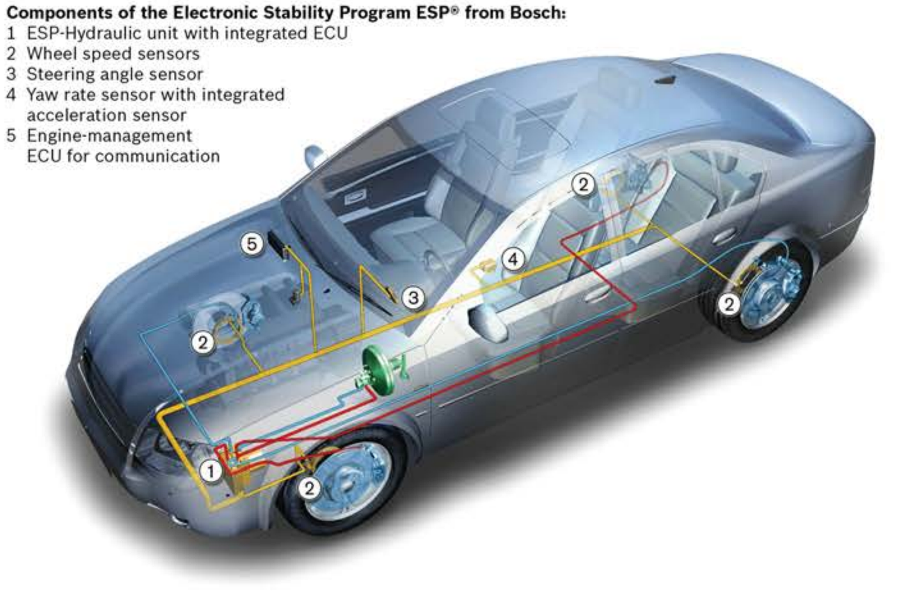

Components of the Electronic Stability Program ESP® from Bosch:
1 ESP-Hydraulic unit with integrated ECU
2 Wheel speed sensors
3 Steering angle sensor
4 Yaw rate sensor with integrated acceleration sensor
5 Engine-management ECU for communication

Figure 15.14 Bosch ESP system and components (Source: Bosch Media)

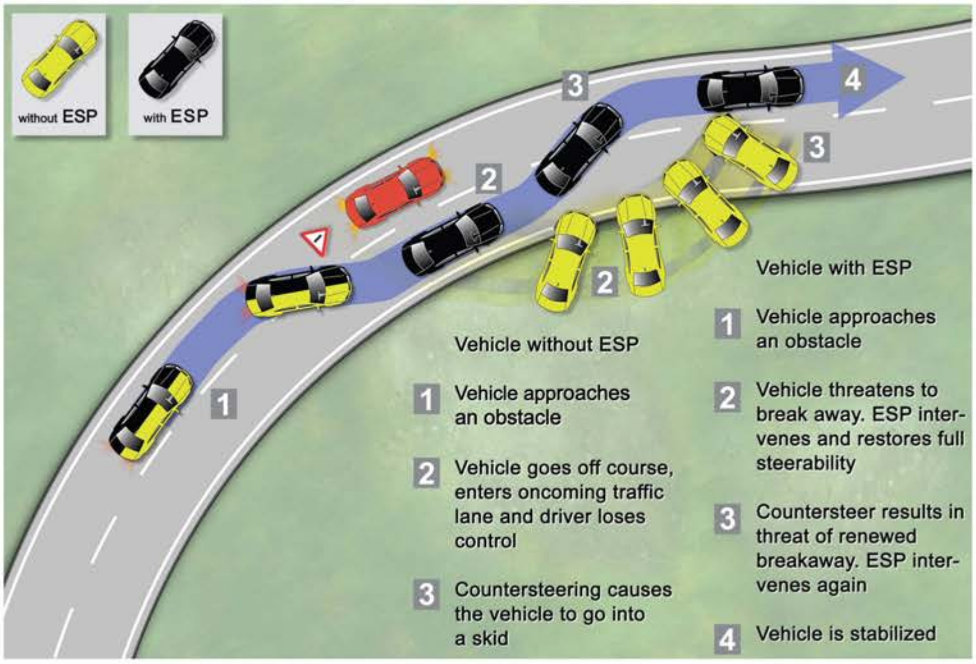
The hydraulic component of an ESP system consists of the hydraulic unit and electronics. The unit also forms a part of the anti-lock brake system (ABS).

ESP improves the safety of a vehicle's handling by detecting and preventing skids. Using a range of sensors, it detects whether the car is following the steering movements. When ESP detects loss of steering control, it automatically applies individual brakes to help control the vehicle and keep it in the direction intended by the driver (Figures 15.15 and 15.16).

Key fact

When ESP detects loss of steering control, it automatically applies individual brakes to help control the vehicle and keep it in the direction intended by the driver.

If the vehicle threatens to leave its track, ESP can apply individual braking forces at each wheel in order to correct vehicle movement. If for instance the vehicle



Vehicle without ESP

1 Vehicle approaches an obstacle

2 Vehicle goes off course, enters oncoming traffic lane and driver loses control

3 Countersteering causes the vehicle to go into a skid

Vehicle with ESP

1 Vehicle approaches an obstacle

2 Vehicle threatens to break away. ESP intervenes and restores full steerability

3 Countersteer results in threat of renewed breakaway. ESP intervenes again

4 Vehicle is stabilized

Figure 15.15 Critical manoeuvre – with and without ESP (Source: Bosch Media)

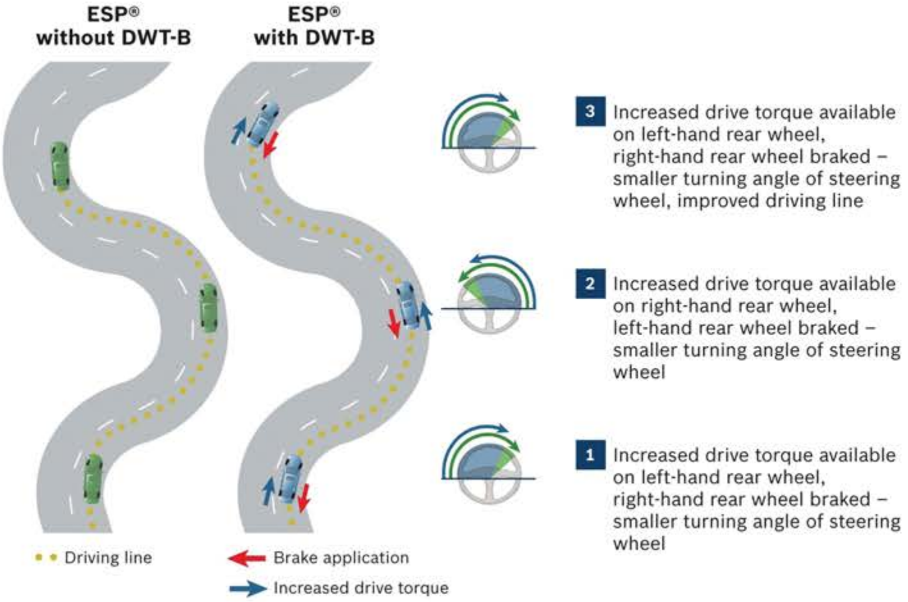


Figure 15.16 Vehicle dynamic management using dynamic wheel control (Source: Bosch Media)

Key fact

ESP is able to influence engine torque and therefore the slip at the driven wheels.

threatens to oversteer in a curve, ESP applies the brakes at the front wheel on the outside of the curve and generates a reaction torque which stabilizes the vehicle. Furthermore, if necessary, ESP is able to influence engine torque and therefore the slip at the driven wheels. ESP reacts very quickly and thus neutralizes critical situations before they have a chance to materialize.

The complex ESP control system would be impossible without high-performance sensors and electronics. The system calculates the desired trajectory based on steering angle, while wheel speed sensors are measuring the rotational velocity of all wheels. A central element within this system is a yaw-rate sensor designed to consistently monitor the vehicle's tendency to rotate around its vertical axis (Figure 15.17).

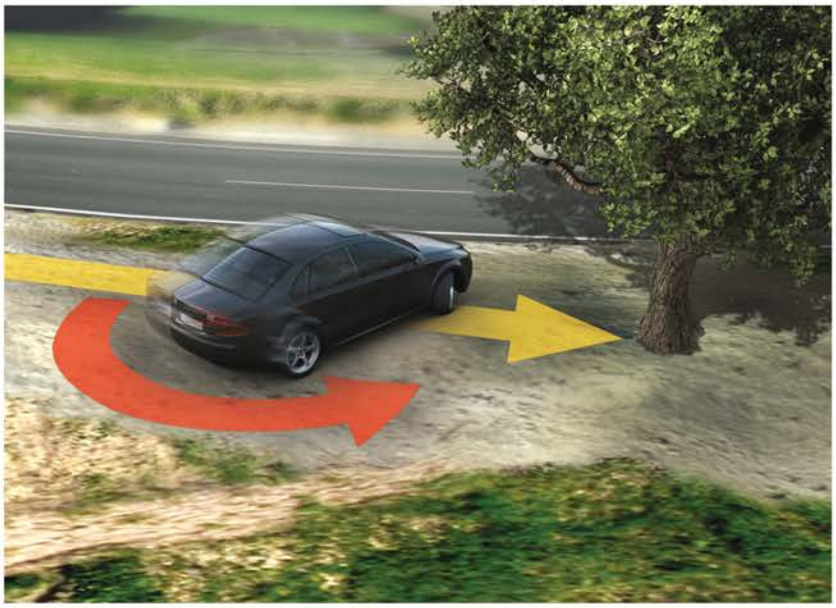


Figure 15.17 Yaw is the turning of a vehicle about a vertical axis (Source: Bosch Media)

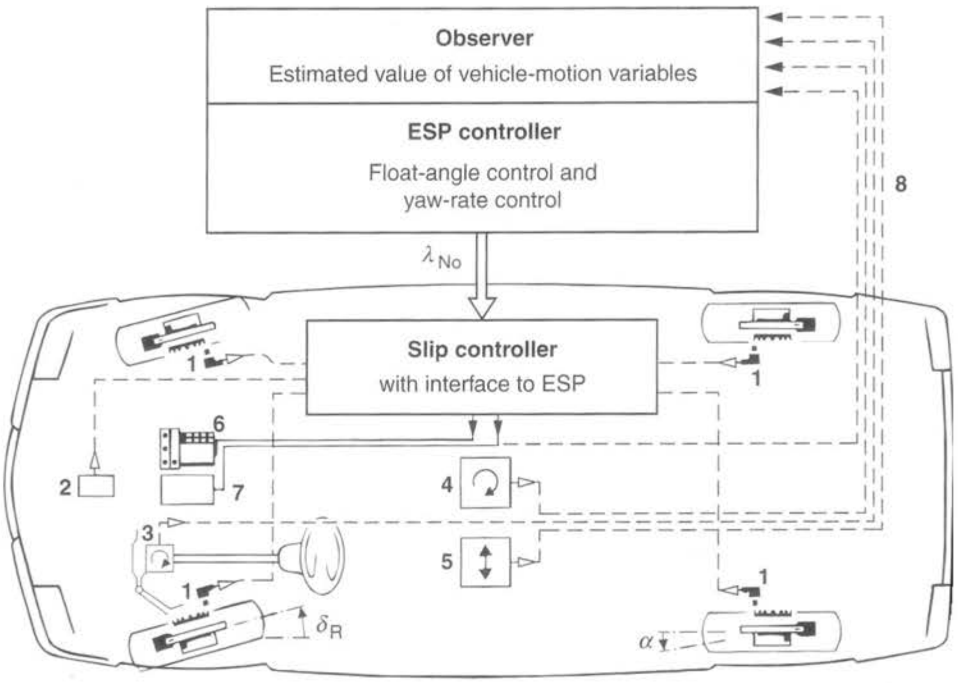


Figure 15.18 Overall ESP control system: 1, Wheel-speed sensors; 2, Brake-pressure sensor; 3, Steering-wheel angle sensor; 4, Yaw-rate sensor; 5, Lateral-acceleration sensor; 6, Brake-pressure modulation; 7, Engine management; 8, Sensor signals for ESP α Tyre slip angle; δ_R Steering angle; λ_{No} Nominal tyre slip (Source: Bosch UK)

ESP saves lives by improving vehicle stability in critical conditions. However, do remember it can't change the laws of physics – so beyond a certain point – a crash will happen no matter how complex the system!

Figure 15.18 shows the overall control system and the operating functions of the ESP system.

The computing power of ESP means that it is now possible to network active and passive safety systems so as to address other causes of crashes. For example, sensors may detect when a vehicle is following too closely and slow down the vehicle, straighten seat backs, and tighten seat belts, to avoid or prepare for a crash.

Safety first

ESP can't change the laws of physics!

15.3 Active suspension

15.3.1 Overview

Electronic control of suspension or active suspension, like many other innovations was born in the world of Formula 1. It is now slowly becoming more popular on production vehicles. Conventional suspension systems are always a compromise between soft springs for comfort and harder springing for better cornering ability. Active systems have the ability to switch between the two extremes.

A traditional or a conventional suspension system, consisting of springs and dampers, is passive. In other words, once it has been installed in the car, its characteristics do not change.

The main advantage of a conventional suspension system is its predictability. Over time the driver will become familiar with a car's suspension and understand

Key fact

Conventional suspension systems are a compromise between soft springs for comfort and hard springs for performance.

Figure 15.19 Jaguar suspension system (Source: Jaguar Media)

its capabilities and limitations. The disadvantage is that the system has no way of compensating for situations beyond its original design. The main benefits of active suspension are therefore as follows:

- Improvements in ride comfort, handling and safety.
- Predictable control of the vehicle under different conditions.
- No change in handling between laden and unladen.

Definition

Active suspension: a system that can change its properties depending on conditions (also known as computerized ride control).

The benefits are considerable and as component prices fall, the system will become available on more vehicles. It is expected that more off-road vehicles may be fitted with active suspension in the near future.

An active suspension system (also known as computerized ride control) has the ability to adjust itself continuously. It monitors and adjusts its characteristics to suit the current road conditions. As with all electronic control systems, sensors supply information to an ECU which in turn outputs to actuators. By changing its characteristics in response to changing road conditions, active suspension offers improved handling, comfort, responsiveness and safety (Figures 15.20 and 15.21).

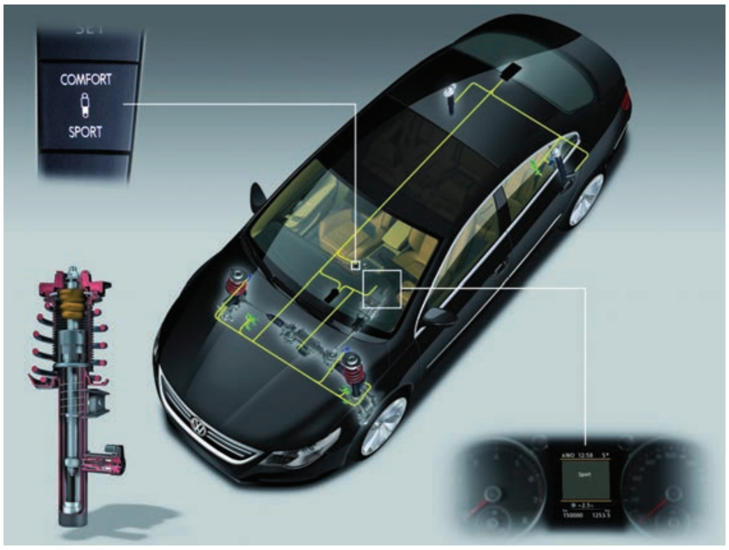


Figure 15.20 Active suspension also allows adjustments, in this case, between sport and comfort settings (Source: Volkswagen Media)

Active suspension systems usually consist of the following components:

- Electronic control unit (ECU).
- Adjustable dampers and springs.
- Sensors at each wheel and throughout the car.
- Levelling compressor (some systems).

Active suspension works by constantly sensing changes in the road surface and feeding that information, to the ECU, which in turn controls the suspension springs and dampers. These components then act upon the system to modify the overall suspension characteristics by adjusting damper stiffness, ride height (in some cases) and spring rate.

There are a number of ways of controlling the suspension. However, in most cases it is done by controlling the oil restriction in the damper. On some systems ride height is controlled by opening a valve and supplying pressurized fluid from an engine driven compressor. Later systems are starting to use special fluid in the dampers that reacts to a magnetic field, which is applied from a simple electromagnetic coil.

The improvements in ride comfort are considerable, which is why active suspension technology is becoming more popular. In simple terms, sensors provide the input to a control system that in turn actuates the suspension dampers in a way that improves stability and comfort.

Key fact

Active suspension works by sensing changes in conditions and feeding that information, to the ECU.

Figure 15.21 Suspension strut and actuator connection (Source: Delphi Media)

15.3.2 Sensors and actuators

To control the hydraulic units to the best advantage, the ECU needs to 'know' certain information. This is determined from sensor readings from various parts of the vehicle. A number of sensors are used to provide information to the suspension ECU (Figure 15.22).

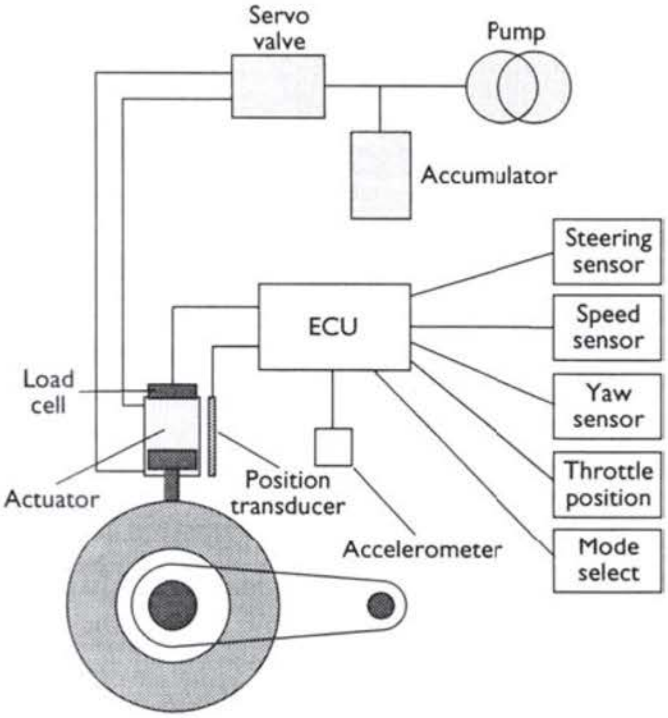


Figure 15.22 Sensors used to provide information to suspension ECU and general layout of an active suspension system

Load sensor

A load cell used to determine whether actual load is positioned on each hydraulic ram.

Displacement and vertical acceleration

This sensor can take a number of forms, as simple as a variable resistor or a more accurate and sensitive linear sensor such as the LVDT (see Chapter 2).

Lateral and longitudinal acceleration

Acceleration can be determined from a pendulum-type sensor using strain gauges linked to a mass, or devices similar to an engine knock sensor.

Yaw transducer

Yaw can be determined from lateral acceleration if the sensor is mounted at the front or rear of the vehicle.

Steering position

As well as steering position, rate of change of position is determined from a rotary position sensor. This device can be a light beam and detector type or similar. If the rate of change of steering position is beyond a threshold the system will switch to a harder suspension setting.

Vehicle speed

The speed of the vehicle is taken from a standard-type sensor as used for operating the speedometer.

Throttle position

Similar to the existing throttle potentiometers. This gives data on the driver's intention to accelerate or decelerate allowing the suspension to switch to a harder setting when appropriate.

Driver mode selection

A switch is provided allowing the driver to choose soft or hard settings. Even if the soft setting is selected, the system will switch to hard, under certain operational conditions.

The layout of the suspension system also shows a simplified view of the hydraulic unit. This is, in effect, a hydraulic ram and can have oil under very high pressure fed to the upper or lower chamber. The actual operation of the whole system is as follows. As a wheel meets a bump in the road there is increased upward acceleration and vertical load. This information is fed to the ECU, which calculates the ideal wheel displacement. A control signal is now sent to the servo valve(s), which control the position of the main hydraulic units. As this process can occur hundreds of times per second, the wheel can follow the contour of the road surface. This cushions the vehicle body from unwanted forces.

Key fact

As a wheel meets a bump in the road there is increased upward acceleration and vertical load.

By considering information from other sensors, such as the lateral acceleration sensor, which gives data relating to cornering, and the longitudinal sensor, which gives data relating to braking or acceleration forwards, the actuators can be moved to provide maximum stability at all times.

15.3.3 Delphi MagneRide

MagneRide was the industry's first semi-active suspension technology that employs no electro-mechanical valves and small moving parts (Figure 15.23). The MagneRide Magneto-Rheological (MR) fluid-based system consists of MR fluid-based single tube struts, shock absorbers (dampers), a sensor set and an on-board controller.

Figure 15.23 MagneRide suspension components (Source: Delphi Media)

Magneto-Rheological (MR) fluid is a suspension of magnetically soft particles such as iron microspheres in a synthetic hydrocarbon base fluid. When MR fluid is in the 'off' state, it is not magnetized, and the particles exhibit a random pattern. But in the 'on' or magnetized state the applied magnetic field aligns the metal particles into fibrous structures, changing the fluid rheology to a near plastic state (Figures 15.24 and 15.25).

By controlling the current to an electromagnetic coil inside the piston of the damper, the MR fluid's shear strength is changed, varying the resistance to fluid flow. Fine tuning of the magnetic current allows for any state between the low forces of 'off' to the high forces of 'on' to be achieved in the damper. The result is continuously variable real time damping.

The layout in Figure 15.26 shows the inputs and outputs of the MagneRide system. Note the connections with the ESP system and how the information is shared over the controller area network (CAN).

The MagneRide system produced by Delphi uses a special fluid in the dampers. The properties of this fluid are changed by a magnetic field. This allows for very close control of the damping characteristics and a significant improvement in ride comfort and quality.

Definition

Rheology is the study of friction between liquids.

Key fact

By controlling the current to an electromagnetic coil inside the piston of the damper, the MR fluid's shear strength is changed, varying the resistance to fluid flow.

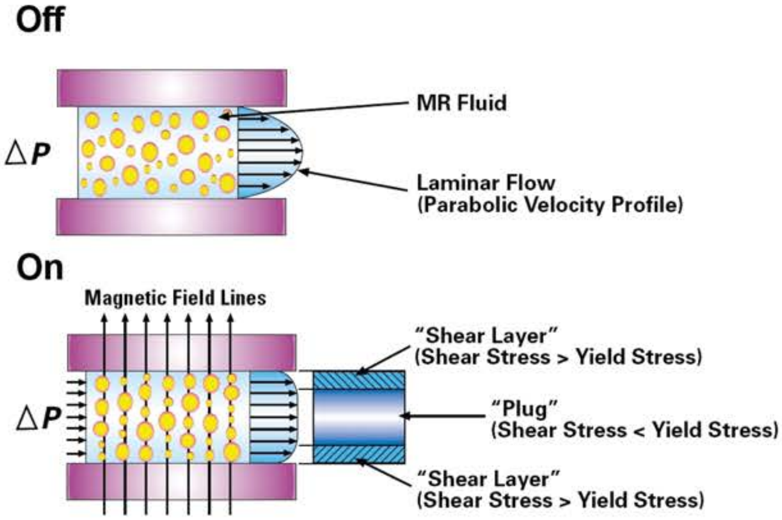


Figure 15.24 Fluid in the on and off states (Source: Delphi Media)

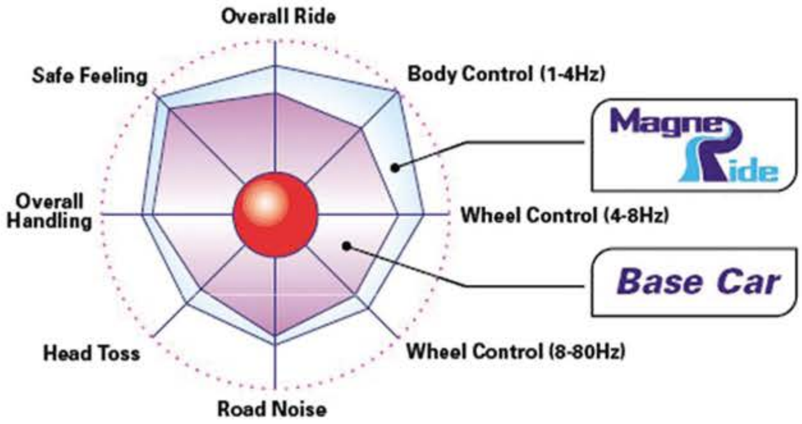


Figure 15.25 Representation of improvements when suspension is controlled (Source: Delphi Media)

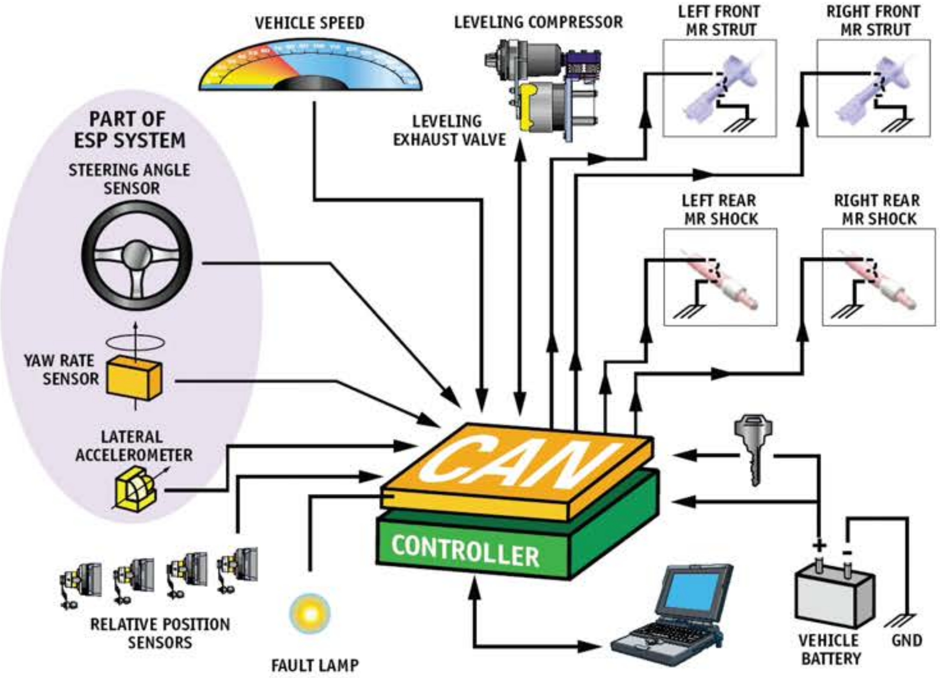


Figure 15.26 Control system (Source: Delphi Media)

Figure 15.27 System layout on an Audi (Source: Audi Media)

15.4 Automatic transmission

15.4.1 Introduction

The main aim of electronically controlled automatic transmission (ECAT) is to improve conventional automatic transmission in the following ways.

- Gear changes should be smoother and quieter.
- Improved performance.
- Reduced fuel consumption.
- Reduction of characteristic changes over system life.
- Increased reliability.

The actual operation of an automatic gearbox is beyond the scope of this book. However, the important points to remember are that gear changes and lock-up of the torque converter are controlled by hydraulic pressure. In an ECAT system, electrically controlled solenoid valves can influence this hydraulic pressure. Figure 15.28 is a block diagram of an ECAT system.

Most ECAT systems now have a transmission ECU (Figure 15.29) that is in communication with the engine control ECU (by a CAN databus in many cases). The system as a whole consists of a number of sensors providing data to the ECU, which in turn is able to control a number of actuators or output devices.

Key fact

In an automatic gearbox, gear changes and lock-up of the torque converter are controlled by hydraulic pressure.

15.4.2 Control of gear shift and torque converter

With an ECAT system, the actual point of gear shift is determined from pre-programmed memory within the ECU. Data from the sensors are used to

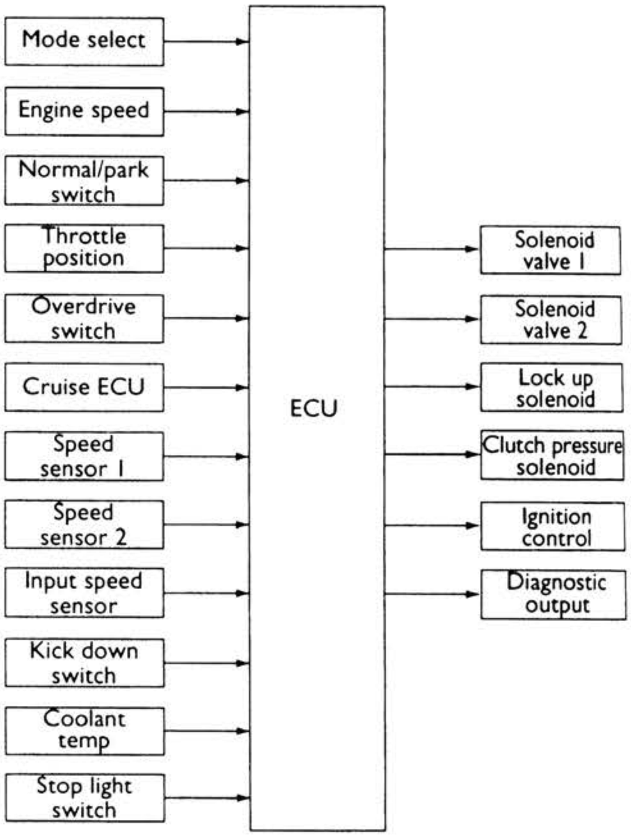


Figure 15.28 Block diagram of an ECAT system

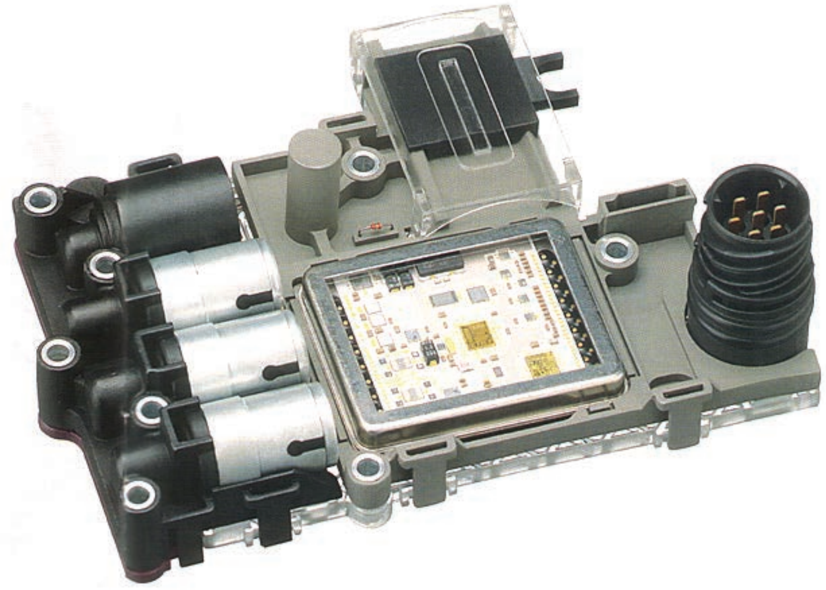


Figure 15.29 Electro-hydraulic valve block for controlling automatic transmission

Key fact

With an ECAT system, the actual point of gear shift is determined from pre-programmed memory within the ECU.

reference a look-up table mainly as a function of engine speed and vehicle speed. Data from other sensors are also taken into consideration. Actual gear shifts are initiated by changes in hydraulic pressure, which is controlled by solenoid valves.

The two main control functions of this system are hydraulic pressure and engine torque. The temporary reduction in engine torque during gear shifting (about 200 ms) allows smooth operation. This is because the peaks of gearbox output torque are suppressed, which causes the characteristic surge as the gears change on conventional automatics. Figure 15.30 shows a comparison of transmission output torque from systems with and without engine torque control. Also shown are the transmission speed and the timing control of the engine. Engine torque control can be by throttle movement, fuel cut-off or ignition timing retardation. The latter seems to have proved the most appropriate for modern systems.

Figure 15.31 shows how control of hydraulic pressure during gear up-shift again prevents a surge in transmission output torque. The hydraulic pressure control is in three stages as shown in the figure.

Basic pressure control

Pressure is set to an optimum value for speed of the gear shift. This can be adapted as the system learns the ideal pressure by monitoring shift time and changing the pressure accordingly.

Feedback control

The ECU detects the deviation of the rotational speed of the input shaft from a target value and adjusts pressure to maintain fine control.

Completion control

Torque converter hydraulic pressure is reduced momentarily so that as the engine torque output control is released, the potential surge is prevented. Because of these control functions, smooth gear shifts are possible and, due to the learning ability of some ECUs, the characteristics remain constant throughout the life of the system.

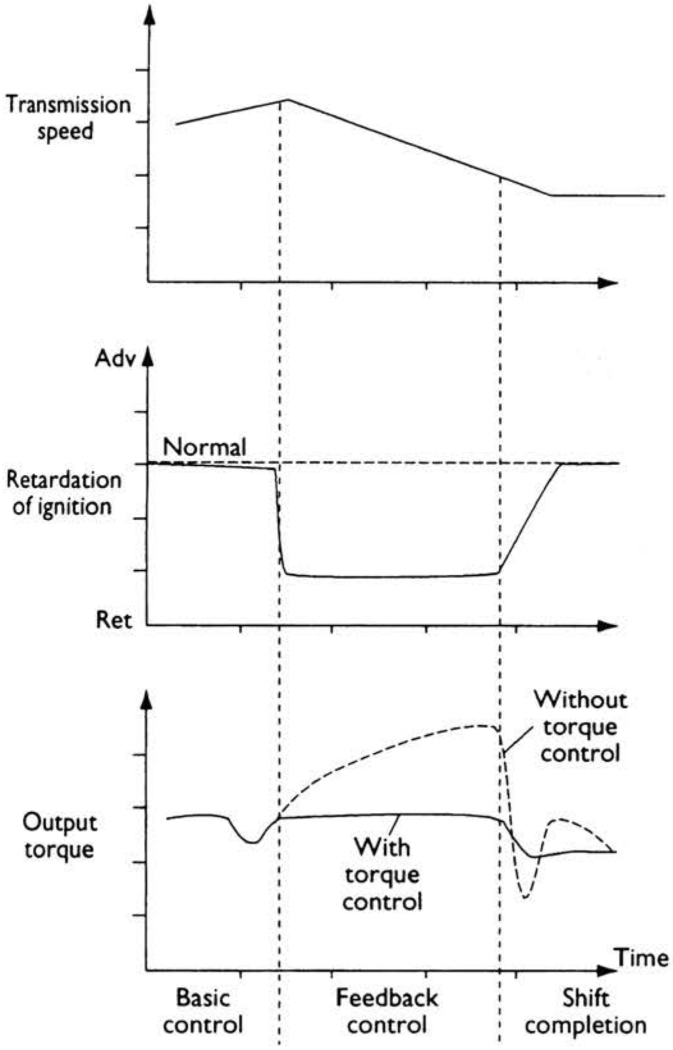


Figure 15.30 Transmission output torque from systems with and without engine torque control

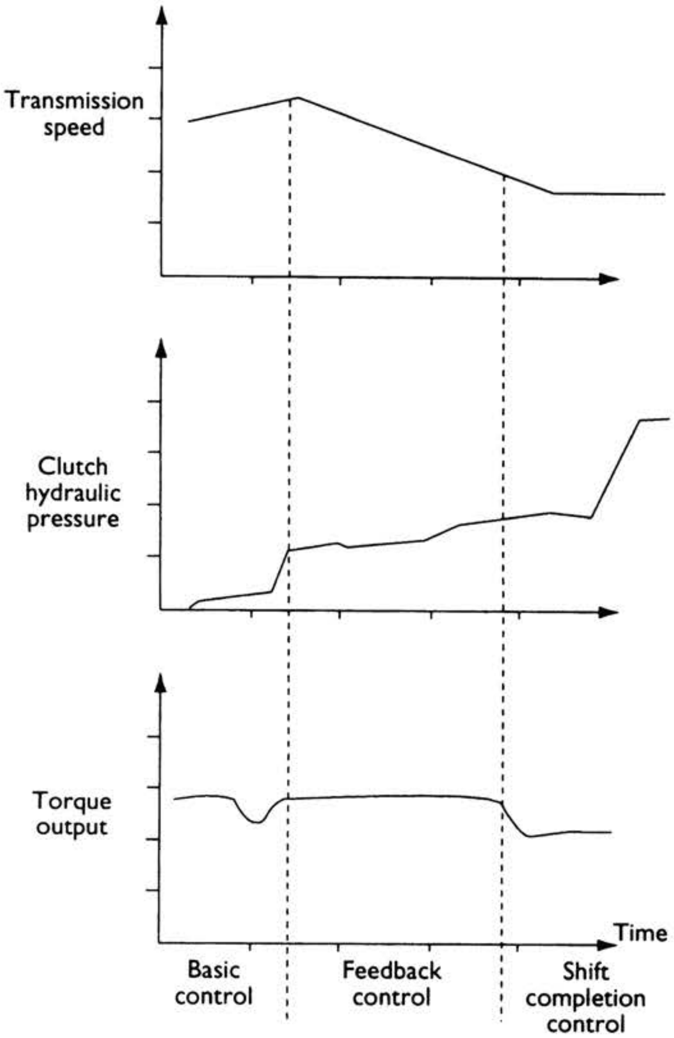


Figure 15.31 Control of hydraulic pressure

Torque converter lock-up

The ability to lock up the torque converter has been used for some time, even on vehicles with more conventional automatic transmission. This gives better fuel economy, quietness and improved driveability. Lock-up is carried out using a hydraulic valve, which can be operated gradually to produce a smooth transition. The timing of lock-up is determined from ECU memory in terms of the vehicle speed and acceleration.

15.4.3 Tiptronic

Developed by Porsche, Tiptronic S is a fully 'intelligent' multi-programme automatic transmission with additional fingertip control. The dual-function, 5-speed, automatic transmission, with active shift driving programmes is controlled by the 'Porsche Tiptronic' control system. As an alternative, and in addition to, the automatic mode, it is also possible to shift manually with fingertip controls.

Tiptronic first appeared in 1990 with technology directly descended from F1 and the 'Le Mans' Porsche 962s, which went on to win the 1994 'Le Mans 24 Hrs'.

Key fact

Tiptronic first appeared in 1990 with technology directly descended from F1 and the 'Le Mans' Porsche 962s, which went on to win the 1994 'Le Mans 24 Hrs'.

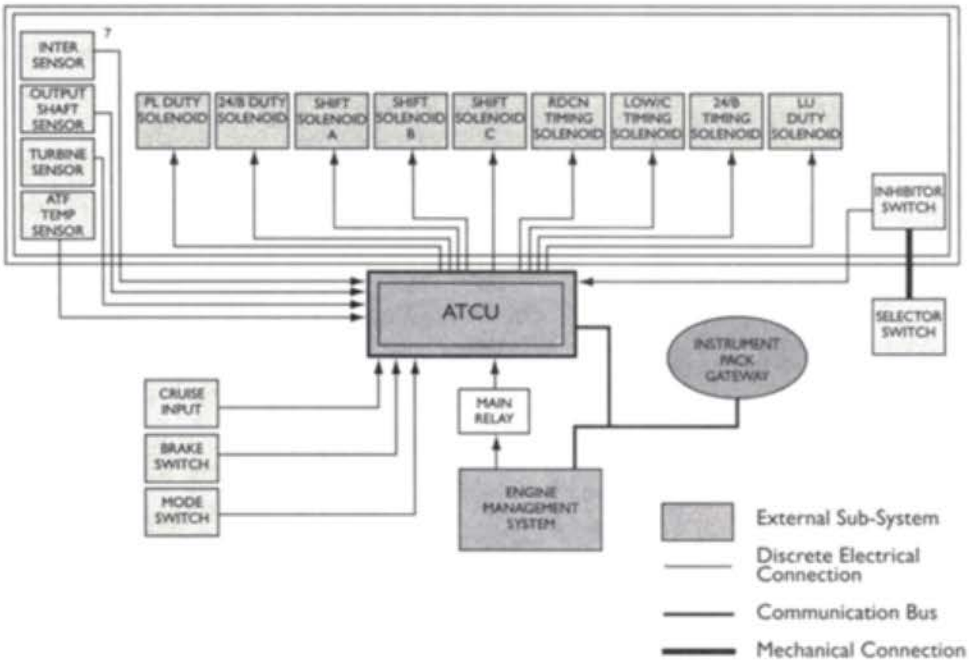


Figure 15.32 Example of a system shown as a block diagram

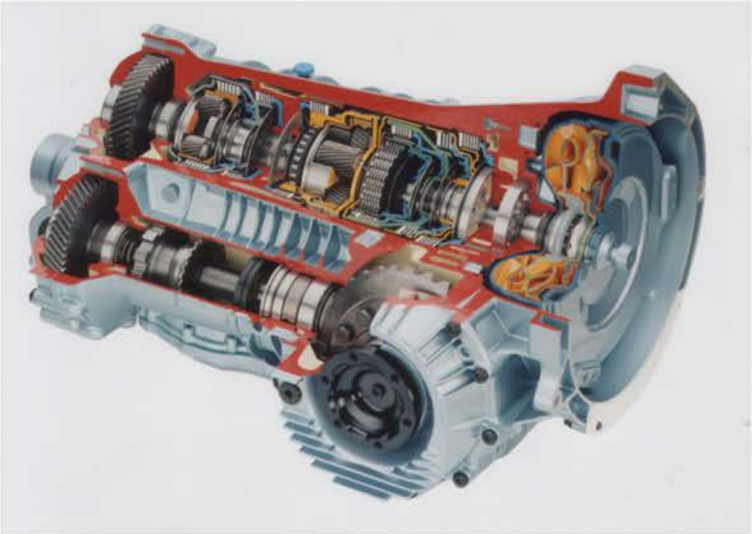


Figure 15.33 Porsche Tiptronic gearbox (Source: Porsche)

The Tiptronic S system features manual shift control integrated into the steering wheel. Once in manual mode, a driver can shift manually by pushing one of the two rocker switches mounted behind the steering wheel. Slight pressure of the thumb is all it takes to shift up – tipping it downward will produce a downshift.

The location and design of the rocker switches, as well as the distinctly perceptible pressure points, in combination with electronic transmission management, rule out any shift errors. The chosen gear is always indicated via a read-out located on the speedometer. The quick responses of the transmission triggered by a thumb push generates a spontaneous driving impression, gear changes being twice as quick as a manual gear-box.

The Tiptronic S system 'learns' about a particular driving style by monitoring eight sensors around the car, which include throttle speed and position, road speed, engine speed and temperature, lateral acceleration and deceleration. Redline-controlled protective programmes in the system prevent engine damage due to shift errors.

The shift patterns range from an economic variant for smooth motoring to dynamic motoring with the engine revving to maximum torque and power in the respective gears before changing, and downshifting appropriately from relatively high engine speeds. Rapid movements of the accelerator pedal, as well as hard acceleration, result in a graduated change of shift maps, right up the most extreme variant. In addition, the system is intelligent enough to react to other driving conditions and, for example, to downshift when braking before corners, which obviously reflects driving style with manual gearboxes.

This intelligent shift program (ISP) is characterized by the following special features in addition to the five automatic electronic shift maps.

- A warm-up program, which suppresses early upshifting to ensure a rapid rise to the engine operating temperature to ensure clean emissions.
- Active shifting – when the accelerator pedal is depressed and released rapidly, the most ‘dynamic’ shift map is available instantaneously.
- Suppression of the overrun up-shift on a sudden lifting of the throttle – e.g. no gear change mid-bend.
- Brake-initiated downshift to the next lower gear for more efficient engine braking.
- Holding onto a gear in curves – i.e. no gear change whilst in mid-bend.
- Graduated up-shifting from lower gears to prevent immediate change over to the top gear, especially after active downshifting.
- Identification of uphill stretches to stay in the lower gears as long as possible when driving up or downhill.
- Slip-induced up-shifting initiated by inertia forces when braking on slippery surfaces (rain, snow) to improve lateral guidance of the driving wheels and consequently driving stability.

15.4.4 Summary

The use of integrated intelligent control of both the engine and transmission allows considerable improvements to the operation of automatic gearboxes. Improvements are possible to efficiency, performance and smoothness of operation. Extra facilities become available to the driver such as being able to select the desired mode of operation. This can be a choice between, for example, performance and economy. The tie up between engine control and transmission control helps to illustrate how it is becoming more difficult to consider vehicle systems as isolated units and how more consideration must be given to the overlap of the system boundaries.

15.5 Other chassis electrical systems

15.5.1 Electric power steering

There are two main ways of using electric power for steering assistance (the second is now the most common):

1. Replacing the conventional system pump with an electric motor whilst the ram remains much the same.

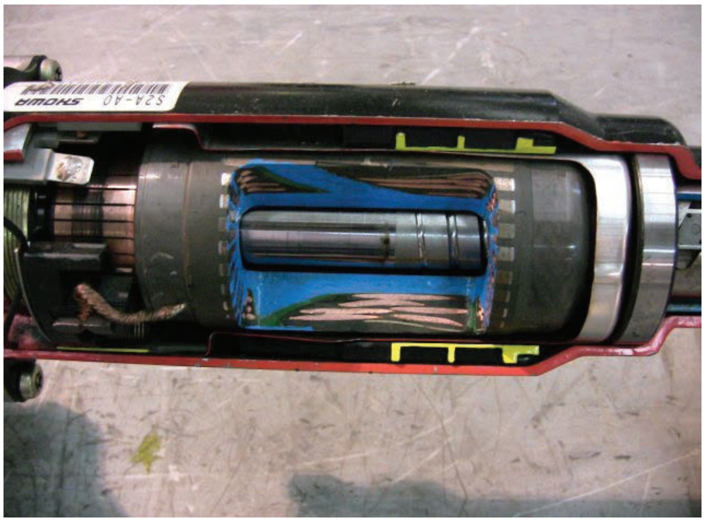


Figure 15.34 Direct acting motor

2. Using a drive motor, which directly assists with the steering and has no hydraulic components.

Key fact

With a direct acting type an electric motor works directly on the steering via an epicyclic gear train.

With a direct acting type an electric motor works directly on the steering via an epicyclic gear train. This completely replaces the hydraulic pump and servo cylinder. This eliminates the fuel penalty of the conventional pump and greatly simplifies the drive arrangements. Engine stall when the power steering is operated at idle speed is also eliminated.

On many systems, an optical torque sensor is used to measure driver effort on the steering wheel (all systems use a sensor of some sort). The sensor works by measuring light from an LED, which is shining through holes. These are aligned in discs at either end of a torsion bar, fitted into the steering column. An optical sensor element identifies the twist of two discs on the steering axis with respect to each other, each disc being provided with appropriate codes. From this sensor information the system calculates the torque as well as the absolute steering angle.

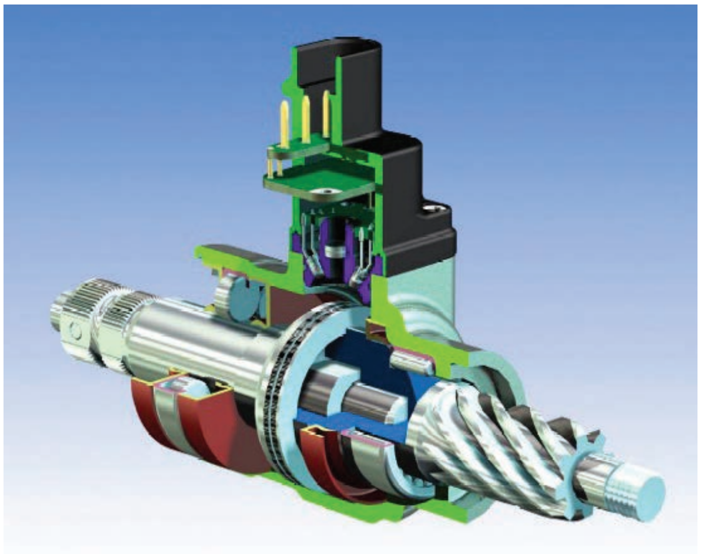


Figure 15.35 Steering sensor (Source: Bosch Media)

Electrical PAS occupies little under bonnet space, something that is at a premium these days, and the 400 W motor only averages about 2 A under urban driving conditions. The cost benefits over conventional hydraulic methods are considerable.

Electric power for steering assistance can be applied in a number of ways. However, using a drive motor, which directly assists with the steering, is the most common. This method uses less power and takes up less space than other methods.

Key fact

Electrical PAS occupies little under bonnet space and the motor only averages about 2A current draw under urban driving conditions.

15.5.2 Robotized manual transmission

The electronic clutch was developed for racing vehicles to improve the getaway performance. For production vehicles, a strategy has been developed to interpret the driver's intention. With greater throttle openings, the strategy changes to prevent abuse and drive line damage. Electrical control of the clutch release bearing position is by a solenoid actuator, which can be modulated by signals from the ECU. This allows the time to reach the ideal take-off position to be reduced and the ability of the clutch to transmit torque to be improved. The efficiency of the whole system can therefore be increased. Figure 15.37 shows the torque transmitted curve for an electronic clutch system. A switch could be provided to change between performance and economy mode.

An automatic clutch eliminates the need for a clutch pedal. Cars equipped with this are therefore more comfortable and more fun to drive as the driver is freed from the tiresome effort of depressing and releasing the clutch pedal every time he or she changes gear. The gear lever remains, however, leaving active control of the car with the driver.

The actuator shown as Figure 15.38 is an add-on system, which can be fitted to conventional manual transmissions. It consists of a clutch actuator, a powerful CPU and specific sensors driven by dedicated software that is optimized for each vehicle type.

The clutch actuator uses electromechanical actuation. It is therefore more compact, weighs less and costs less compared with hydraulic systems. Its internal compensation spring allows for very fast response time (de-clutching time: 70–100 ms) combined with low electrical consumption (20 W). Its 16-bit electronic control unit and power electronics were developed and produced by Valeo Electronics.

Figure 15.36 Electric PAS (Source: Ford Media)

Key fact

The electronic clutch was developed for racing vehicles to improve the getaway performance.

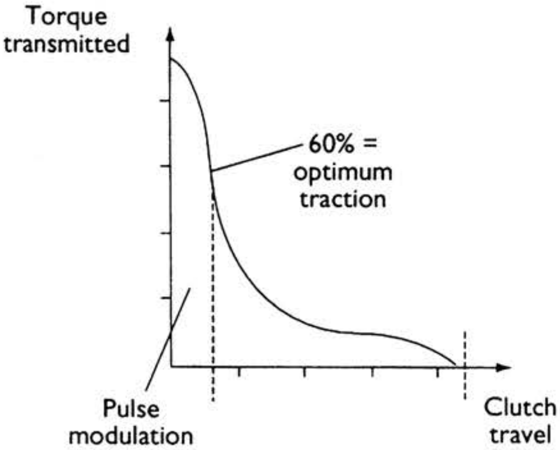


Figure 15.37 Torque transmitted curve for an electronic clutch system

Figure 15.38 Automatic clutch actuator (Source: Valeo)

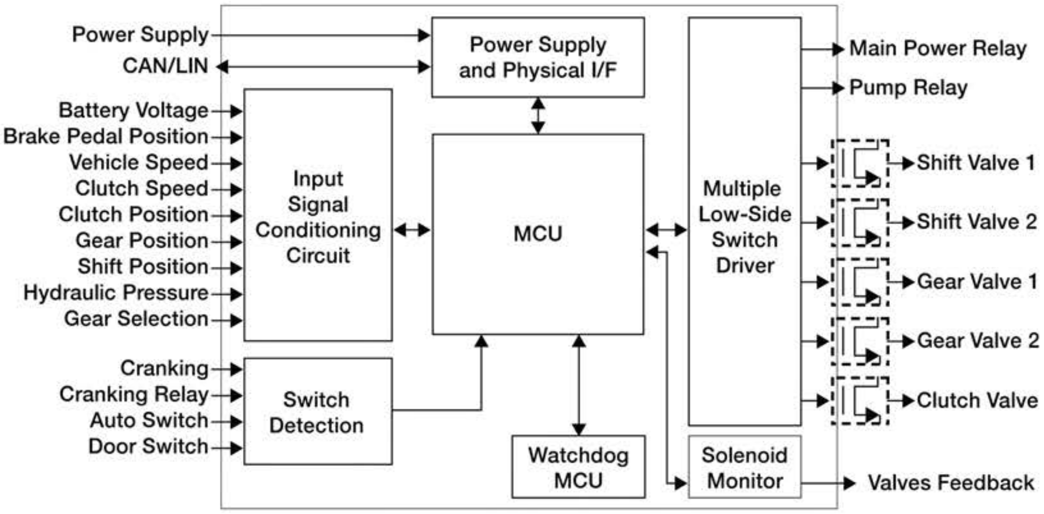


Figure 15.39 Block diagram showing inputs, control and outputs of an automated manual transmission (Source: Freescale Electronics)

Key fact

When combined with a gear shift actuator, fully automatic operation is possible with a 'manual' gearbox.

When combined with a gear shift actuator, fully automatic operation is possible with a 'manual' gearbox. This is known as robotized manual transmission.

15.5.3 Active roll reduction

The conventional anti-roll bar, as fitted to many vehicles, is replaced with a bar containing a rotary actuator. The actuators are hydraulically operated from a dedicated pump. Lateral acceleration is calculated by the ECU from steering angle and road speed. Hydraulic pressure is then regulated as required to the front and rear actuators such as to provide a force on the roll bar preventing the body of the vehicle from tilting. A good use for this system is on larger panel vans although it is being offered as an option to a range of vehicles. Figure 15.40 shows the positioning of one of the actuators for active roll reduction.

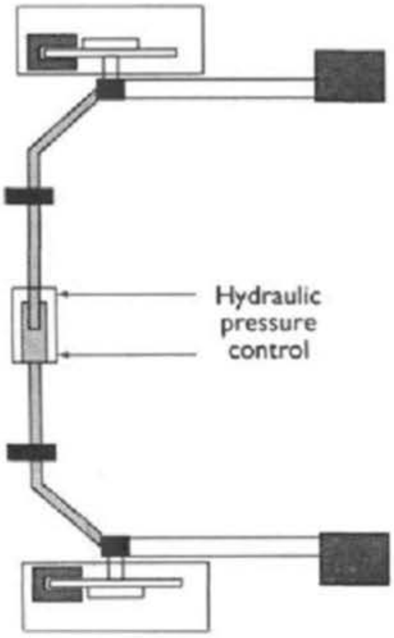


Figure 15.40 Positioning of one of the actuators for active roll reduction

15.5.4 Electronic limited slip differential

Conventional limited slip differentials (LSDs) cannot be designed for optimum performance because of the effect on the vehicle when cornering. Their characteristics cannot be changed when driving. Front-wheel drive vehicles have even more problems due to the adverse effect on the steering. These issues have prompted the development of electronic control for the LSD.

The slip limiting action is controlled by a multi-disc clutch positioned between the crown wheel and the differential housing. It is able, if required, to lock the axle fully. The pressure on the clutch plates is controlled by hydraulic pressure, which in turn is controlled by a solenoid valve under the influence of an ECU. Data are provided to the ECU from standard ABS-type wheel sensors. Figure 15.41 shows a block diagram of a final drive and differential unit with electronic control.

Key fact

The pressure on the clutch plates is controlled by hydraulic pressure, which in turn is controlled by a solenoid valve under the influence of an ECU.

15.5.5 Brake assist systems

Brake assist systems may be developed because of evidence showing that drivers are not realizing the full benefit of anti-lock braking (ABS). The introduction of ABS has not resulted in the reduction of accidents that had been hoped for. The reason for this is debatable; one view is that many drivers do not push hard enough on the brake pedal during an emergency; therefore the tyres do not slip sufficiently to engage the anti-lock system. To counteract this problem, companies are developing brake assist systems that apply more hydraulic pressure than normal if an emergency condition is sensed. The system's ability to discern whether a braking operation is an emergency, or not, is critical. Pedal force sensors only, as well as pedal force sensors in conjunction with apply rate sensors, are under development, as are the control strategies. If field tests produce satisfactory results, brake assist systems could be introduced relatively quickly into mass production.

Key fact

Brake assist systems may be developed because of evidence showing that drivers are not realizing the full benefit of anti-lock braking.

Electric actuators may even begin to take the place of conventional wheel cylinders. Precisely controlled DC motors operating on drum brakes have the

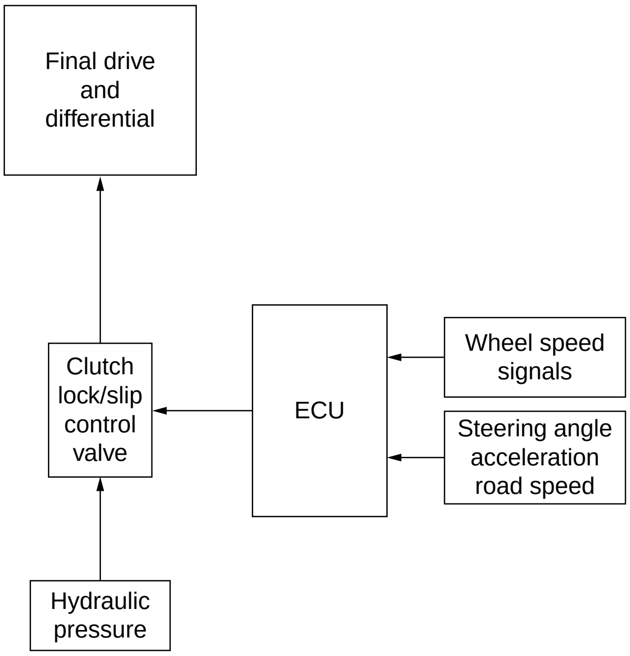


Figure 15.41 Final drive and differential unit electrical control

potential advantages of lower total system weight and cost. Developments are occurring in the area of magnetic braking, which has the potential to remove all wearing components from the vehicle!

Electric parking brakes (EPB) systems are in use on some vehicles. Some simply use a motor to drive, via a reduction gear a screw system that pulls on the existing handbrake cables. A further development is to use a small motor on two (or more) calipers that screws a plunger to press the pads on to the disc. Both systems suffer in that an electrical failure can cause the brake to stick on – or off. A manual override is used in many cases.

15.5.6 X-by-wire

15.5.6.1 Introduction

Key fact

The term 'X-by-wire', is used to represent any mechanical technology on the vehicle that is operated electrically.

The term 'X-by-wire', is used to represent any mechanical technology on the vehicle that is operated electrically. In some areas this has been used for many years. 'Window lift-by-wire', would be a good example. However, the term tends to be used now to represent systems that have not, traditionally, been electrically operated: brake-by-wire and steer-by-wire are two such areas. The industry is going through a development stage that will lead to some level of standardization in the deployment of X-by-wire systems. In particular the areas of failure tolerance and communication protocols are developing. This section will examine the emerging technologies relating to gas-, steer- and brake-by-wire systems.

Interestingly, X-by-wire systems have been in use for many years in the aircraft industry and appear to be readily accepted by the general public. However, the concept of brake-by-wire on a car seems to cause great concern over safety. This may be due to the lack of regulation in the repair and service industry!

15.5.6.2 Gas-by-wire

The concept of gas-by-wire is already accepted and in use on many vehicles. This includes injection, EGR, electric supercharging and throttle-by-wire (sometimes referred to as ETC). Injectors have been electrically operated for many years as has the actuator for the EGR system. In the throttle-by-wire system, a sensor and an actuator replace the traditional cable. The sensor is, in most cases, a variable resistor with suitable spring pressure built in to maintain an appropriate feel. The actuator designs vary but a stepper motor is a common choice because of its degree of control.

Figure 15.42 Gas-by-wire GDI components (Source: Bosch Press)

Electric supercharging is an interesting development; it is particularly useful for gasoline direct injection engines, where it improves performance generally but also prevents 'turbo lag'. The development of gas-by-wire systems will continue because of the pressure to reduce CO_2 emissions. This is leading to the development of smaller, more efficient, engines.

15.5.6.3 Steer-by-wire

Currently all series production power steering systems maintain a mechanical connection between the vehicle's front wheels and the steering wheel. If the assistance system, be it electric or hydraulic, were to fail, the mechanical link still works as a backup. Further, current regulations require this mechanical connection to be in place. However, the rigid mechanical connection is a major drawback as far as the systems functional features are concerned. Issues such as noise, vibration and harshness (NVH) and crashworthiness are also drawbacks of the rigid system.

Advances in mechatronic systems mean that the rigid link may soon be replaced – with wires. Steer-by-wire vehicles transfer the rotation of the steering wheel to front wheel movement by using sensors and use an electronically controlled actuator in place of the conventional steering rack. Feedback, an important characteristic of a steering system, is generated for the driver by a force feedback actuator behind the steering wheel. A change to the regulations relating to the rigid link are being replaced by a regulation that relates to integrity requirements.

Clearly the development of a steer-by-wire system is determined by the reliability of the components used. Many current developments relate to 'fault-tolerant system architecture'. A failure rate of less than 10^{-7} fatal failures per hour of operating time is the aim. This cannot currently be achieved using single-channel electronic control units (ECUs). To achieve an integrity value comparable with mechanical link systems, steer-by-wire must be able to tolerate single electrical or electronic faults in any of its sub-systems. It must also include a method of detecting these faults. This tolerance would therefore exclude the possibility of sudden fatal failure. Appropriate fault handling may involve a limit to vehicle speed or in critical conditions would prevent the vehicle from being driven.

Key fact

To achieve an integrity value comparable with mechanical link systems, steer-by-wire must be able to tolerate single electrical or electronic faults in any of its sub-systems.

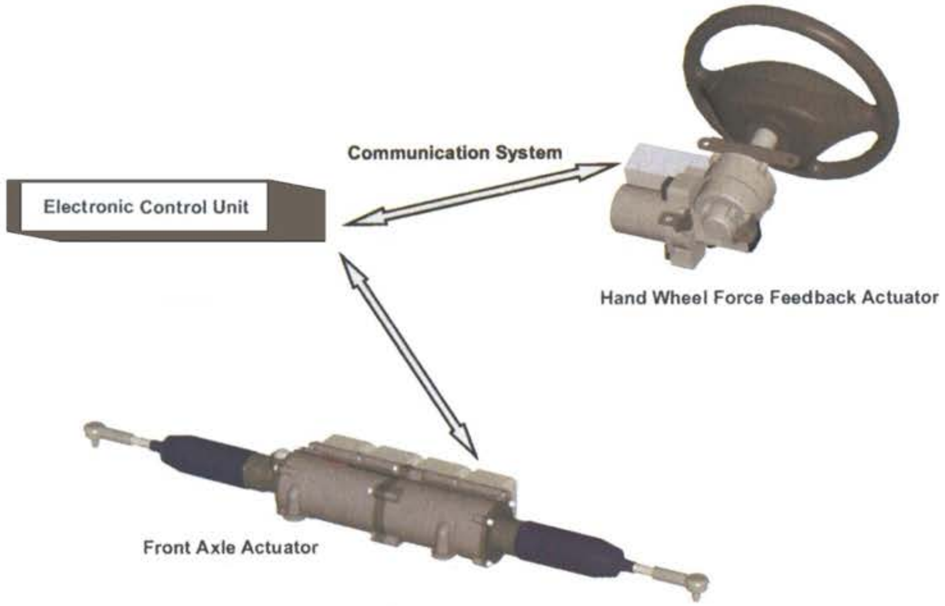


Figure 15.43 Fault-tolerant steer-by-wire system layout (Source: TRW Automotive)

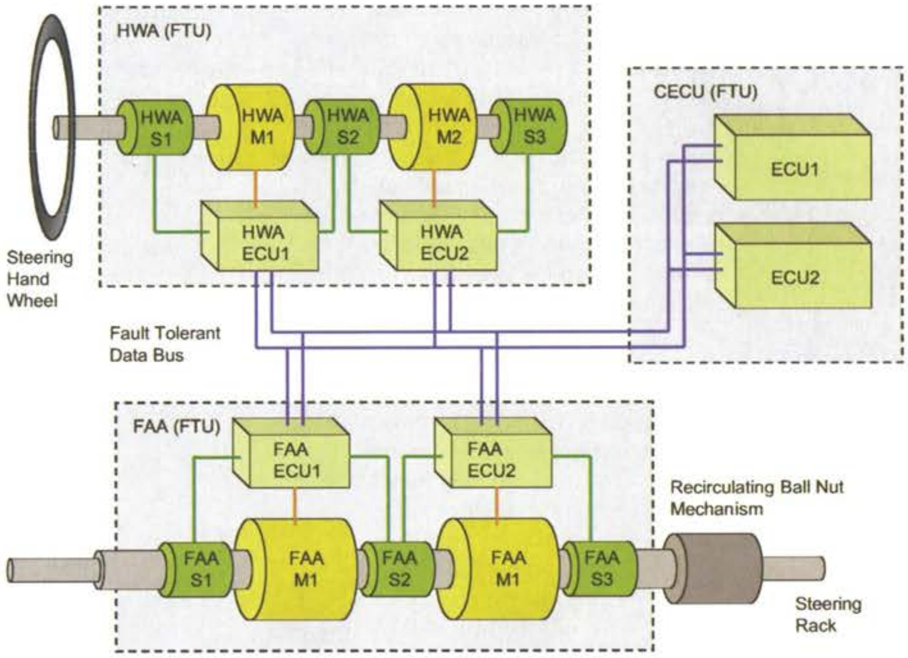


Figure 15.44 Steer-by-wire system architecture (Source: TRW Automotive)

The force feedback to the steering wheel is often considered to be less safety-critical. However, for high speed passenger vehicle use, the response time of the driver may be critical. For this reason, the feedback actuator must also be part of the fault-tolerant system. The overall architecture of a fault- tolerant steer-by-wire system must include significant redundancy. In simple terms, almost all components are duplicated and must include a fault-tolerant power supply.

In Figure 15.44, the force-feedback actuators are labelled HWA and the 'steering rack' actuators FAA. The steer-by-wire electronic control unit is labelled CECU and contains two ECUs. The labels M# and S# relate to motors and sensors respectively.

The power supply is critical for any X-by-wire system. A mid-sized car will require a peak power output of about 1000 W for full steering performance. The overall consumption is relatively low but because of the peak loading. The electric drive units and the ECUs a dual redundant supply.

The many advantages that steer-by-wire will bring tend to suggest that is will soon be available. TRW Automotive, a well-known and respected OEM, said in 2003 that steer-by-wire would be ready for production by 2007 (Heitzer 2003).

15.5.6.4 Brake-by-wire

Many aspects of the brake-by-wire field are already quite advanced. However, it is becoming clear that full electrical operation of the brakes, i.e. removal of the hydraulic/mechanical link, is some years away yet. However, the functions of the braking system are undergoing a smooth and continuous evolution (Kelling and Leteinturier 2003).

There are significant operational and constructional advantages to having full brake-by-wire. Some of these are as follows:

- Safety – Improved reaction time of just 0.5 s could decrease death from front end collisions by between 30 and 50%.
- Environment – Hydraulic brake fluid is poisonous and requires changing during the vehicle lifetime.

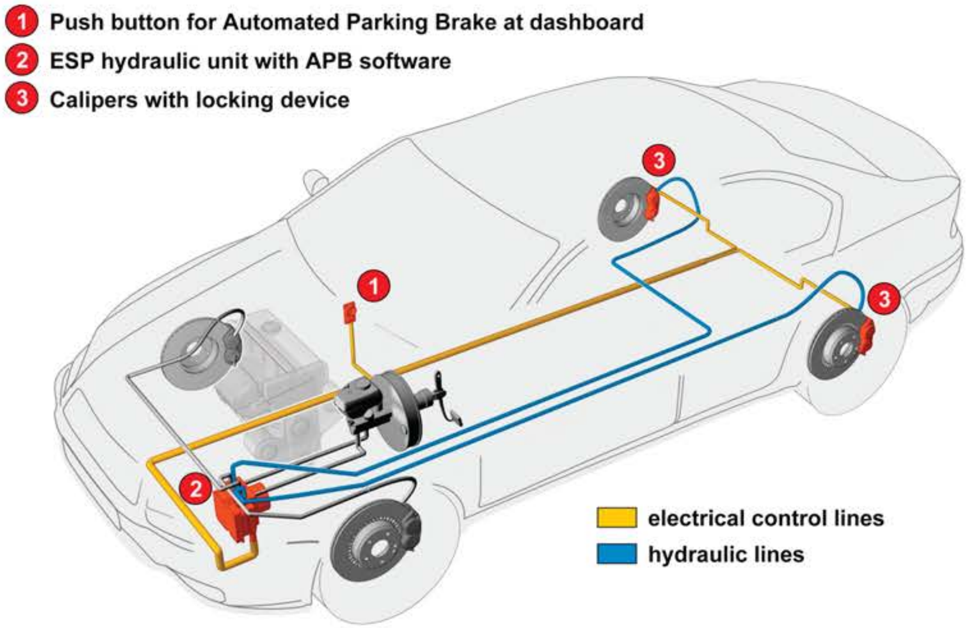


Figure 15.45 Electrohydraulic Brakes - EHB (Source: Bosch Press)

- Control – A consistent and integrated approach will enhance other functions such as adaptive cruise control and stability control.
- Comfort – Lower and adjustable pedal force as well as features such as 'hill-holding' enhance the driver experience.

The need for a fault-tolerant electrical system and its additional cost, mean that all current developments have retained the hydraulic system. The following figure shows the evolution and future projection for brake system developments.

In 1978, Bosch launched the first electronically controlled anti-lock braking system (ABS); nine years later came the traction control system (TCS). The next innovation was the Electronic Stability Program (ESP) in 1995. The most advanced system in current production is the electro-hydraulic brake (EHB), also known as Sensotronic Brake Control (SBC) (Plapp and Dufour 2003).

The brake system can perform all the driver-related additional functions by electro-hydraulic means (by wire), without requiring complex and expensive changes to the vehicle's electrical system. A range of new safety and/or convenience features is under development:

- Electronic Brake Prefill – If the driver lifts their foot suddenly off the accelerator, the brake system will deduce that there is a potential emergency. The brake pads are immediately moved into contact with the brake discs, so that there is no delay in slowing the vehicle down if emergency braking is undertaken.
- Brake Disc Wiping – In heavy rain, a film of moisture forms on the brake discs. The brake pads are made to touch the discs briefly on a regular basis, removing the film of water and helping the brakes to bite more quickly.
- Soft-Stop – This facilitates smooth, jerk-free stopping by reducing braking pressure shortly before the vehicle comes to rest.
- Hill Hold Control – prevents unintentional rolling backwards on hill starts. The brake system automatically maintains the braking pressure and stops the vehicle rolling backwards until the driver presses the accelerator again.
- Stop & Go – expands the Adaptive Cruise Control (ACC) distance control system. Using data from distance sensors, this function can automatically

bring the vehicle to a complete halt and then move it forward again when traffic allows, without the driver needing to do anything.

An automatic parking brake (APB) is another attractive function offering the driver increased comfort and convenience. Since the handbrake lever is dispensed with, car manufacturers have more freedom of choice as to where they site the operating parts within the car. The technical principle involved can be compared with that of a ball-point pen, where the ink cartridge is pushed out by finger pressure and then held in position with a locking mechanism, until the button is pressed once again.

When the driver presses the switch to activate the parking brake, the ESP unit automatically generates pressure in the braking system and presses the brake pads against the disc. The calipers are then locked in position – an electrically controlled magnetic valve built into the caliper operates the locking mechanism hydraulically. The caliper then remains locked without any hydraulic pressure. To release the brake, the ESP briefly generates pressure again, slightly more than was needed to lock the caliper (Figure 15.46).

Development of brake-by-wire will not stop – because it has the potential to improve significantly the way in which the vehicle will stop!

15.5.7 Diagnosing chassis electrical system faults

Table 15.1 lists some common symptoms of chassis electrical system malfunctions together with suggestions for the possible fault. The faults are generic but will serve as a good reminder.

‘Chassis electrical systems’ covers a large area of the vehicle. The generic fault-finding lists presented in other chapters may be relevant but the technique that will be covered here is known as ‘black box fault-finding’. This is an excellent technique and can be applied to many vehicle systems from engine management and ABS to cruise control and instrumentation.

As most systems now revolve around an ECU, the ECU is considered to be a ‘black box’, in other words we know what it should do but how it does it is

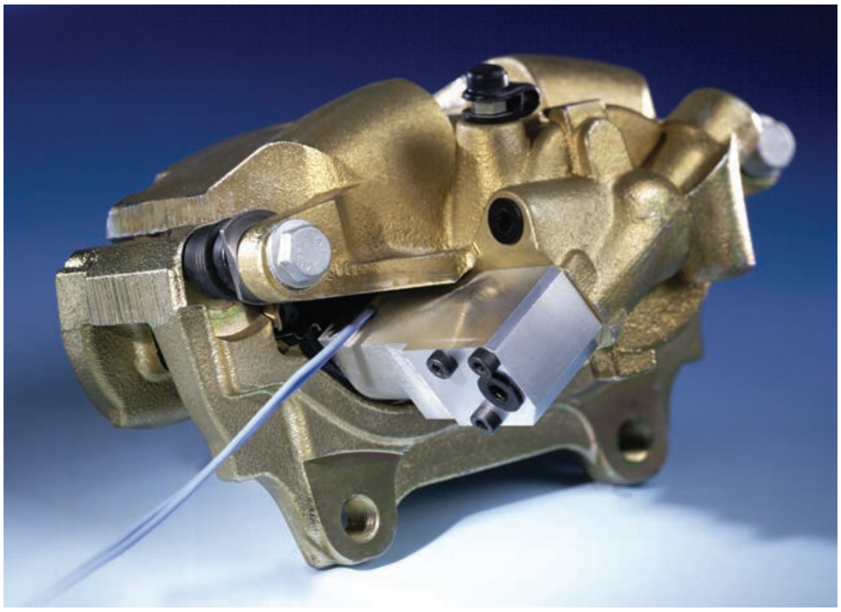


Figure 15.46 Electrically operated parking brake caliper (Source: Bosch Press)

Table 15.1 Chassis system symptoms and possible faults

Symptom	Possible fault
ABS not working and/or warning light on	Wheel sensor or associated wiring open circuit/high resistance Wheel sensor air gap incorrect Power supply/earth to ECU low or not present Connections to modulator open circuit No supply/earth connection to pump motor Modulator windings open circuit or high resistance
Traction control inoperative	Wheel sensor or associated wiring open circuit / high resistance Wheel sensor air gap incorrect Power supply/earth to ECU low or not present ABS system fault Throttle actuator inoperative or open circuit connections Communication link between ECUs open circuit
ECAT system reduced performance or not working	Communication link between engine and transmission ECUs open circuit Power supply/earth to ECU low or not present Transmission mechanical fault Gear selector switch open/short circuit Speed sensor inoperative
Power steering assistance low or not working	Power supply/earth to ECU low or not present Mechanical fault Power supply/earth to drive motor low or not present Steering sensor inoperative

irrelevant! ‘Any colour, so long as it’s black,’ said Henry Ford in the 1920s. I doubt that he was referring to ECUs though.

Figure 15.47 shows a block diagram that could be used to represent any number of automobile electrical or electronic systems. Treating the ECU as a ‘black box’ allows us to ignore its complexity. The theory is that if all the sensors and associated wiring to the ‘black box’ are OK, all the output actuators and their wiring are OK and the supply/earth connections are OK, then the fault must be the ‘black box’. Most ECUs are very reliable, however, and it is far more likely that the fault will be found in the inputs or outputs.

Normal fault-finding or testing techniques can be applied to the sensors and actuators. For example, if an ABS system uses four inductive-type wheel speed sensors, then an easy test is to measure their resistance. Even if the correct value were not known, it would be very unlikely for all four to be wrong at the same time, so a comparison can be made. If the same resistance reading is obtained on the end of the sensor wires at the ECU, almost all of the ‘inputs’ have been tested with just a few ohmmeter readings.

The same technique will often work with ‘outputs’. If the resistance of all the operating windings in, say, a hydraulic modulator were the same, then it would be reasonable to assume the figure was correct.

Sometimes, however, it is almost an advantage not to know the manufacturer’s recommended readings. If the ‘book’ says the value should be between 800 and 900, what do you do when your ohmmeter reads 915? Answers on a postcard please…

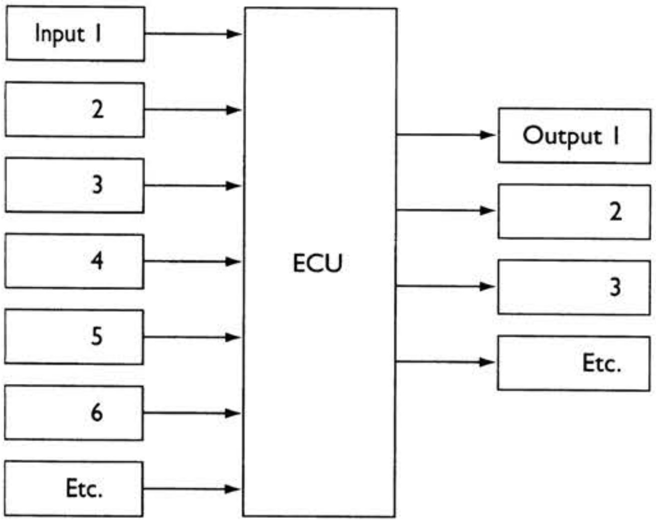


Figure 15.47 Block diagram representing many electrical systems

Finally, don't forget that no matter how complex the electronics in an ECU, they will not work without a good power supply and an earth.

15.6 Advanced chassis systems technology

15.6.1 Road surface and tyre friction

The friction between the tyre and the road surface is a key issue when considering anti-lock brakes. Frictional forces must be transferred between the tyre contact patch and the road surface when the vehicle is accelerating or braking. The normal rules for friction between solid bodies have to be adapted because of the springy nature of rubber tyres. To get round this complicated problem which involves a molecular theory, the term 'slip' is used to describe the action of tyre and road.

Slip occurs when braking effort is applied to a rotating wheel. This can be defined as follows:

$$\lambda = \frac{\omega_0 - \omega}{\omega_0} \times 100\%$$

or

$$\lambda = \frac{V_v - V_r}{V_v} \times 100\%$$

0% is a free rolling wheel and 100% is a locked wheel

where: λ *= slip,* ω_o *= angular velocity of freely rotating wheel,* ω *= angular velocity of braked wheel,* V_v *= vehicle road speed (*$\omega_o r_d$*),* V_r *= circumferential velocity of braked wheel (*ωr_d*),* r_d *= dynamic rolling radius of the wheel.*

The braking force or the adhesion coefficient of braking force (µF), measured in the direction the wheel is turning, is a function of slip. µF depends on a number of factors; the main ones are listed:

- Road surface material/condition.
- Tyre material, inflation pressure, tread depth, tread pattern and construction.
- Contact weight.

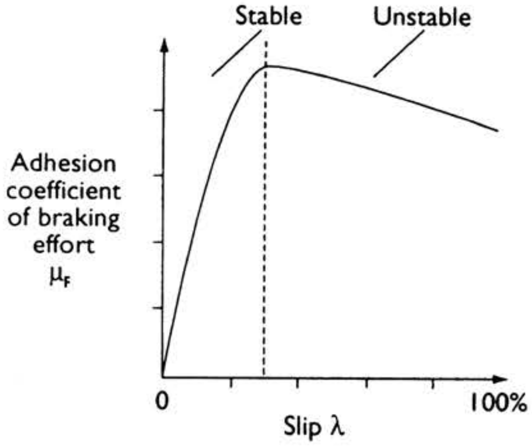


Figure 15.48 Relationship between adhesion and coefficient of braking effort and amount of slip

Figure 15.48 shows the relationship between the adhesion coefficient of braking effort and the amount of slip. Note that the graph is divided into two areas, stable and unstable. In the stable zone a balance exists between the braking effort applied and the adhesion of the road surface. Non-slip braking is therefore possible. In the unstable zone when the critical slip (l_c) is passed, no balance exists and the wheel will lock unless the braking force is reduced.

The value of critical slip (l_c) can vary between about 8 and 30% depending on the tyres and the road surface conditions. Figure 15.49 shows the difference between road surface conditions. This serves to highlight that a fixed slip threshold as a reference point, for when ABS should operate, would not make the best use of the available adhesion coefficient.

Lateral slip of the vehicle wheels must also be considered. This occurs when the wheel centre line forms an angle of drift with the intended path of the wheel centre. The directional movement of the vehicle is defined as the correlation between the slip angle and the lateral force. This is shown as Figure 15.50 which is a graph of the coefficient of adhesion for lateral force, designated as μ_L, against slip angle (α). The critical slip angle (α_c) lies in general between 12 and 15°.

To regulate braking, it is essential that braking force and lateral guidance forces be considered. Figure 15.51 shows the combination of adhesion coefficient (μ_F), and the lateral adhesion coefficient (μ_L), against braking slip (λ). The slip angle

Key fact

The directional movement of a vehicle is defined as the correlation between the slip angle and the lateral force.

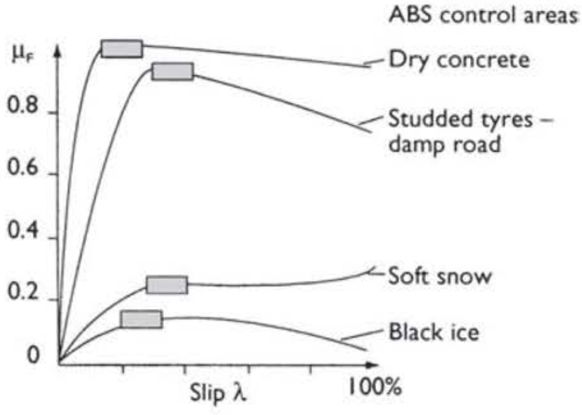


Figure 15.49 Difference between road surface conditions

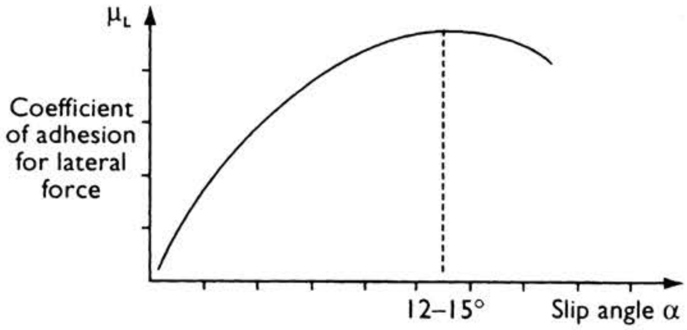


Figure 15.50 Graph of the coefficient of adhesion for lateral force against slip angle

is shown at 2^0 and 10^0 and the test is on a dry road. Note the considerable reduction in lateral adhesion (μ_L) when the braking slip (λ) increases. When $\lambda = 1$. the value of μ_L is as a result of the steered angle of the wheel. This can be calculated:

$$\mu_L \text{ (min)} = \mu_F \sin\alpha$$

This serves to demonstrate how a locked wheel provides little steering effect. ABS control must be extended for larger slip angles. If full braking occurs when the vehicle is experiencing high lateral acceleration (larger α) then ABS must intervene early and progressively allow greater slip as the vehicle speed decreases. This data is stored in look-up tables in a read-only memory in the electronic control unit.

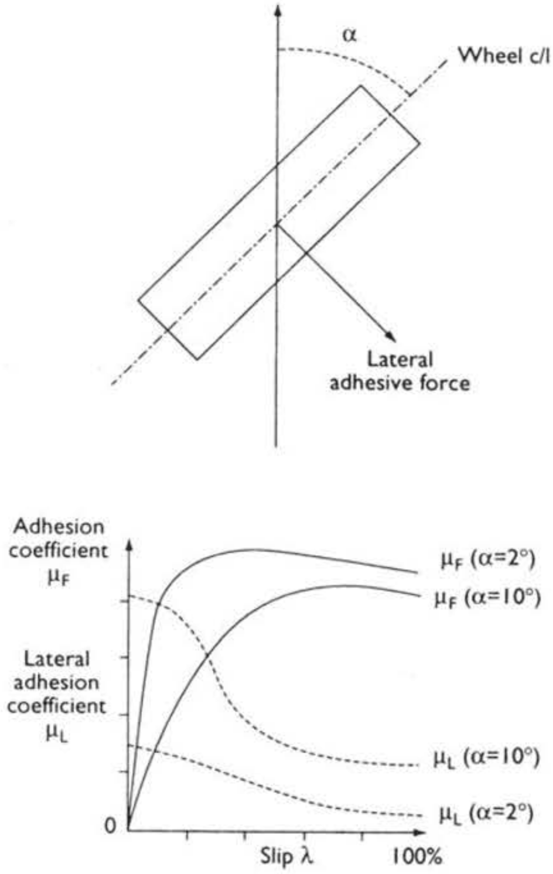


Figure 15.51 The combination of adhesion coefficient and lateral adhesion coefficient against braking

15.6.2 ABS control cycles

Figure 15.52 shows the braking control cycle for a high adhesion road (good grip) and Figure 15.53 shows the same for a low adhesion (slippery) road. The figures are split into eight phases which are described below.

High adhesion

1. Initial braking, ABS not yet activated.
2. Wheel speed exceeds the threshold calculated from the vehicle reference speed and brake pressure is held at a constant value.
3. Wheel deceleration falls below a threshold (a) and brake pressure is reduced.
4. Brake pressure holding is now occurring and wheel speed will increase.
5. Wheel acceleration exceeds the upper limit (A) so brake pressure is now allowed to increase.
6. Pressure is again held constant as the limit (a) is exceeded.
7. Brake pressure is now increased in stages until wheel speed threshold (a) is exceeded.
8. Brake pressure is decreased again and then held constant when (a) is reached.

The process as above continues until the brake pedal is released or the vehicle speed is less than a set minimum, at which time the wheels will lock to bring the vehicle finally to rest.

Low adhesion

1. Initial braking, ABS not yet activated.
2. Wheel speed exceeds the threshold calculated from the vehicle reference speed and brake pressure is held at a constant value.
3. During this phase a short holding time is followed by a reduction in brake pressure. The wheel speed is compared with, and found to be less than, the calculated slip threshold so pressure is reduced again followed by a second holding time. A second comparison takes place and the pressure is reduced again.

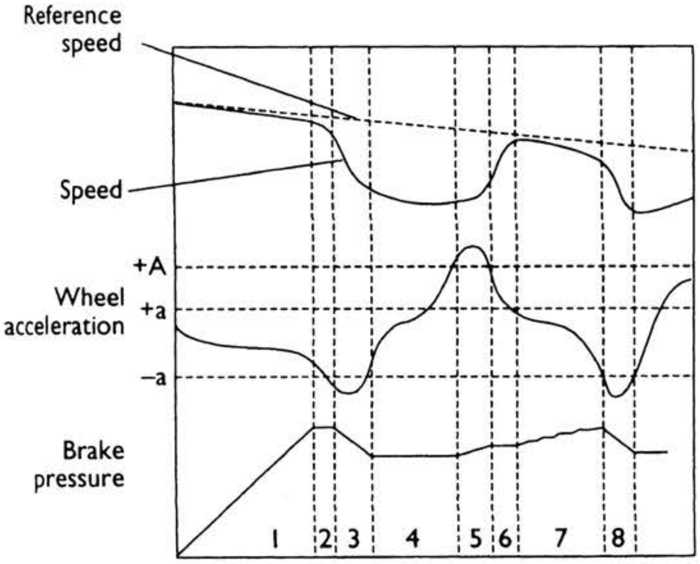


Figure 15.52 Braking control cycle for a high-adhesion surface

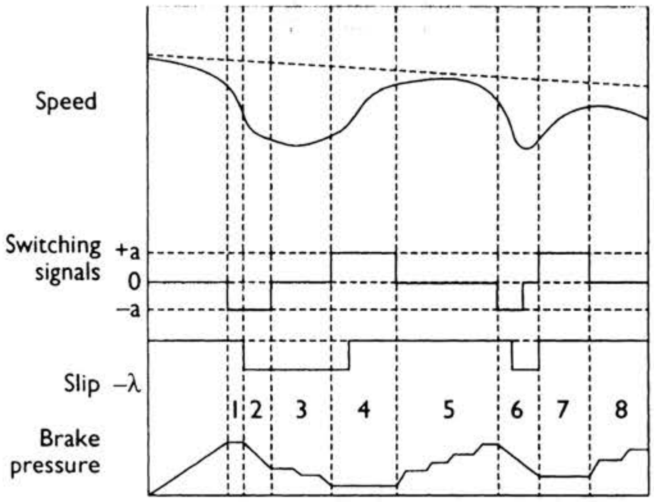


Figure 15.53 Braking control cycles for a low-adhesion surface

4. A brake pressure holding phase allows the wheel speed to increase.
5. There is a gradual introduction of increased brake pressure and holding pressure in steps until the wheel again slips.
6. Brake pressure is decreased allowing wheel speed to increase.
7. Pressure holding as the calculated slip value is reached.
8. Stepped increase in pressure with holding phases to keep high slip periods to a minimum. This ensures maximum stability.

The process again continues until the brake pedal is released or the vehicle comes to rest.

15.6.3 Traction control calculations

Figure 15.54 shows the forces acting on the wheels of a vehicle when accelerating on a non-homogeneous road surface. The maximum propulsion force can be calculated:

$$F = FH + FL = 2FL + FB$$

where: F = total propulsion force, FH = force transmitted to μH part of road, FL = force transmitted to μL part of road, FB = braking force, μH = high braking force coefficient, μL = low braking force coefficient.

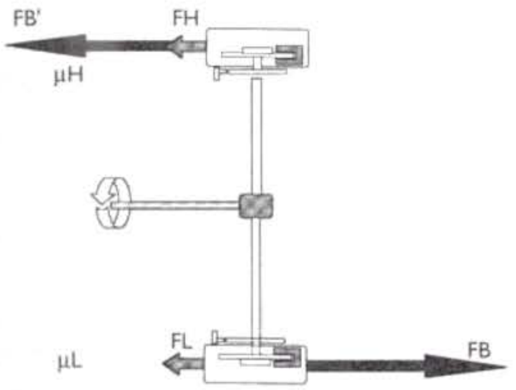


Figure 15.54 Forces acting on wheels of a vehicle when accelerating on a non-homogenous road surface

CHAPTER 16

Comfort and safety

16.1 Seats, mirrors and sun-roofs

16.1.1 Introduction

Electrical movement of seats, mirrors and the sun-roof is included in one chapter as the operation of each system is quite similar. The operation of electric windows and central door locking is also much the same.

Fundamentally, all the above-mentioned systems operate using one or several permanent magnet motors, together with a supply reversing circuit. A typical motor reverse circuit is shown in Figure 16.1. When the switch is moved, one of the relays will operate and this changes the polarity of the supply to one side of the motor. If the switch is moved the other way, then the polarity of the other side of the motor is changed. When at rest, both sides of the motor are at the same potential. This has the effect of regenerative braking so that when the motor stops it will do so instantly.

Further refinements are used to enhance the operation of these systems. Limit switches, position memories and force limitations are the most common.

16.1.2 Electric seat adjustment

Adjustment of the seat is achieved by using a number of motors to allow positioning of different parts of the seat. Movement is possible in the following ways.

- Front to rear.
- Cushion height rear.

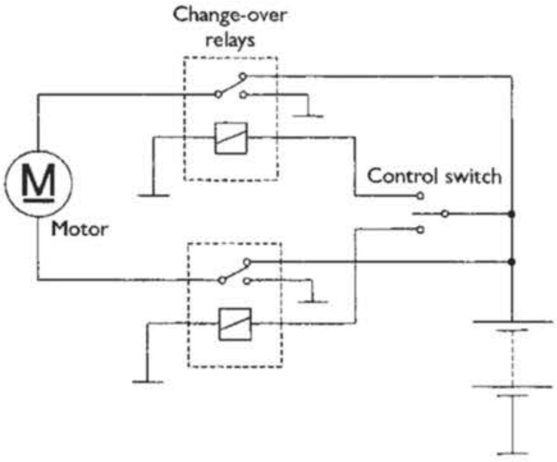


Figure 16.1 Typical motor reverse circuit

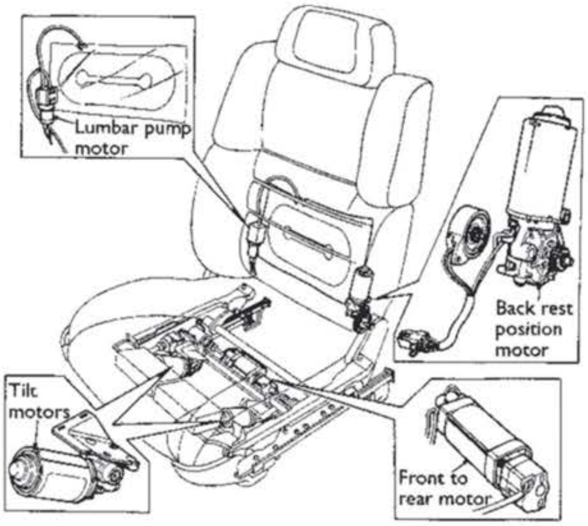


Figure 16.2 Electrically controlled seat

- Cushion height front.
- Backrest tilt.
- Headrest height.
- Lumber support.

Figure 16.2 shows a typical electrically controlled seat. This system uses four positioning motors and one smaller motor to operate a pump, which controls the lumber support bag. Each motor can be considered to operate by a simple rocker-type switch that controls two relays as described above. Nine relays are required for this, two for each motor and one to control the main supply.

When the seat position is set, some vehicles have set position memories to allow automatic re-positioning if the seat has been moved. This is often combined with electric mirror adjustment. Figure 16.3 shows how the circuit is constructed to allow position memory. As the seat is moved a variable resistor, mechanically linked to the motor, is also moved. The resistance value provides feedback to

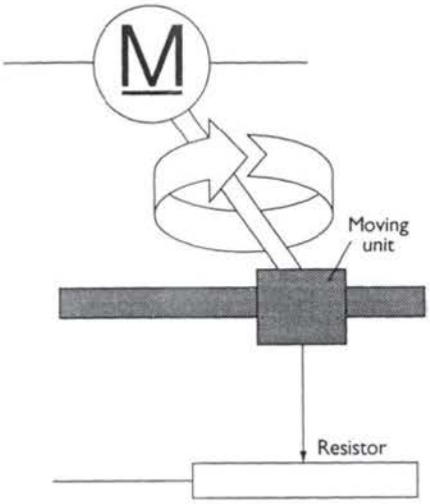


Figure 16.3 Position memory for electric seats

Figure 16.4 Seat motors

an electronic control unit. This can be 'remembered' in a number of ways; the best technique is to supply the resistor with a fixed voltage such that the output relative to the seat position is proportional to position. This voltage can then be 'analog-to-digital' converted, which produces a simple 'number' to store in a digital memory. When the driver presses a memory recall switch, the motor relays are activated by the ECU until the number in memory and the number fed back from the seat are equal. This facility is often isolated when the engine is running to prevent the seat moving into a dangerous position as the car is being driven. The position of the seats can still be adjusted by operating the switches as normal.

16.1.3 Electric mirrors

Many vehicles have electric adjustment of mirrors, particularly on the passenger side. The system used is much the same as has been discussed above in relation to seat movement. Two small motors are used to move the mirror vertically or horizontally. Many mirrors also contain a small heating element on the rear of the glass. This is operated for a few minutes when the ignition is first switched on and can also be linked to the heated rear window circuit. Figure 16.5 shows an electrically operated mirror circuit, which includes feedback resistors for positional memory.

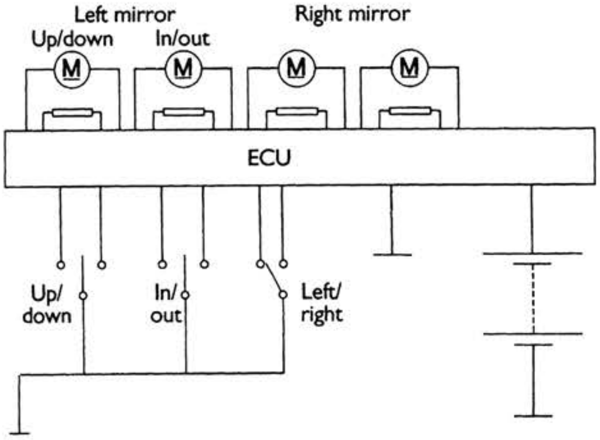


Figure 16.5 Feedback resistors for positional memory and the circuit

Figure 16.6 Mirror movement mechanism

16.1.4 Electric sun-roof operation

Key fact

A latching relay works in much the same way as a normal relay except that it locks into position each time it is energized.

Key fact

A micro switch is mechanically positioned such as to operate when the roof is in its closed position.

The operation of an electric sun-roof is similar to the motor reverse circuit discussed earlier in this chapter. However, further components and circuitry are needed to allow the roof to slide, tilt and stop in the closed position. The extra components used are a micro switch and a latching relay. A latching relay works in much the same way as a normal relay except that it locks into position each time it is energized. The mechanism used to achieve this is much like that used in ball-point pens that use a button on top.

The micro switch is mechanically positioned such as to operate when the roof is in its closed position. A rocker switch allows the driver to adjust the roof. The circuit for an electrically operated sun-roof is shown in Figure 16.8. The switch provides the supply to the motor to run it in the chosen direction. The roof will be caused to open or tilt. When the switch is operated to close the roof, the motor is run in the appropriate direction until the micro switch closes when the roof is in its closed position. This causes the latching relay to change over, which stops the motor. The control switch has now to be released. If the switch is pressed again, the latching relay will once more change over and the motor will be allowed to run.

Figure 16.7 Sun-roof

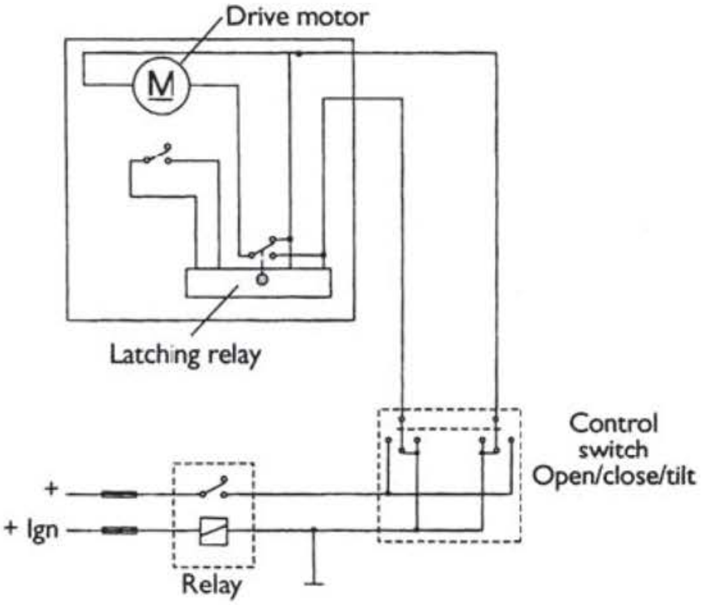


Figure 16.8 Sun-roof circuit

16.1.5 Seat control circuit

The circuit diagram shown as Figure 16.9 is an example from a modern Ford car. The front seats have a side air bag and height adjustable head restraint.

The seats in this vehicle are equipped with the following:

- power seat height adjustment;
- forward and rearward track movement;
- backrest recline;
- adjustable lumbar mat;
- heated and ventilated cushion and backrest.

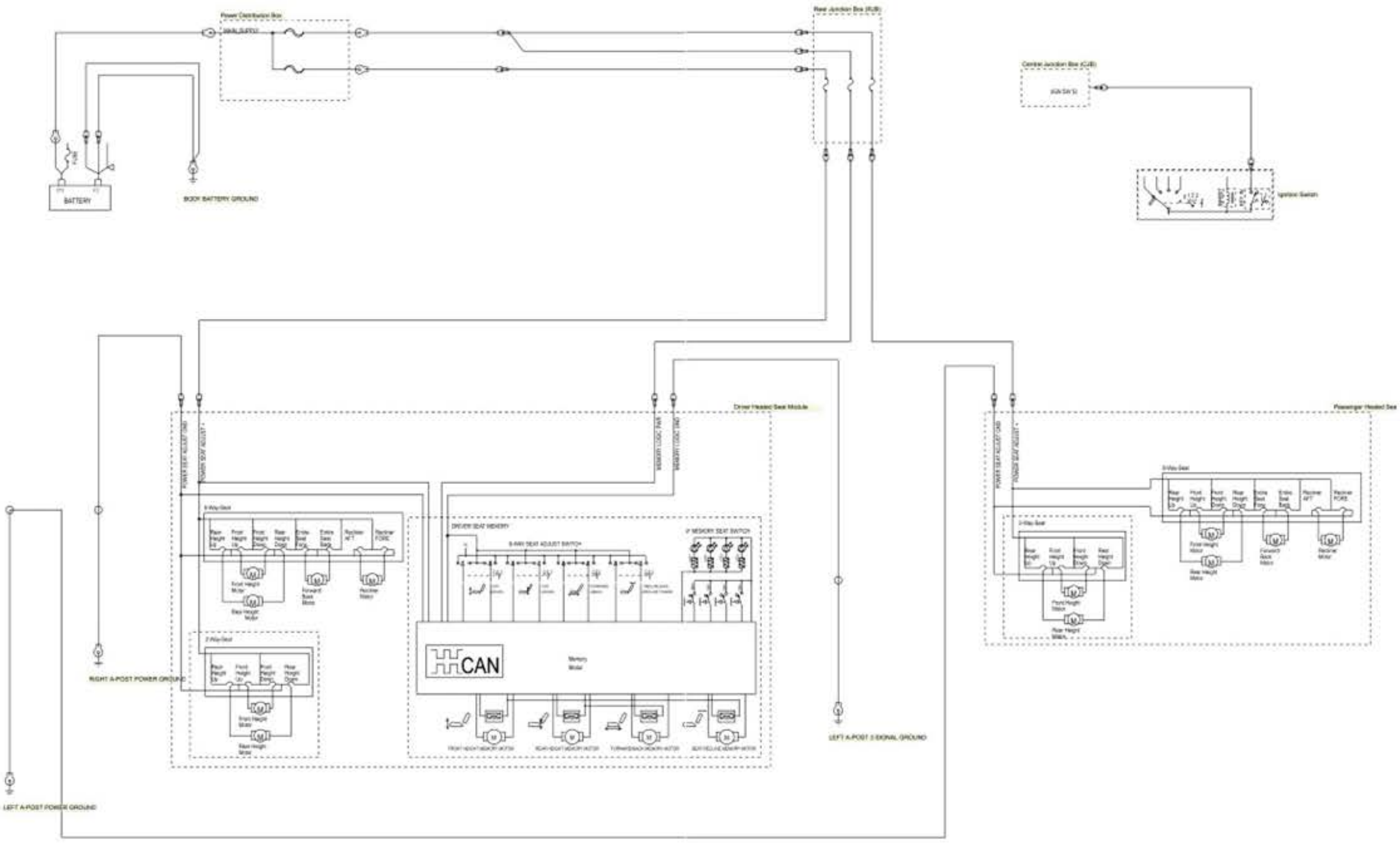


Figure 16.9 Seat control system circuit diagram (Source: Ford Motor Company)

16.2 Central locking and electric windows

16.2.1 Door locking circuit

When the key is turned in the driver's door lock, all the other doors on the vehicle should also lock. Motors or solenoids in each door achieve this. If the system can only be operated from the driver's door key, then an actuator is not required in this door. If the system can be operated from either front door or by remote control, then all the doors need an actuator. Vehicles with sophisticated alarm systems often lock all the doors as the alarm is set.

Figure 16.10 shows a basic door locking circuit. The main control unit contains two change-over relays (as in Figure 16.1), which are actuated by either the door

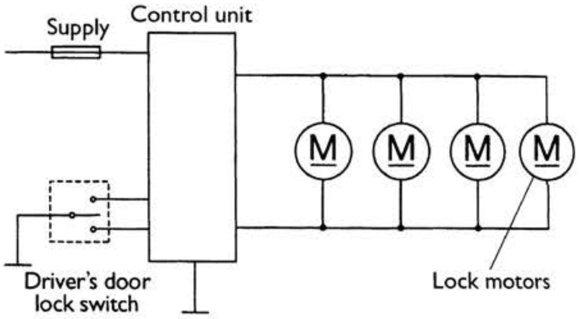


Figure 16.10 Door lock circuit

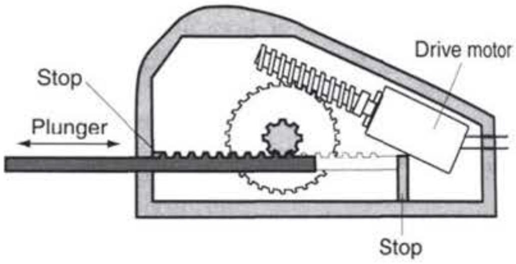


Figure 16.11 Door lock actuator

Key fact

Most door actuators are small motors.

Key fact

Remote central door locking is controlled by a small hand-held transmitter and a receiver unit as well as a decoder in the main control unit.

Definition

MAC: Message authentication code.

lock switch or the remote key. The motors for each door lock are simply wired in parallel and all operate at the same time.

Most door actuators are small motors which, via suitable gear reduction, operate a linear rod in either direction to lock or unlock the doors. A simple motor reverse circuit is used to achieve the required action. Figure 16.11 shows a typical door lock actuator.

Remote central door locking is controlled by a small hand-held transmitter and a receiver unit as well as a decoder in the main control unit. This layout will vary slightly between different manufacturers. When the remote key is operated by pressing a small switch, a complex code is transmitted. The number of codes used is well in excess of 50 000. The sensor picks up this code and sends it in an electrical form to the main control unit. If the received code is correct, the relays are triggered and the door locks are either locked or unlocked. If an incorrect code is received on three consecutive occasions when attempting to unlock the doors, then the infrared system will switch itself off until the door is opened by the key. This will also reset the system and allow the correct code to operate the locks again. This technique prevents a scanning type transmitter unit from being used to open the doors. Figure 16.12 shows a flow diagram representing the operation of a system that uses a ‘rolling code’ (MAC stands for Message Authentication Code).

16.2.2 Electric window operation

The basic form of electric window operation is similar to many of the systems discussed so far in this chapter; that is, a motor reversing system that is operated either by relays or directly by a switch.

More sophisticated systems are now becoming more popular for reasons of safety as well as improved comfort. The following features are now available from many manufacturers:

- One shot up or down.
- Inch up or down.
- Lazy lock.
- Back-off.

The complete system consists of an electronic control unit containing the window motor relays, switch packs and a link to the door lock and sun-roof circuits. This is represented in the form of a block diagram in Figure 16.13.

When a window is operated in one-shot or one-touch mode the window is driven in the chosen direction until either the switch position is reversed, the motor stalls or the ECU receives a signal from the door lock circuit. The problem with one-shot operation is that if a child, for example, should become trapped in the

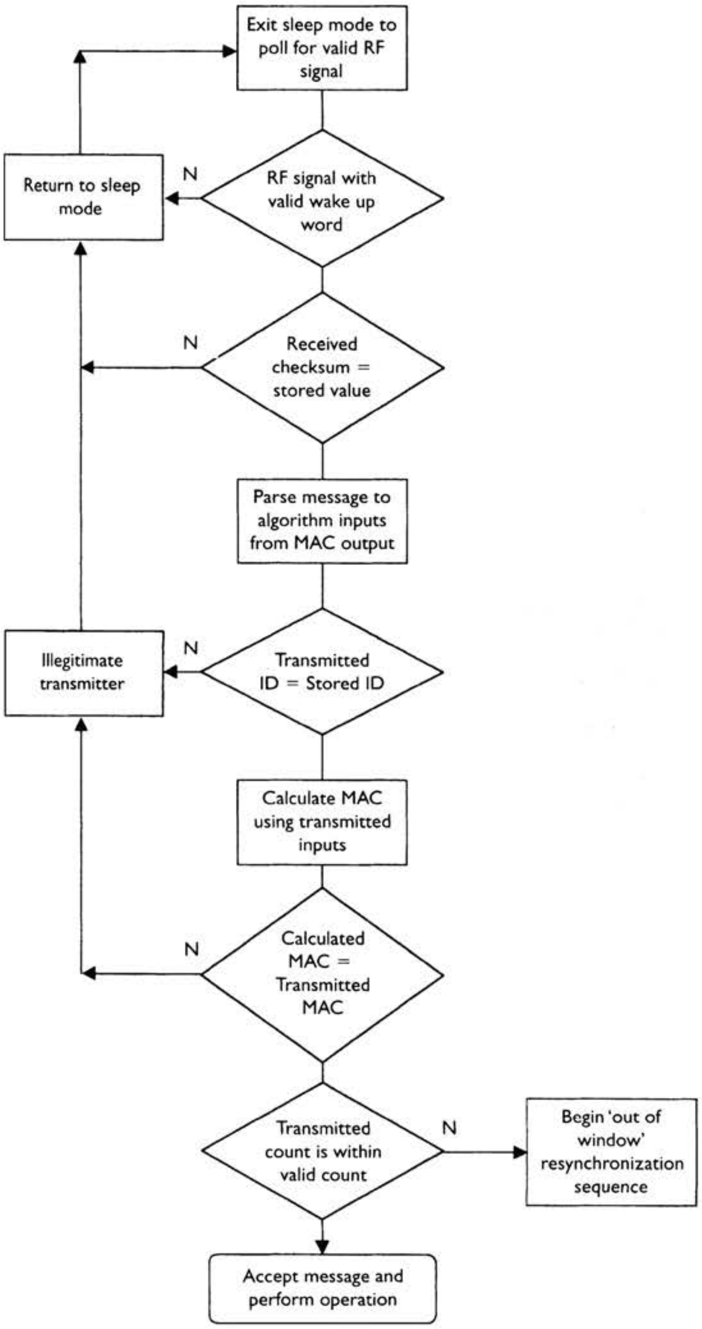


Figure 16.12 Flow diagram representing the 'Rolling Code' system

window there is a serious risk of injury. To prevent this, the back-off feature is used. An extra commutator is fitted to the motor armature and produces a signal via two brushes, proportional to the motor speed. If the rate of change of speed of the motor is detected as being below a certain threshold when closing, then the ECU will reverse the motor until the window is fully open.

By counting the number of pulses received, the ECU can also determine the window position. This is important, as the window must not reverse when it stalls in the closed position. In order for the ECU to know the window position it must be initialized. This is often done simply by operating the motor to drive the window first fully open, and then fully closed. If this is not done then the one-shot close will not operate.

Key fact

If the rate of change of speed of the motor is detected as being below a certain threshold when closing, then the ECU will reverse the motor until the window is fully open.

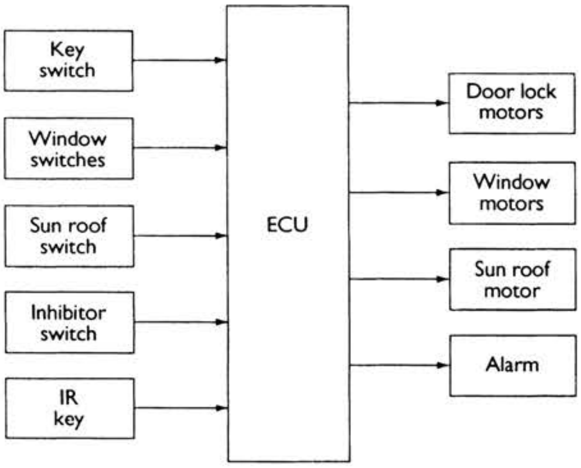


Figure 16.13 Block diagram showing links between door locks, windows and sun-roof – controlled by an infrared key

Key fact

On some systems, Hall Effect sensors are used to detect motor speed.

On some systems, Hall Effect sensors are used to detect motor speed. Other systems sense the current being drawn by the motor and use this as an indication of speed.

The lazy lock feature allows the car to be fully secured by one operation of a remote key. This is done by the link between the door lock ECU and the window and sun-roof ECUs. A signal is supplied and causes all the windows to close in turn, then the sun-roof, and finally it locks the doors. The alarm will also be set if required. The windows may close in turn to prevent the excessive current demand that could occur if they all operate at the same time.

A circuit for electric windows is shown in Figure 16.14. Note the connections to other systems such as door locking and the rear window isolation switch. This is commonly fitted to allow the driver to prevent rear window operation for child safety, for example.

Figure 16.15 shows a typical window lift motor used for cable or arm-lift systems.

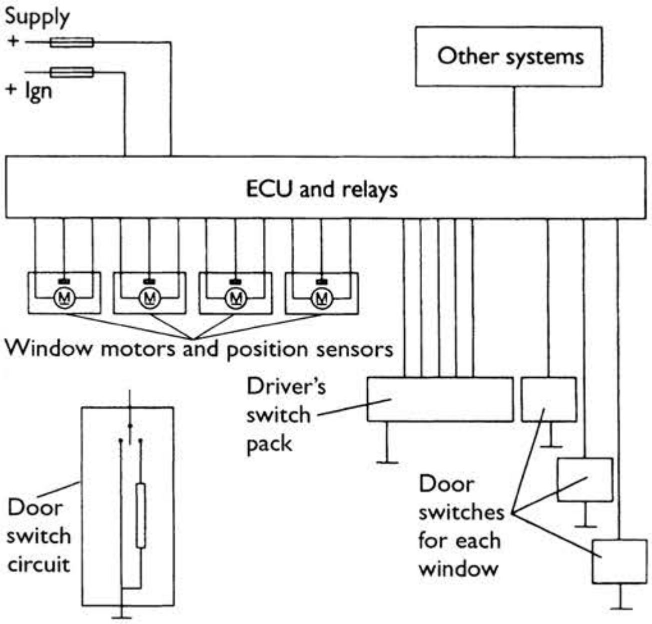


Figure 16.14 Electric window control circuit

Figure 16.15 Window lift motor cable or arm-lift systems

16.2.3 Electric windows example circuit

The circuit of the electric window system used by some vehicles is shown in Figure 16.16. The windows will only operate when the ignition is switched on. When the ignition is switched on, the window lift relay is energized by the supply from fuse 18 in the passenger compartment fuse-box on the LG wire, which passes to earth on a B wire. With the relay energized, the battery supply from fusible link 4 on the N wire feeds the four window lift fuses on an N/U wire.

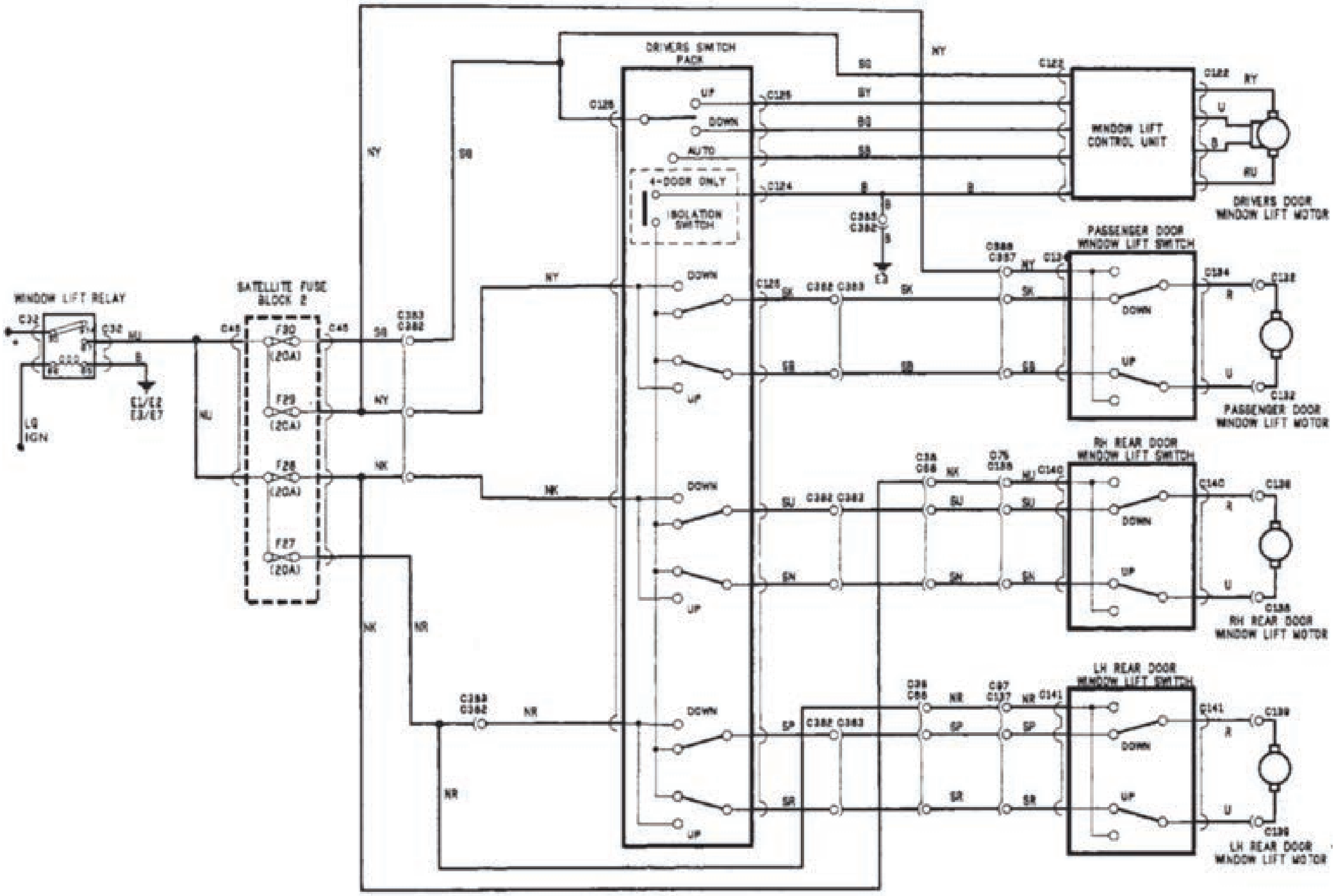


Figure 16.16 Electric window system circuit

The driver's window can only be operated from a switch block on the driver's door, which is supplied from fuse 30 in satellite fuse block 2, on an S/G wire. When the 'up' switch is pressed, the feed from the fuse crosses the window lift switch and provides a feed to the control unit on a B/Y wire. The control unit will now provide a positive supply to the window lift motor on a R/U wire and an earth path on an R/Y wire. The window will now move upwards until the switch is released or it reaches the end of its travel.

When the 'down' switch is pressed the supply from fuse 30 in satellite fuse block 2 provides a feed to the control unit on an S/G wire. The control unit will now connect a positive feed to the window lift motor on an R/Y wire and an earth path on a R/U wire. The window will now move downwards until the switch is released or the window reaches the end of its travel.

The driver's door window may be fully opened by moving the driver's door window switch fully downwards then releasing it. This will allow a supply to cross the closed switch contacts and feed the control unit on an S/B wire. The control unit will now operate the window lift motor in the downward direction until the window reaches the end of its travel. The front passenger's window can be operated from the driver's door switchback or the passenger's door switchback.

When the 'up' switch is pressed, the supply from fuse 29 in satellite fuse block 2 on the N/Y wire crosses the window lift switch out to the passenger's window lift switch on an S/B wire, then onto the window lift motor on a U wire. The earth path for the window lift motor on an R wire crosses the passenger's window lift switch out to the driver's door master switch on an S/K wire, through the isolator switch and to earth on a B wire.

When the 'down' switch is pressed, the supply from fuse 29 in satellite fuse block 2 on an N/Y wire crosses the window lift switch out to the passenger's window lift switch on an S/K wire, then onto the window lift motor on an R wire. The earth path for the window lift motor on a U wire crosses the passenger's window lift switch out to the driver's door master switch on an S/B wire, through the isolator switch and to earth on a B wire.

When the 'up' switch is pressed, the supply from fuse 29 in satellite fuse block 2 supplies the passenger's window lift switch on an N/Y wire, then onto the window lift motor on a U wire. The earth path for the window lift motor on an R wire crosses the passenger's window lift switch out to the driver's door master switch on an S/K wire, through the isolator switch and to earth on a B wire.

When the 'down' switch is pressed, the supply from fuse 29 in satellite fuse block 2 supplies the passenger's window lift switch on an N/Y wire, then onto the window lift motor on an R wire. The earth path for the window lift motor on a U wire crosses the passenger's window lift switch out to the driver's door master switch on an S/B wire, through the isolator switch and to earth on a B wire.

Each rear window can be operated from the driver's door switchback or, provided that the isolation switch in the driver's door switchback has not been pressed, from the switch on each rear door. The operation of the rear windows is similar in operation to the front passenger's window.

Key fact

Cruise control is the ideal example of a closed loop control system.

16.3 Cruise control

16.3.1 Introduction

Cruise control is the ideal example of a closed loop control system. Figure 16.17 illustrates this in the form of a block diagram. The purpose of cruise control

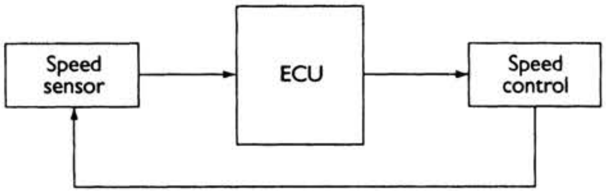


Figure 16.17 Cruise control closed loop system

is to allow the driver to set the vehicle speed and let the system maintain it automatically.

The system reacts to the measured speed of the vehicle and adjusts the throttle accordingly. The reaction time is important so that the vehicle's speed does not feel to be surging up and down.

Other facilities are included such as allowing the speed to be gradually increased or decreased at the touch of a button. Most systems also remember the last set speed and will resume this again at the touch of a button.

To summarize and to add further refinements, the following is the list of functional requirements for a good cruise control system.

- Hold the vehicle speed at the selected value.
- Hold the speed with minimum surging.
- Allow the vehicle to change speed.
- Relinquish control immediately the brakes are applied.
- Store the last set speed.
- Contain built in safety features.

16.3.2 System description

The main switch turns on the cruise control, this in turn is ignition controlled. Most systems do not retain the speed setting in memory when the main switch has been turned off. Operating the 'set' switch programs the memory but this normally will only work if conditions similar to the following are met.

- Vehicle speed is greater than 40 km/h.
- Vehicle speed is less than 12 km/h.
- Change of speed is less than 8 km/h/s.
- Automatics must be in 'drive'.
- Brakes or clutch are not being operated.
- Engine speed is stable.

Once the system is set, the speed is maintained to within about 2–3 km/h until it is deactivated by pressing the brake or clutch pedal, pressing the 'resume' switch or turning off the main control switch. The last 'set' speed is retained in memory except when the main switch is turned off.

If the cruise control system is required again then either the 'set' button will hold the vehicle at its current speed or the 'resume' button will accelerate the vehicle to the previous 'set' speed. When cruising at a set speed, the driver can press and hold the 'set' button to accelerate the vehicle until the desired speed is reached when the button is released.

If the driver accelerates from the set speed to overtake, for example, then when the throttle is released, the vehicle will slow down until it reaches the last set position.

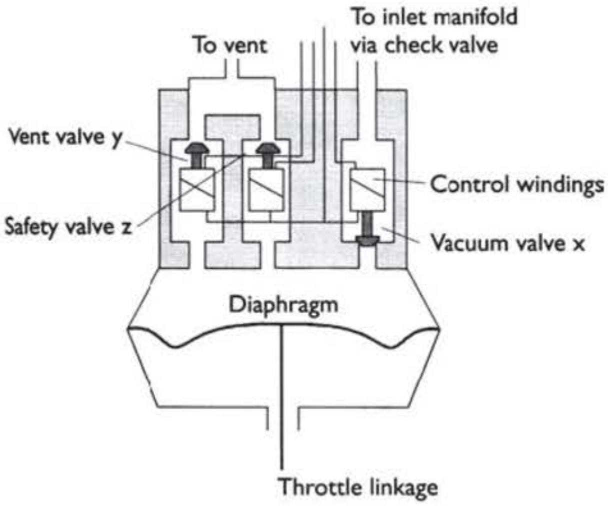


Figure 16.18 Cruise control 'vacuum' actuator

16.3.3 Components

The main components of a typical cruise control system are as follows.

Actuator

Figure 16.19 Motor-driven actuator

A number of methods are used to control the throttle position. Vehicles fitted with driven by-wire systems allow the cruise control to operate the same actuator. A motor (Figure 16.19) can be used to control the throttle cable or, in many cases, a vacuum-operated diaphragm is used which is controlled by three simple valves. This technique is shown in Figure 16.18. When the speed needs to be increased, valve 'x' is opened allowing low pressure from the inlet manifold to one side of the diaphragm. The atmospheric pressure on the other side will move the diaphragm and hence the throttle. To move the other way, valve 'x' is closed and valve 'y' is opened allowing atmospheric pressure to enter the chamber. The spring moves the diaphragm back. If both valves are closed then the throttle position is held. Valve 'x' is normally closed and valve 'y' normally open; thus, in the event of electrical failure cruise control will not remain engaged and the manifold vacuum is not disturbed. Valve 'z' provides extra safety and is controlled by the brake and clutch pedals.

Main switch and warning lamp

This is a simple on/off switch located within easy reach of the driver on the dashboard. The warning lamp can be part of this switch or part of the main instrument display as long as it is in the driver's field of vision.

Set and resume switches

These are fitted either on the steering wheel or on a stalk from the steering column. When the switches are part of the steering wheel, slip rings are needed to transfer the connection. The 'set' button programs the speed into memory and can also be used to increase the vehicle and memory speed. The 'resume' button allows the vehicle to reach its last set speed or temporarily to deactivate the control.

Brake switch

This switch is very important, as it would be dangerous braking if the cruise control system was trying to maintain the vehicle speed. This switch is normally of superior quality and is fitted in place or as a supplement to the brake light switch activated by the brake pedal. Adjustment of this switch is important.

Clutch or automatic gearbox switch

The clutch switch is fitted in a similar manner to the brake switch. It deactivates the cruise system to prevent the engine speed increasing if the clutch is pressed. The automatic gearbox switch will only allow the cruise to be engaged when it is in the ‘drive’ position. This is again to prevent the engine over-speeding if the cruise control tried to accelerate to a high road speed with the gear selector in the ‘1’ or ‘2’ position. The gearbox will still change gear if accelerating back up to a set speed as long as it ‘knows’ top gear is available.

Speed sensor

This will often be the same sensor that is used for the speedometer. If not, several types are available – the most common produces a pulsed signal, the frequency of which is proportional to the vehicle speed.

16.3.4 Adaptive cruise control

Conventional cruise control has now developed to a high degree of quality. It is, however, not always very practical on many European roads as the speed of the general traffic varies constantly and traffic is often very heavy. The driver has to take over from the cruise control system on many occasions to speed up or slow down. Adaptive cruise control can automatically adjust the vehicle speed to the current traffic situation. The system has three main aims.

- Maintain a speed as set by the driver.
- Adapt this speed and maintain a safe distance from the vehicles in front.
- Provide a warning if there is a risk of collision.

The main extra components, compared to standard cruise control, are the ‘headway’ sensor and the steering angle sensor; the first of these is clearly

Key fact

Adaptive cruise control can automatically adjust the vehicle speed to the current traffic situation.

Figure 16.20 Adaptive cruise control operation

Figure 16.21 Adaptive cruise control

the most important. Information on steering angle is used to enhance further the data from the headway sensor by allowing greater discrimination between hazards and spurious signals. Two types of the headway sensor are in use, the radar and the lidar. Both contain transmitter and receiver units. The radar system uses microwave signals at about 35 GHz, and the reflection time of these gives the distance to the object in front. Lidar uses a laser diode to produce infrared light signals, the reflections of which are detected by a photodiode. Figure 16.21 shows a lidar sensor.

These two types of sensors have advantages and disadvantages. The radar system is not affected by rain and fog but the lidar can be more selective by recognizing the standard reflectors on the rear of the vehicle in front. Radar can produce strong reflections from bridges, trees, posts and other normal roadside items. It can also suffer loss of signal return due to multipath reflections. Under ideal weather conditions, the lidar system appears to be the best but it becomes very unreliable when the weather changes. A beam divergence of about 2.5° vertically and horizontally has been found to be the most suitable whatever headway sensor is used. An important consideration is that signals from other vehicles fitted with this system must not produce erroneous results. Figure 16.22 shows a typical headway sensor and control electronics.

Figure 16.22 Headway sensor is fitted at the front of a vehicle (Source: Bosch Media)

Fundamentally, the operation of an adaptive cruise system is the same as a conventional system except when a signal from the headway sensor detects an obstruction, in which case the vehicle speed is decreased. If the optimum stopping distance cannot be achieved by just backing off the throttle, a warning is supplied to the driver. A more complex system can also take control of the vehicle transmission and brakes but this, while very promising, is further behind in development. It is important to note that adaptive cruise control is designed to relieve the burden on the driver, not take full control of the vehicle!

Key fact

It is important to note that adaptive cruise control is designed to relieve the burden on the driver, not take full control of the vehicle!

16.4 In-car multimedia

16.4.1 Introduction

These days it would be almost unthinkable not to have at least a radio cassette/CD player in our vehicles. It was not long ago, however, that these were an optional extra. Looking back just a little further, the in-car record player must have been interesting to operate – it was evidently quite successful in large American cars in the US but left a bit to be desired in British vehicles and on British roads.

Figure 16.23 ICE multimedia

We now have ICE systems fitted to standard production cars, which are of good hi-fi quality. Facilities such as compact disc players and multiple compact disc changers together with automatic station search and re-tune are popular.

We have seen the rise and fall of the CB radio and the first car telephones – which were so large the main unit had to be fitted in the car boot. 'Hands-free' car telephones, which allow both hands to be kept free to control the car, are in common use and voice activation of other systems is developing.

16.4.2 Speakers

Good ICE systems usually include at least six speakers, two larger speakers in the rear parcel shelf to produce good low-frequency reproduction, two front door speakers for the mid-range and two front door tweeters for high-frequency notes. A separate sub-woofer may also be used. Figure 16.24 shows a Pioneer sub-woofer speaker, which may be mounted in a cabinet in the boot for example, as very low frequencies are not directional (i.e. the stereo effect is not important).

Key fact

Good ICE systems usually include at least six speakers.

Figure 16.24 Speakers

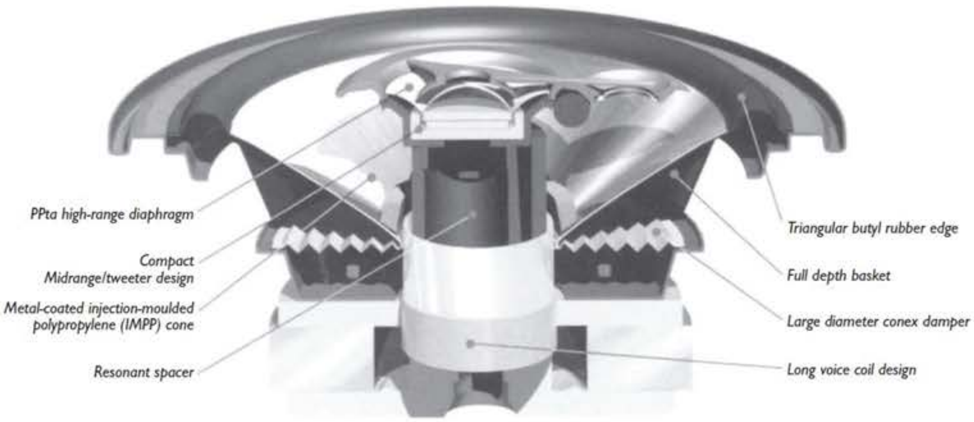


Figure 16.25 Speaker construction

Speakers are a very important part of a sound system. No matter how good the receiver, MP3 player or CD player is, the sound quality will be reduced if inferior speakers are used. Equally, if the speakers are of a lower power output rating than the set, distortion will result at best, and damage to the speakers at worst. Speakers generally fall into the following four categories.

- Tweeters – high-frequency reproduction.
- Mid-range – middle-range frequency reproduction (treble).
- Woofers – low-frequency reproduction (bass).
- Sub-woofers – very-low-frequency reproduction.

Figure 16.25 shows the construction of a speaker.

Key fact

Most modern sets now also have an auxiliary input for an MP3 player, which can be a simple 3.5 mm stereo jack (like headphones) or in some cases a USB connection is possible.

16.4.3 In-car entertainment (ICE)

Controls on most ICE sets will include volume, treble, bass, balance and fade. On old systems, cassette tape options will include Dolby filters to reduce hiss and other tape selections such as chrome or metal. A digital display will provide a visual output of the operating condition. This is also linked into the vehicle lighting to prevent glare at night. Track selection and programming for one or several compact discs is possible. Most modern sets now also have an auxiliary input for an MP3 player, which can be a simple 3.5 mm stereo jack (like headphones) or in some cases a USB connection is possible.

Many ICE systems are coded to deter theft. The code is activated if the main supply is disconnected and will not allow the set to work until the correct code has been re-entered. Some systems included a plug-in electronic ‘key card’, which makes the set worthless when removed.

16.4.4 Radio data system (RDS)

RDS has become a standard on many radio sets. It is an extra inaudible digital signal, which is sent with FM broadcasts in a similar way to how teletext is sent with TV signals. RDS provides information so a receiver can appear to act intelligently. The possibilities available when RDS is used are as follows.

- The station name can be displayed in place of the frequency.
- Automatic tuning is possible to the best available signal for the chosen radio station. For example, in the UK, a journey from the south of England to

Figure 16.26 ICE unit for a 24 V truck (Source: Bosch Media)

Scotland would mean the radio would have to be re-tuned up to ten times. RDS will do this without the driver even knowing.

- Traffic information broadcasts can be identified and a setting made so that whatever you are listening to at the time can be interrupted.

Key fact

RDS is an extra inaudible digital signal, which is sent with FM broadcasts.

RDS has six main features, which are listed here with a brief explanation.

1. Programme identification to allow the re-tune facility to follow the correct broadcasts.
2. Alternative frequencies, again to allow the receiver to try other signals for re-tuning as required.
3. Programme service name for displaying the name of the station on the radio set.
4. Traffic information, which provides for two codes to work in conjunction with route finding equipment.
5. Traffic programme, which allows the set to indicate that the station broadcasts traffic information.
6. A traffic announcement is transmitted when an announcement is being broadcast. This allows the receiver either to adjust the volume, switch over from the cassette during the announcement, lift an audio mute or, of course, if the driver wishes it, to do nothing.

16.4.5 Radio broadcast data system (RBDS)

The Radio Broadcast Data System is an extension of the Radio Data System (RDS), which has been in use in Europe since 1984. The system allows the broadcaster to transmit text information at the rate of about 1200 bits per second. The information is transmitted on a 57 kHz suppressed sub-carrier as part of the FM multiplexed (MPX) signal.

RBDS was developed for the North American market by the National Radio Systems Committee (NRSC), a joint committee composed of the Electronic

Industries Association (EIA) and the National Association of Broadcasters (NAB). The applications for the transmission of text to the vehicle are interesting.

- Song title and artist.
- Traffic, accident and road hazard information.
- Stock information.
- Weather.

In emergency situations, the audio system can be enabled to interrupt the CD or normal radio broadcast to alert the user.

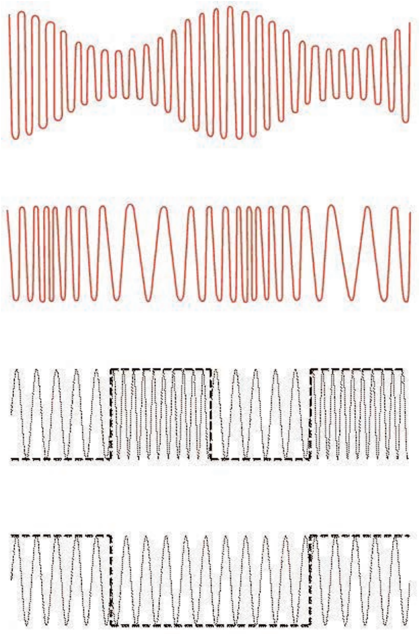


Figure 16.27 Signal modulation: Top to bottom; AM analog, FM analog, FM digital, Phase shift digital

16.4.6 Radio reception

There are three main types of radio signal transmitted; these are amplitude modulation (AM), frequency modulation (FM) and digitally modulated for digital audio broadcast (DAB). Figure 16.27 shows the difference between these signals.

Amplitude modulation is a technique for varying the height, or amplitude, of a wave in order to transmit information. Some radio broadcasts still use amplitude modulation. A convenient and efficient means of transmitting information is by the propagation of waves of electromagnetic radiation. Sound waves in the audible range, such as speech and music, have a frequency that is too low for efficient transmission through the air for significant distances. By the process of modulation, however, this low-frequency audio information can be impressed on a carrier wave that has a much higher frequency and can propagate through space for great distances. The transmitter at a radio station generates a carrier wave having constant characteristics, such as amplitude and frequency. The signal containing the desired information is then used to modulate the carrier.

This new wave, called the modulated wave, will contain the information of the signal. In AM, it is the amplitude of the carrier wave that is made to vary so that it will contain the information of the signal. When the modulated wave reaches a radio receiver tuned to the proper frequency, it is demodulated, which is essentially the opposite of modulation. The set can then reproduce the desired sound via an amplifier and the loudspeakers. AM radio is still a popular form of radio broadcasting, but it does have a number of disadvantages. The quality of reproduction is relatively poor because of inherent limitations in the technique and because of interference from other stations and other electrical signals, such as those produced by lightning or by electronic devices – of which the car has more than its fair share. Some of these drawbacks can be overcome by using FM.

Key fact

Frequency modulation is a method of modulation in which the frequency of a wave is varied in response to a modulating wave.

Frequency modulation is a method of modulation in which the frequency of a wave is varied in response to a modulating wave. The wave in which frequency is varied is called the carrier, and the modulating wave is called the signal. Frequency modulation requires a higher-frequency carrier wave and a more complex method for transmitting information than does AM; however, FM has an important advantage in that it has constant amplitude; it is therefore much less susceptible to interference from both natural and artificial sources. Such sources cause static in an amplitude-modulated radio.

Both types of modulation, however, are used in radio broadcasting. FM radio is generally a far better source of high fidelity music. This is because the quality of AM reception, as well as the problems outlined above, is limited by the narrow bandwidth of the signal. During the winter months, reception of AM signals becomes worse due to changes in the atmosphere. FM does, however, present problems with reception when mobile. As most vehicles use a rod aerial, which is omnidirectional, it will receive signals from all directions. Because of this, reflections from buildings, hills and other vehicles can reach the set all at the same time. This can distort the signal and is heard as a series of clicks or signal flutter as the signal is constantly enhanced or reduced. The best FM reception is considered to be line-of-sight from the transmitter. In general, the coverage or footprint of FM transmitters is quite extensive and, especially with the advent of RDS, the reception when mobile is quite acceptable.

In digital modulation, an analog carrier signal is modulated by a digital bit stream. Digital modulation methods can be considered as digital-to-analog conversion, and the corresponding demodulation or detection as analog-to-digital conversion

Key fact

In digital modulation, an analog carrier signal is modulated by a digital bit stream.

DAB is advertised as being interference-free and is supposed to be easier to receive than FM. Unfortunately the reality is that DAB's reception is far from perfect. Interference in the form of multipath can reduce the signal strength at the main carrier frequency and this result in too many bits of the digital signal being in error and a sound that is commonly called 'bubbling' or 'boiling mud'. This is quite a common problem with reception of DAB signals. If the multipath interference gets any worse than when you hear the bubbling sounds then the signal will drop out altogether and the audio mutes. This too is not uncommon, but tends not to be mentioned in the DAB advertisements!

A test has shown that for stationary listening that the audio quality on lower bit rate DAB is lower than FM stereo. However, an upgraded version of the system was released in February 2007, which is called DAB+. DAB is not forward compatible with DAB+, which means that DAB-only receivers will not be able to receive DAB+ broadcasts. DAB+ is approximately twice as efficient as DAB due to the adoption of a better audio codec, and DAB+ can provide high quality audio with as low as 64 kbit/s. Reception quality will also be more robust on DAB+ than on DAB due to the addition of error correction coding.

Definition

Codec: is short for encoder-decoder. Different video and audio formats have different compression ratios so codecs are necessary to ensure these files play on a computer or similar device.

16.4.7 Digital audio broadcast (DAB)

Digital Audio Broadcasting is designed to provide high-quality, multiservice digital radio broadcasting for reception by stationary and mobile receivers. It is being designed to operate at any frequency up to 3 GHz.

The system uses digital techniques to remove redundancy and perceptually irrelevant information from the audio source signal. It then applies closely controlled redundancy to the transmitted signal for error correction. All transmitted information is then spread in both the frequency and the time domains (multiplexed) so a high quality signal is obtained in the receiver, even under poor conditions.

The method of processing the DAB signal is shown in Figure 16.28. FIC is a fast information channel.

Frequency reallocation will permit broadcasters to extend services, using additional transmitters, all operating on the same radiated frequency. A common worldwide frequency in the L band (around 1.5 GHz) is being considered, but some disagreement still exists. The possibilities make the implementation of DAB inevitable.

Definition

FIC: F fast information chennel.

16.4.8 Interference suppression

The process of interference suppression on a vehicle is aimed at reducing the amount of unwanted noise produced from the speakers of an ICE system. This,

Definition

OFDM: Orthogonal Frequency Division Multiplexing - A modulation technology that separates the data stream into a number of lower-speed data streams, which are then transmitted in parallel. It is used for wireless (Wi-Fi) and DAB.

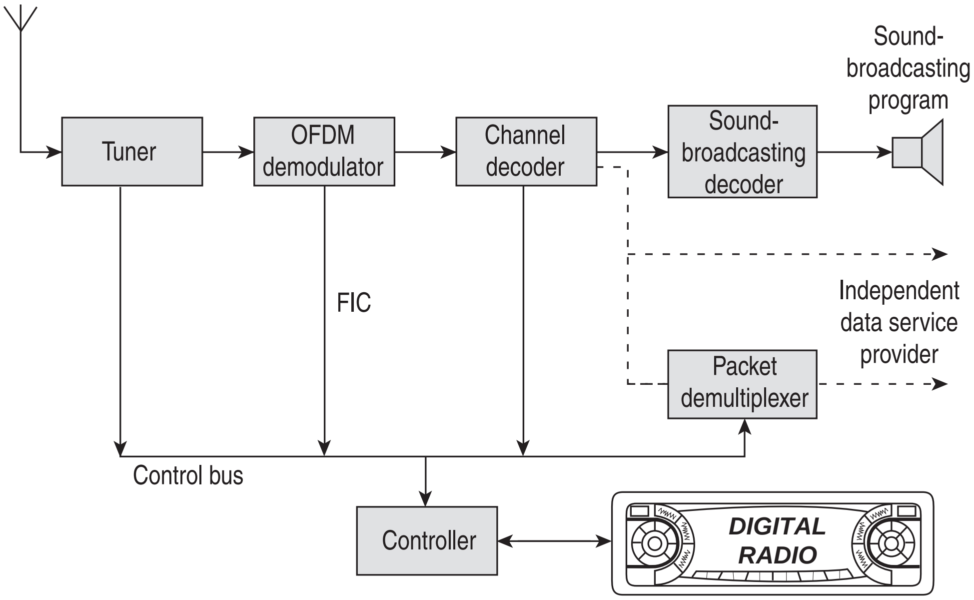


Figure 16.28 DAB receiver components

Key fact

The process of interference suppression on a vehicle is aimed at reducing the amount of unwanted noise produced from the speakers of an ICE system.

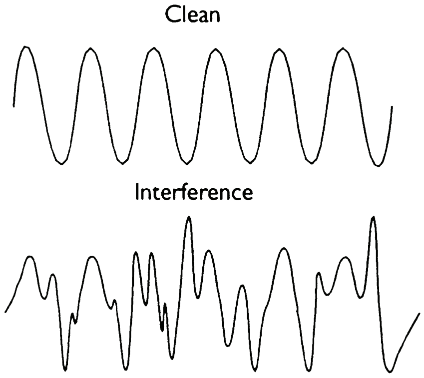


Figure 16.29 Two signals, one clean and the other suffering from interference

however, can be quite difficult. To aid the discussion, it is necessary first to understand the different types of interference. Figure 16.29 shows two signals, one clean and the other suffering from interference. The amount of interference can be stated as a signal-to-noise ratio. This is the useful field strength compared with the interference field strength at the receiver. This should be as high as possible but a value in excess of 22.1 for radio reception is accepted as a working figure. Interference is an electromagnetic compatibility (EMC) issue and further details can be found in Chapter 4.

There are two issues to be considered relating to suppression of interference on a vehicle:

1. Short range – the effect of interference on the vehicle's radio system.
2. Long range – the effect of the vehicle on external receivers such as domestic televisions. This is covered by legislation making it illegal to cause disturbance to radios or televisions when using a vehicle.

Interference can propagate in one of four ways.

- Line borne, conducted through the wires.
- Air borne, radiated through the air to the aerial.
- Capacitive coupling by an electric field.
- Inductive coupling magnetic linking.

The sources of interference in the motor vehicle can be summarized quite simply as any circuit, which is switched or interrupted suddenly. This includes the action of a switch and the commutation process in a motor, both of which produce rapidly increasing signals. The secret of suppression is to slow down this increase. Interference is produced from four main areas of the vehicle.

- Ignition system.
- Charging system.
- Motors and switches.
- Static discharges.

The ignition system of a vehicle is the largest source of interference, particularly the high tension side. Voltages up to 50 kV are now common and the peak current for a fraction of a second when the spark plug fires can peak in excess of 100 A. The interference caused by the ignition system is mostly above 30 MHz and the energy can peak, for fractions of a second, of the order of 500 kW.

The charging system produces noise because of the sparking at the brushes. Electronic regulators produce little problems but regulators with vibrating contacts can cause trouble.

Any motor or switch, including relays, is likely to produce some interference. The most popular sources are the wiper motor and heater motor. The starter is not considered due to its short usage time.

The build-up of static electricity is due to friction between the vehicle and the air, and the tyres and the road. If the static on, say, the bonnet builds up more than the wing then a spark can be discharged. Using bonding straps to ensure all panels stay at the same potential easily prevents this. Due to the action of the tyres, a potential can build up between the wheel rims and the chassis unless suitable bonding straps are fitted. The arc to ground can be as much as 10 kV.

There are five main techniques for suppressing radio interference.

- Resistors.
- Bonding.
- Screening.
- Capacitors.
- Inductors.

Resistance is used exclusively in the ignition HT circuit, up to a maximum of about 20 kΩ per lead. This has the effect of limiting the peak current, which in turn limits the peak electromagnetic radiation. Providing excessive resistance is not used, the spark quality is not affected. These resistors effectively damp down the interference waves. Even coil on plug (COP) systems use resistive plugs.

Bonding (electrically connecting on panel to another) is done to ensure all parts of the vehicle are at the same electrical potential to prevent sparking due to the build-up of static.

Screening is generally only used for specialist applications such as emergency services and the military. It involves completely enclosing the ignition system and other major sources of noise, in a conductive screen, which is connected to the vehicle's chassis earth. This prevents interference waves escaping; it is a very effective technique but expensive. Often, a limited amount of screening – metal covers on the plugs for example – can be used to good effect.

Capacitors and inductors are used to act as filters. This is achieved by using the changing value of 'resistance' to alternating signals as the frequency increases. The correct term for this resistance is either capacitive or inductive reactance. By choosing suitable values of a capacitor in parallel and or an inductor in series it is possible to filter out unwanted signals of certain frequencies.

The aerial is worth a mention at this stage. Several types are in use; the most popular still being the rod aerial, which is often telescopic. The advantage of a rod aerial is that it extends beyond the interference field of the vehicle. For reception in the AM bands the aerial represents a capacitance of 80 pF with a shunt resistance of about 1 MΩ. The set will often incorporate a trimmer to ensure the aerial is matched to the set. Contact resistance between all parts of the aerial should be less than 20 mΩ. This is particularly important for the earth connection.

When receiving in the FM range, the length of the aerial is very important. The ideal length of a rod aerial for FM reception is one quarter of the wavelength.

In the middle of the FM band (94 MHz) this is about 80 cm. Due to the magnetic and electrical field of the vehicle and the effect of the coaxial cable, the most practical length is about 1 m. Some smaller aerials are available but whilst these may be more practical the signal strength is reduced. Aerials embedded into the vehicle windows or using the heated rear window element are good from the damage prevention aspect and insensitivity to moisture, but produce a weaker signal, often requiring an aerial amplifier to be included. Note that this will also amplify interference. Some top-range vehicles use a rod aerial and a screen aerial, the set being able to detect and use the strongest signal. This reduces the effect of reflected signals and causes less flutter.

Consideration must be given to the position of an external aerial. This has to be a compromise taking into account the following factors.

- Rod length – 1 m if possible.
- Coaxial cable length – longer cable reduces the signal strength.
- Position – as far away as reasonably possible from the ignition system.
- Potential for vandalism – out of easy reach.
- Aesthetic appearance – does it fit with the style of the vehicle?
- Angle of fitting – vertical is best for AM, horizontal for FM.

Most quality sets also include a system known as interference absorption. This is a circuit built into the set consisting of high quality filters. Digital sets include error checking and correction systems.

Key fact

Most quality analog radios include an interference absorption system. Digital sets include error checking and correction systems.

Figure 16.30 shows a circuit of an earlier ICE system. An electric aerial is included and also the connection to a multi compact disc unit via a data bus.

Shown in Figure 16.31 is a block diagram showing inputs, control and outputs of a modern ‘infotainment’ system. Note the ‘Mic In’ (microphone input) so as to allow integration with a mobile phone and for voice control.

16.4.9 Mobile communications

If the success of the cellular industry is any indication of how much use we can make of the telephone, the future promises an even greater expansion. Cellular technology started to become useful in the 1980s and has continued to develop from then – very quickly!

The need and desire we perceive to keep in touch with each other is so great that an increasing number of business people now have several telephone numbers:

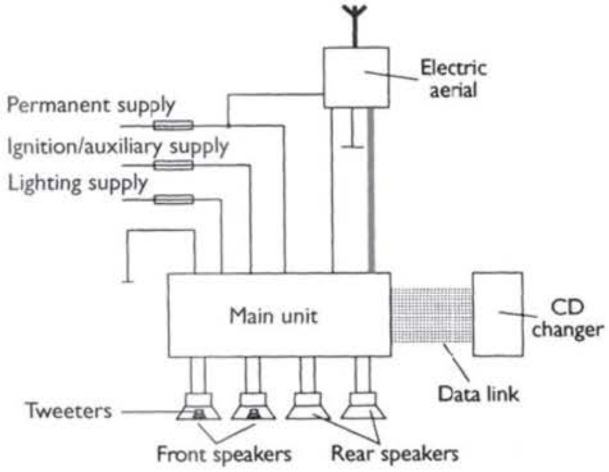


Figure 16.30 ICE system wiring

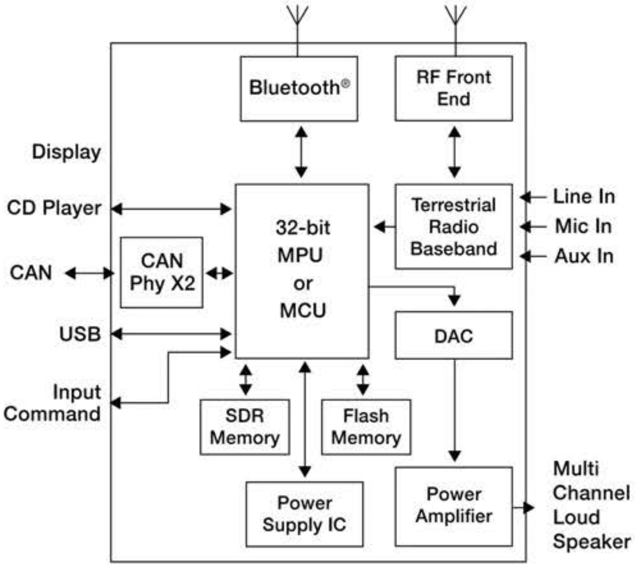


Figure 16.31 Block diagram showing inputs, control and outputs of an 'infotainment' system (Source: Freescale Electronics)

home, office, pager, fax and mobile/cellular (and more than one of these in some cases!).

But where does this leave communication systems relating to the vehicle? It is my opinion that 'in-vehicle' communication equipment for normal business and personal use will be by the smartphone and that there is no further market for the car telephone. Hands-free conversions will still be important as will how the phone connects to the car (see Chapter 4 for additional information).

CB radios and short-range two-way systems such as used by taxi firms and service industries will still have a place for the time being. However, even these may decline as the cellular network becomes cheaper and more convenient to use.

16.5 Security

16.5.1 Introduction

Stolen cars and theft from cars account for about a quarter of all reported crime. A huge number of cars are reported missing each year and over 20% are never recovered. Even when returned many are damaged. Most car thieves are opportunists, so even a basic alarm system can serve as a deterrent.

Car and alarm manufacturers are constantly fighting to improve security. Building the alarm system as an integral part of the vehicle electronics has made significant improvements. Even so, retrofit systems can still be very effective. Three main types of intruder alarm are used.

- Switch operated on all entry points.
- Battery voltage sensed.
- Volumetric sensing.

There are three common ways to disable the vehicle.

- Ignition circuit cut off.
- Starter circuit cut off.
- Engine ECU code lock.

Key fact

Stolen cars and theft from cars account for about a quarter of all reported crime.

Figure 16.32 Alarm bonnet/hood switch

A separate switch or remote transmitter can be used to set an alarm system. Often, they are set automatically when the doors are locked.

16.5.2 Basic security

To help introduce the principles of a vehicle alarm, this section will describe a very simple system, First, the requirements of this particular alarm system:

- It must activate when a door is opened.
- The ignition to be disabled.
- The existing horn is used as the warning.
- Once triggered, the horn must continue even when the door is closed.
- It must reset after 15 seconds.

The design will be based around a simple relay circuit. When a door is opened, the switches make an earth/ground connection (Figure 16.32). This will be used to trigger the relay, which in turn will operate the horn. The delay must be built in using a capacitor, which will keep the relay energized even after the door closes, for a further 15 seconds.

An external key switch is to be used to arm and disarm whilst isolating the ignition supply. Figure 16.33 shows a simple alarm circuit, which should achieve some of the aims. The delay is achieved by using a CR circuit; the 'R' is the resistance of the relay coil. Using the following data the capacitor value can be calculated.

- Time delay 15 s.
- Relay coil 120 Ω.
- Supply voltage 12 V.
- Relay drop out 8 V.

A capacitor will discharge to about 66% of its full value in *CR* seconds. The supply voltage is 12 V, so 66% of this is 8 V.

Therefore, if $CR = 15$, then, $C = 15/120$

$C = 125$ mF

This seems an ideal simple solution – but it is not. The main reason being that if the door is closed again – the alarm stops. However, it does illustrate the principle.

16.5.3 Top of the range security

The following is an overview of an alarm system now available either as a retro-fit or factory fitted. Most systems have electronic sirens (or use the vehicle

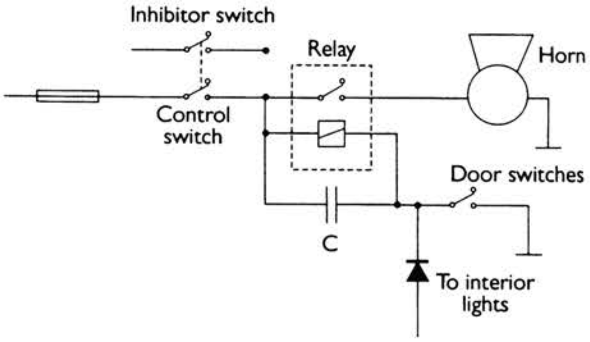


Figure 16.33 Simple alarm circuit the entry delay is made by using a CR circuit

horn) and may give an audible signal when arming and disarming. They are all triggered when the car door opens and will automatically reset after a period of time. The alarms are triggered instantly when an entry point is breached. Most systems can be considered as two pieces, with a separate control unit and siren; most will have the control unit in the passenger compartment and the siren under the bonnet.

Most systems now come with two remote 'keys' that use small button-type batteries and may have an LED that shows when the signal is being sent. They operate with one vehicle only. On aftermarket kits, intrusion sensors such as car movement and volumetric sensing can be adjusted for sensitivity.

When operating with flashing lights most systems draw about 5 A. Without flashing lights (siren only) the current drawn is less than 1 A. The sirens produce a sound level of about 95 dB, when measured 2 m in front of the vehicle. The alarm should only sounds for one minute then re-arm after 45 seconds, until deactivated.

Key fact

The alarm should only sounds for one minute then re-arm after 45 seconds, until deactivated.

Figure 16.34 shows a block diagram of a complex alarm system. The system, as is usual, can be considered as a series of inputs and outputs.

Inputs

- Ignition supply.
- Engine crank signal.
- Volumetric sensor.
- Bonnet switch.
- Trembler switch.
- IR/RF remote (Figure 16.27).
- Doors switches.
- Control switch.

Outputs

- Volumetric transmitter.
- System LED.
- Horn or siren.
- Hazard lights.
- Ignition immobilizer.

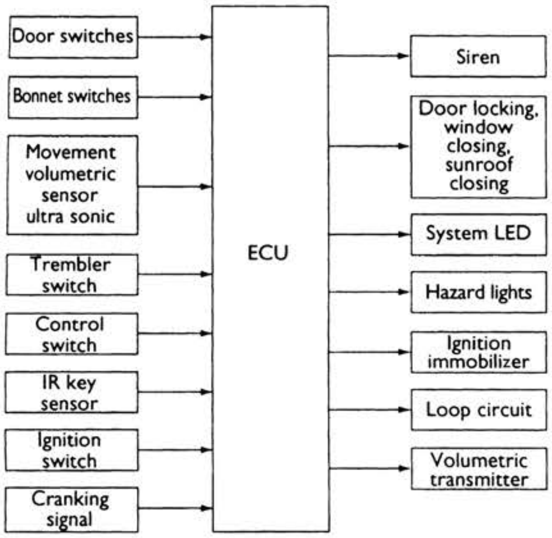


Figure 16.34 Block diagram of a complex alarm system

Key fact

Most factory fitted alarms are combined with the central door locking system.

- Loop circuit.
- Electric windows, sun-roof and door locks.

Most factory fitted alarms are combined with the central door locking system. This allows the facility mentioned in a previous section known as lazy lock. Pressing the button on the remote unit, and as well as setting the alarm, the windows and sun-roof close, and the doors lock.

16.5.4 Security-coded ECUs

A security code in the engine electronic control unit is a powerful deterrent. This can only be 'unlocked' to allow the engine to start when it receives a coded signal. Ford and other manufacturers use a special ignition key that is programmed with the required information. Even the correct 'cut' key will not start the engine.

Of course nothing will stop the car being lifted on to a truck and driven away, but this technique will mean a new engine control ECU will be needed by the thieves. The cost will be high and also questions may be asked as to why a new ECU is required.

16.5.5 Alarms and immobilizers

A modern anti-theft alarm circuit shown as Figure 16.35. As with all complex systems it can be considered as a black box with inputs and outputs. The inputs are signals from key and lock switches as well as monitoring sensors. The outputs are the alarm horn and the hazard lights but also starter inhibitor relays, etc.

Alarm system

This system can be operated by remote control or using the key in a door lock. When first activated, the system checks that the doors and tailgate are closed by monitoring the appropriate switches. If all is in order, the anti-theft system is then activated after a 20-second delay. The function indicator LED flashes rapidly during this time and then slowly once the system is fully active.

The alarm can be triggered in a number of ways:

- Opening a door, the tailgate or the bonnet/hood.
- Removal of the radio connector loop.
- Switching on the ignition.
- Movement inside the vehicle.

If the alarm is triggered the horn operates for 30 seconds and the hazard lights for 5 minutes. This stops if the remote key or door key is used to unlock the vehicle.

Passive anti-theft system (PATS)

This system is a vehicle immobilizer developed by Ford (Figure 16.36). It is activated directly through the ignition switch by means of an electronic code stored in a special key. Each key has a transponder that stores the code, which does not require a battery. The key code is read by the receiver (which is part of the ignition switch) when the key is turned from position 0 to 1 or 2 (usually marked as I or II). If the code matches the one stored in the module, then it allows the engine to start. These systems operate independently of the alarm.

Key programming

Some keys and/or remotes for later vehicles may need to be reprogrammed if, for example, the battery goes flat or a new key is required. There are several

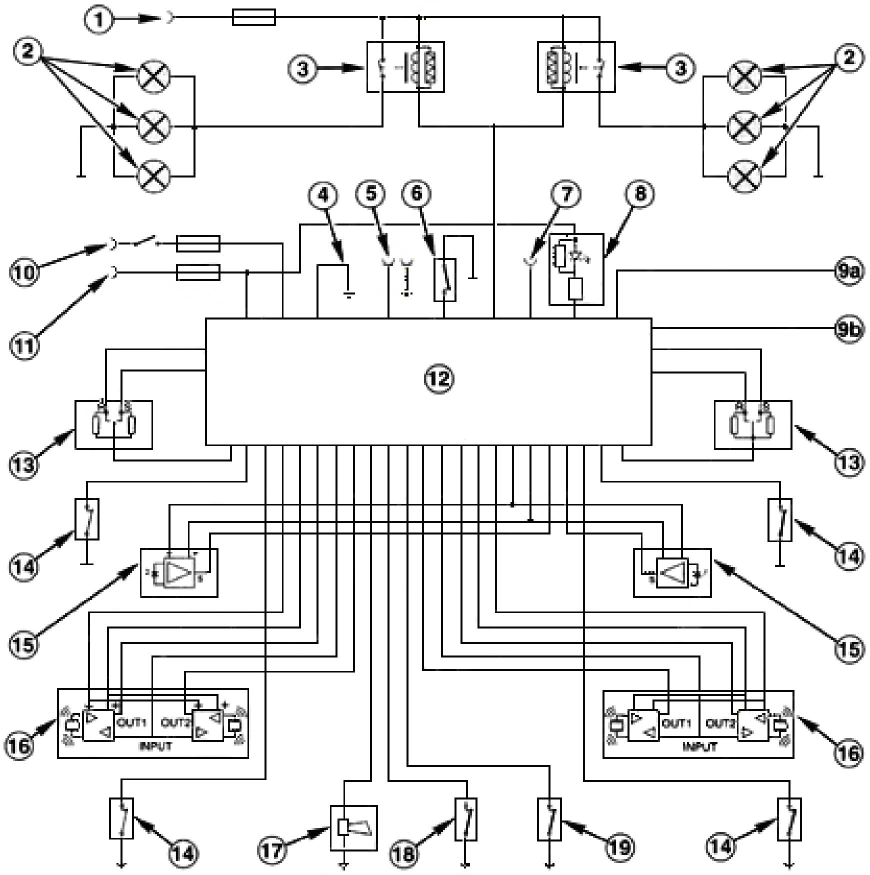


Figure 16.35 Anti-theft alarm system with remote control and interior monitoring: 1, Fused supply. 2, Hazard lights. 3. Hazard lights alarm relay. 4. Earth/Ground. 5. Diagnostic connector. 6. Bonnet/Hood switch. 7. Connector for radio theft protection. 8. Function light. 9. Input signal locked/unlocked. 10. Ignition supply. 11. Battery supply. 12. Anti-theft alarm ECU. 13. Left/Right door key switch. 14. Door switches. 15. Infrared receiver. 16. Ultrasound sensors. 17. Horn. 18, Tailgate switch. 19, Tailgate key switch (Source: Ford Motor Company)

methods of programming remote keys. However, different manufacturers use various methods and it is therefore not possible to cover all of these. A few methods are described here as examples. PATS key programming:

In earlier systems a red key is used as a master; it is exactly the same as the other keys apart from its colour. This key is the only one that can program new keys – if lost the whole system has to be reprogrammed by a dealer – and a new master supplied.

To program a red key system insert the master key into the ignition and turn it to position II. When the light on the clock goes out remove the key. The light will come back on if the master key was used. While the light is still on, insert the new key and turn to position II. The light will flash twice and the key is programmed.

To program a two key system both of the original keys are needed. Insert the keys one after the other in the ignition, turn to position II and then remove. After the second key is removed, insert the new un-programmed key, switch to position II and then remove it. The new key is now programmed.

Remember not to put an un-programmed PATS key in the ignition unless following the above procedure – it will immobilize the vehicle for 30 minutes!

Key fact

Some keys and/or remotes may need to be reprogrammed if, for example, the battery goes flat or a new key is required.

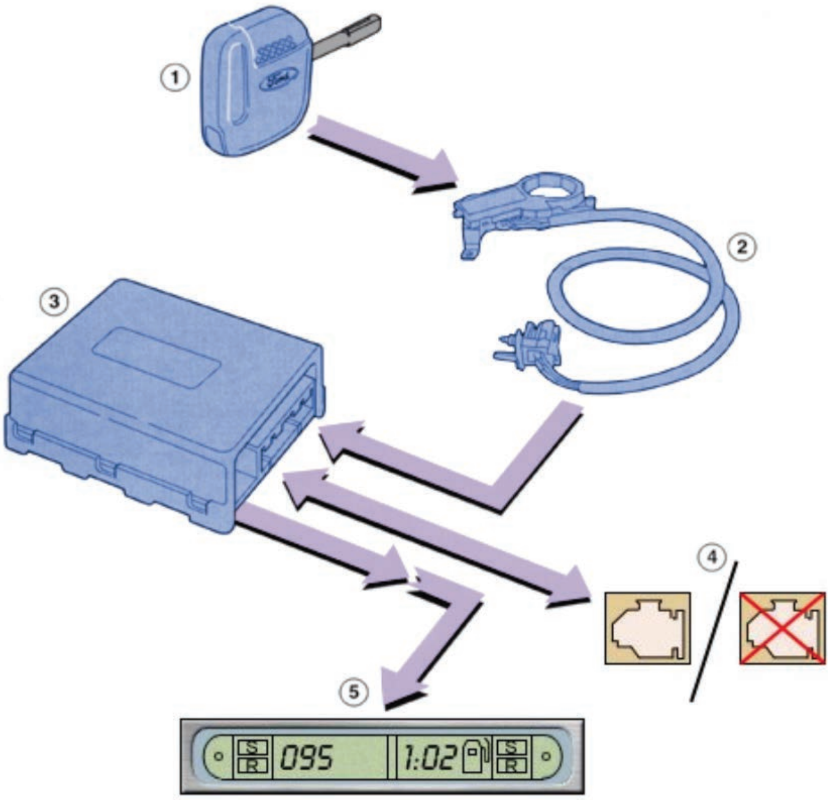


Figure 16.36 PATS components: 1. Key integrated transponder. 2. Transmitter/Receiver. 3. PATS module. 4. Engine start – yes/no. 5. Clock with integrated function indicator (Source: Ford)

Remote keys (example only)

Switch the ignition from I to II quickly 4 times – this illuminates the alarm warning light. Remove the key from the ignition and point it at the remote sensor (interior mirror usually). Press and hold one of the buttons until the light on the remote flashes. Keep holding the first button, press the other button 3 times and finally release both buttons. The light on the remote and the warning light will flash 5 times – the remote key is now programmed.

On some vehicles, switching the ignition from I to II quickly 4 times will activate a chime. Remove the key and press any of the buttons to activate another chime. Finally, replace the key and turn the ignition to position II – the remote key is now programmed.

A useful tip is that on many remotes changing the batteries within 15 seconds will mean they do not need to be reprogrammed.

Fault diagnosis

Many vehicle manufacturers use equipment connected to a diagnostic link connector (DLC) to check several systems, including alarms. This is the same DLC as used for engine management diagnostics. See the sections on OBD for more details. Test equipment is becoming available that can be used by independent repairers. However, it is not often cost effective to purchase this for specific vehicles.

As with others, an alarm system can be treated as a black box system. In other words, checking the inputs and outputs for correct operation means the complexity inside the ECU can be largely ignored. Note that most alarms will not set if the module is receiving an incorrect input signal when it is activated (door switch open/closed for example).

Key fact

Many vehicle manufacturers use equipment connected to a diagnostic link connector (DLC) to check several systems, including alarms.

16.5.6 Keys

Remote keyless entry (RKE)

Remote keyless entry has been a feature on many cars for a number of years. Remote keys work by transmitting either radio frequency or infrared signals. Door locking is controlled by a small handheld transmitter and a receiver unit, as well as a decoder in the main control unit. This layout varies slightly between different manufacturers.

When the remote key is operated (by pressing a small switch), a complex code is transmitted. The number of codes used is in excess of 50 000. The receiver sensor picks up this code and sends it in an electrical form to the main control unit. If the received code is correct, the relays are triggered and the doors are either locked or unlocked. On some systems, if an incorrect code is received on three consecutive occasions when attempting to unlock the doors, the system will switch itself off until the door is opened by the key. This action resets the system and allows the correct code to operate the locks again. This technique prevents a scanning type transmitter unit from being used to open the doors.

Passive keyless entry (PKE)

Passive keyless entry systems mean the driver doesn't even need to press a button to unlock the vehicle! The electronic key is simply carried in a pocket, on a belt clip or in a bag. The controllers in the doors communicate with the key using radio frequency (RF). This action determines if the correct key is present and, if it is, the doors are unlocked.

This communication event is triggered by lifting the door handle, or in some cases the vehicle will even unlock as the key holder approaches it.

PKE systems need the same level of security as any other remote locking method. Conventional RKE is a unidirectional process. In other words, signals are only sent from the key to the receiver. With PKE the communication is two-way. This is because the PKE system carries out an 'identity friend or foe' (IFF) operation for security purposes. The vehicle sends a random challenge to the key; the key encrypts this value and sends it back to the vehicle. The vehicle then

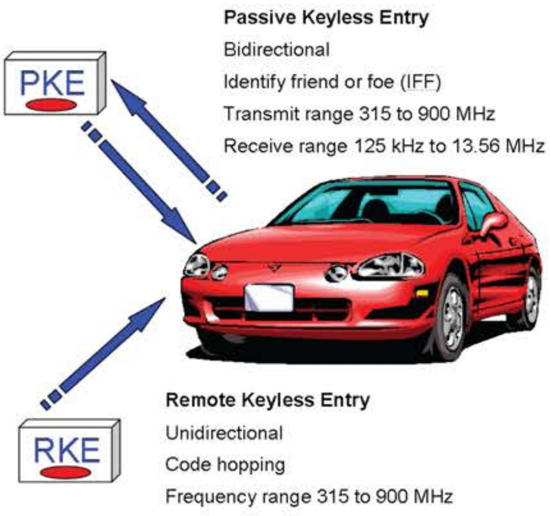


Figure 16.37 Remote and passive key entry systems

performs the same encryption, compares the result with that sent by the key, and unlocks the doors if the values match.

Battery life is a critical issue for PKE. To obtain the required range of operation, 1.5 m (5 ft), the detection circuit in the key needs to be sensitive enough to detect just a few mV; this consumes significant power. There is also an issue with power consumption for the base station (vehicle) if the doors are designed to unlock as the key approaches. To achieve this the base station must poll continuously. In other words, it must keep looking for the key. This consumes battery power, which could be an issue if the vehicle was left for a long period. However, this method does have the advantage that the doors will always be locked unless a key is present.

If the method of lifting a handle is used as a trigger, then no power is consumed until needed. The down side of this method is that the user will want to feel the door unlock as the handle is lifted. However, Texas Instruments has developed a low-frequency RF chip. With a standby current of 5 A and less than 10 mV peak-to-peak sensitivity, the chip therefore provides a long battery life. It comes in an industry-standard package small enough to fit into a key fob or credit card device. This type of system is likely to become very common. Some PKE systems can even be set up to recognize multiple keys. The car could even be programmed to ‘know’ who was driving and set seat and mirror positions automatically!

Passive keyless go and exit

When the driver enters a car the key remains in a ‘pocket’ or at least it will be inside the vehicle. This means assuming that the key is being recognized, engine starting can be by a simple start button. As the button is pressed the same authentication process that takes place for the door locks starts. The engine can only be started if the key is inside the car, which is a technical challenge for the designers. For example, the key could be in a jacket hanging above the back seat, or it could be in the jacket outside on the roof.

Philips Semiconductors have produced a system with receive signal strength identification (RSSI), which can detect whether the key is inside or outside the vehicle. After the occupants have left the vehicle, the doors can be locked by pressing a handle or as the driver leaves the vicinity. ‘Inside/outside’ detection is also necessary for this scenario so the key cannot be locked in the car.

Keypad entry

In vehicles equipped with a keypad entry system, the vehicle doors and the boot can be locked and unlocked without using a key. Before unlocking the boot or a passenger door, the driver’s door must be unlocked. Usually, if more than five seconds pass between pressing numbers on the keypad, the system will shut down and the code has to be entered again.

To unlock the driver’s door, the factory code or a personal code is entered. All codes have five numbers. After the fifth number is pressed, the driver’s door unlocks. The passenger doors can then be unlocked by pressing the 3/4 button within five seconds of unlocking the driver’s door. To unlock the boot, the 5/6 button must also be pressed within five seconds. If this time is exceeded, the code to open the driver’s door must be re-entered.

The keypad can also be used to lock the doors. To lock all of the car doors at the same time, 7/8 and 9/0 need to be pressed at the same time. It is not necessary to enter the keypad code. This will also arm the anti-theft system if fitted.

Figure 16.38 Standard key and remote transmitter combined

16.6 Airbags and belt tensioners

16.6.1 Introduction

A seat-belt, seat-belt tensioner and an airbag are, at present, the most effective restraint system in the event of a serious accident. At collision speeds in excess of 40 km/h the seat-belt alone is no longer adequate. Research after a number of accidents has determined that in 68% of cases an airbag provides a significant improvement. It is suggested that if all cars in the world were fitted with an airbag then the number of fatalities annually would be reduced by well over 50 000. Some airbag safety issues have been apparent in the USA where airbags are larger and more powerful. This is because in many areas the wearing of seat-belts is less frequent.

The method becoming most popular for an airbag system is that of building most of the required components into one unit. This reduces the amount of wiring and connections, thus improving reliability. An important aspect is that some form of system monitoring must be built-in, as the operation cannot be tested – it only ever works once. Figure 16.39 shows airbags operating – on dummies.

Safety first

In collisions in excess of 40 km/h the seat-belt alone is no longer adequate protection.

16.6.2 Operation of the system

The sequence of events in the case of a frontal impact at about 35 km/h, as shown in Figure 16.40, is as follows.

1. The driver is in the normal seating position prior to impact.
2. About 15 ms after the impact the vehicle is strongly decelerated and the threshold for triggering the airbag is reached. The igniter ignites the fuel tablets in the inflater.
3. After about 30 ms the airbag unfolds and the driver will have moved forwards as the vehicle's crumple zones collapse. The seat-belt will have locked or been tensioned depending on the system.
4. At 40 ms after impact the airbag will be fully inflated and the driver's momentum will be absorbed by the airbag.
5. About 120 ms after impact the driver will be moved back into the seat and the airbag will have almost deflated through the side vents, allowing driver visibility.

Passenger airbag events are similar to the above description. A number of arrangements are used with the mounting of all components in the steering

Figure 16.39 Don't be a crash test dummy!

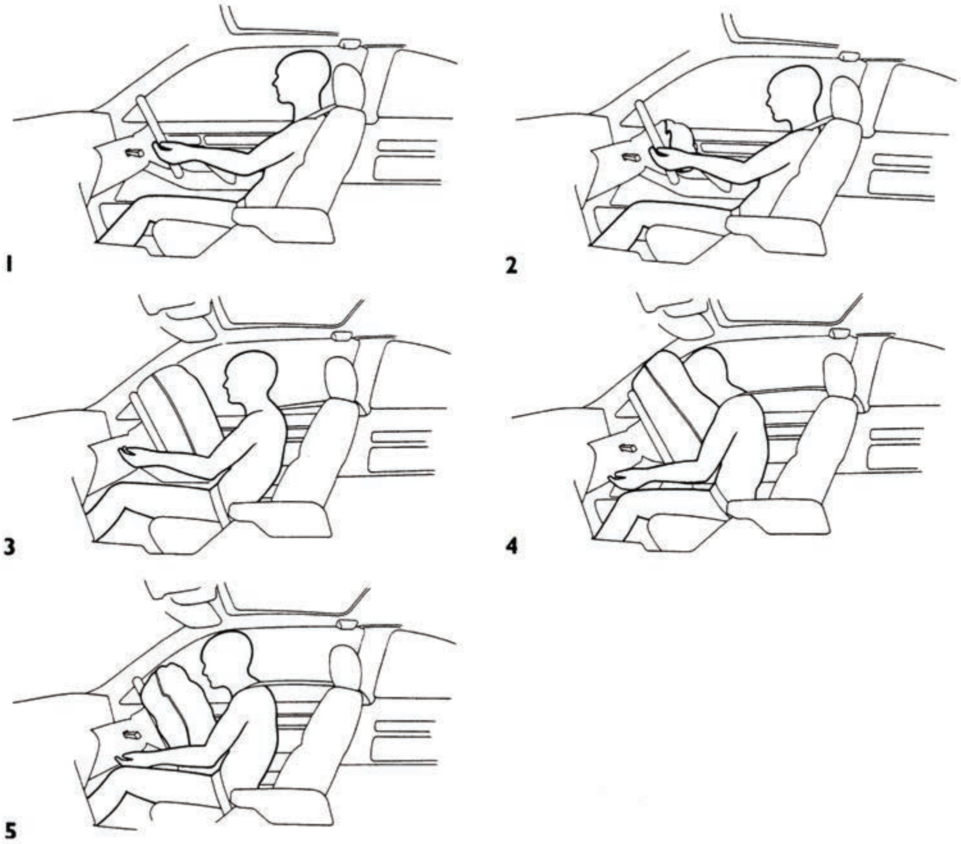


Figure 16.40 Airbag in action

wheel centre becoming the most popular. Nonetheless, the basic principle of operation is the same.

16.6.3 Components and circuit

The main components of a basic airbag system are as follows.

- Driver and passenger airbags.
- Warning light.
- Passenger seat switches.
- Pyrotechnic inflaters (squibs).
- Igniter.
- Crash sensor(s).
- Electronic control unit.

The airbag is made of a nylon fabric with a coating on the inside. Prior to inflation the airbag is folded up under suitable padding that has specially designed break lines built-in. Holes are provided in the side of the airbag to allow rapid deflation after deployment. The driver's air has a volume of about 60 litres and the passenger airbag about 160 litres.

Key fact

The airbag is made of a nylon fabric with a coating on the inside.

A warning light is used as part of the system monitoring circuit. This gives an indication of a potential malfunction and is an important part of the circuit. Some manufacturers use two bulbs for added reliability.

Consideration is being given to the use of a seat switch on the passenger side to prevent deployment when not occupied. This may be more appropriate to side-impact airbags mentioned in the next section.

The pyrotechnic inflater and the igniter can be considered together. The inflater in the case of the driver is located in the centre of the steering wheel (Figure 16.41). It contains a number of fuel tablets in a combustion chamber. The igniter consists of charged capacitors, which produce the ignition spark. The fuel tablets burn very rapidly and produce a given quantity of nitrogen gas at a given pressure. This gas is forced into the airbag through a filter and the bag inflates breaking through the padding in the wheel centre. After deployment, a small amount of sodium hydroxide will be present in the airbag and vehicle interior. Personal protection equipment must be used when removing the old system and cleaning the vehicle interior.

Key fact

The fuel tablets burn very rapidly and produce a given quantity of nitrogen gas at a given pressure.

The crash sensor can take a number of forms; these can be described as mechanical or electronic. The mechanical system (Figure 16.42) works by a

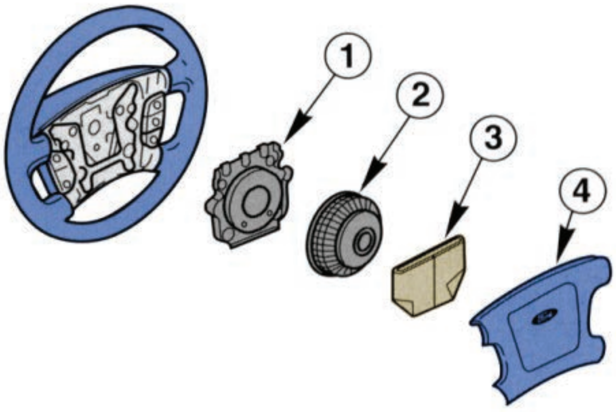


Figure 16.41 Inflater in the steering wheel: 1, Base and connector; 2, Inflater; 3, Airbag; 4, Cover

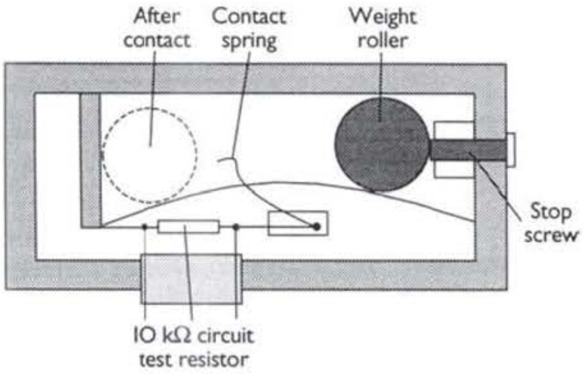


Figure 16.42 The mechanical impact sensor works by a sprung holding a roller

spring holding a roller in a set position until an impact above a predetermined limit, provides enough force to overcome the spring and the roller moves, triggering a micro switch. The switch is normally open with a resistor in parallel to allow the system to be monitored. Two switches similar to this may be used to ensure the bag is deployed only in the case of sufficient frontal impact. Note that the airbag is not deployed in the event of a roll over.

The other main type of crash sensor can be described as an accelerometer. This will sense deceleration, which is negative acceleration. Figure 16.43 is a sensor based on strain gauges.

Figure 16.44 shows two types of piezoelectric crystal accelerometers, one much like an engine knock sensor and the other using spring elements. A severe change in speed of the vehicle will cause an output from these sensors as the seismic mass moves or the springs bend. Suitable electronic circuits can monitor this and be pre-programmed to react further when a signal beyond a set threshold is reached. The advantage of this technique is that the sensors do not have to be designed for specific vehicles, as the changes can be software-based.

The final component to be considered is the electronic control unit or diagnostic control unit. When a mechanical-type crash sensor is used, in theory no electronic unit would be required. A simple circuit could be used to deploy

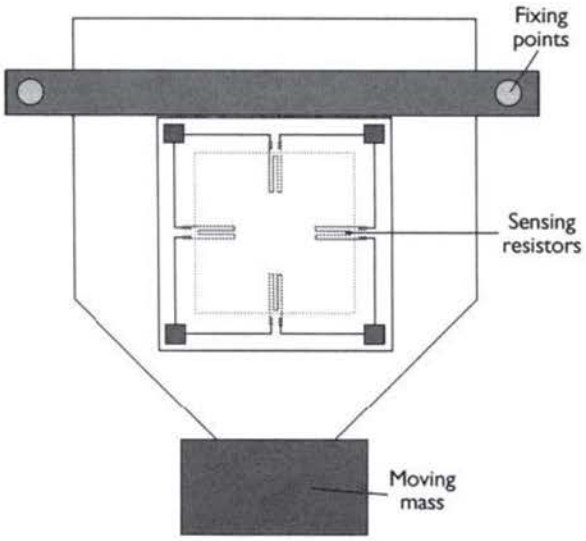


Figure 16.43 Strain gauges accelerometer

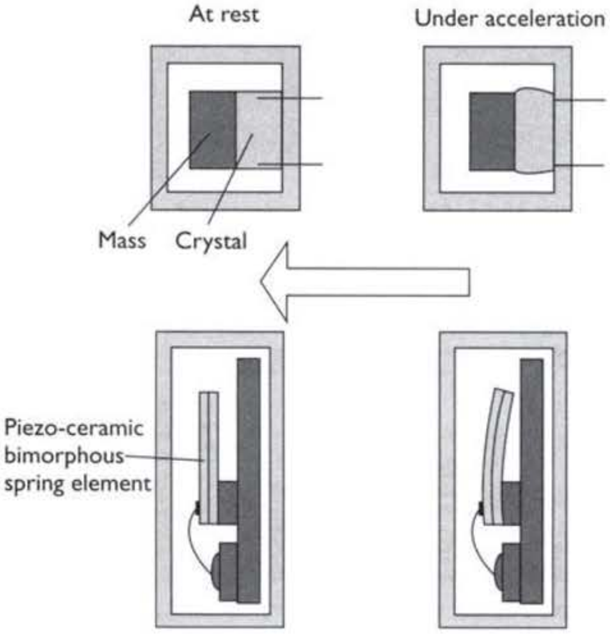


Figure 16.44 Piezoelectric crystal accelerometer

the airbag when the sensor switch was operated. However, it is the system monitoring or diagnostic part of the ECU, which is most important. If a failure is detected in any part of the circuit then the warning light will be operated. Up to five or more faults can be stored in the ECU memory, which can be accessed by blink code or serial fault readers. Conventional testing of the system with a multimeter and jump wires is not to be recommended as it might cause the airbag to deploy! Figure 16.45 shows an airbag ECU as well as its inputs and outputs.

Safety first

Conventional testing of the system with a multimeter and jump wires is not to be recommended as it might cause the airbag to deploy!

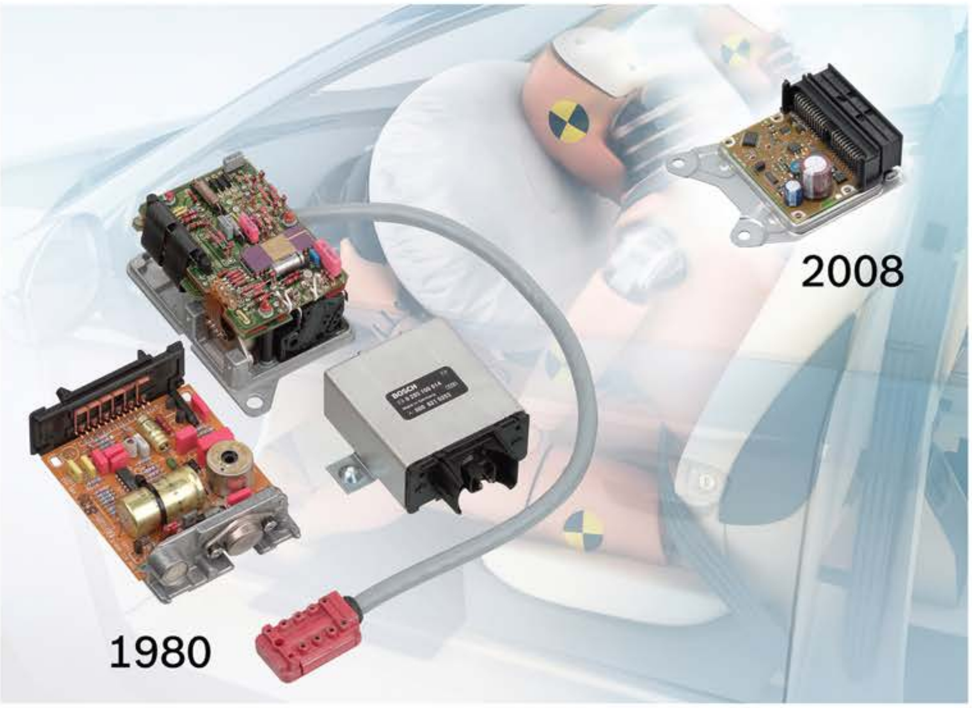


Figure 16.45 Airbag ECU development

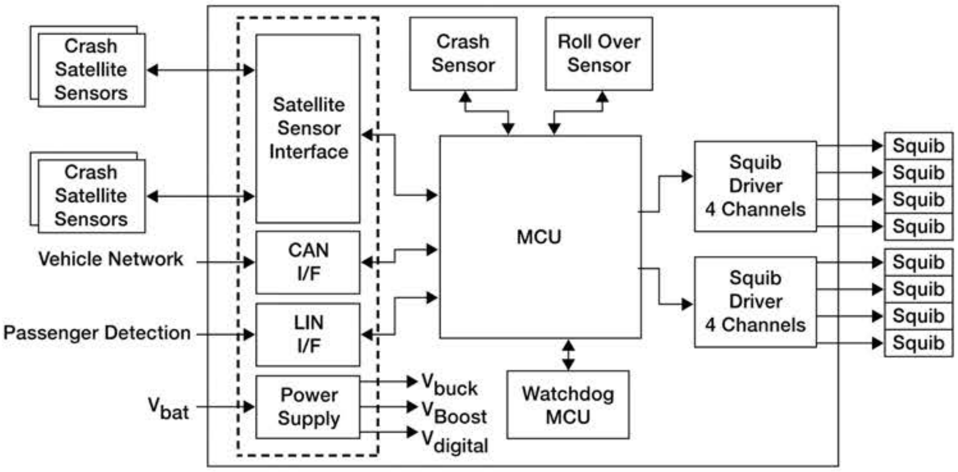


Figure 16.46 Airbag circuit block diagram showing inputs, control and outputs of an airbag system (Source: Freescale Electronics)

A block diagram of an airbag circuit is shown in Figure 16.46. Note the 'watchdog' system which ensures correct operation at all times. A digital-based system using electronic sensors has about 10 ms at a vehicle speed of 50 km/h, to decide if the restraint systems should be activated. In this time more than 10 000 computing operations are necessary. Data for the development of these algorithms are based on computer simulations but digital systems can also remember the events during a crash, allowing real data to be collected.

16.6.4 Seat-belt tensioners

Taking the 'slack' out of a seat-belt in the event of an impact is a good contribution to vehicle passenger safety. The decision to take this action is the same as for the airbag inflation. The two main types of tensioners are:

- Spring tension.
- Pyrotechnic.

The mechanism used by one type of seat-belt tensioner is shown in Figure 16.47. When the explosive charge is fired, the cable pulls a lever on the seat-belt reel, which in turn tightens the belt. The unit must be replaced once deployed. This feature is sometimes described as anti-submarining.

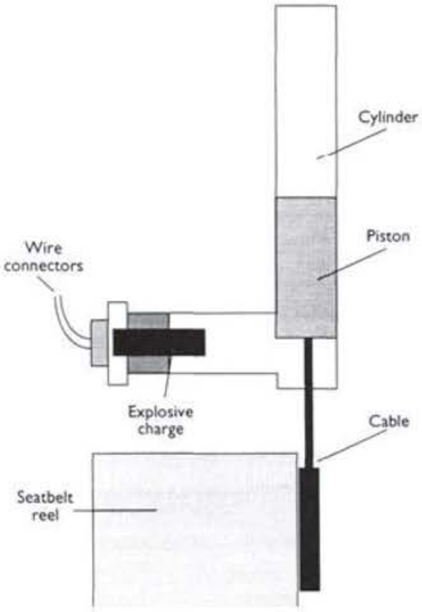


Figure 16.47 The mechanism used by one type of seat-belt tensioner

16.6.5 Side airbags

Airbags working on the same techniques to those described previously are being used to protect against side impacts. In some cases bags are stowed in the door pillars or the edge of the roof. Figure 16.48 shows this system.

Figure 16.49 shows a full seat-belt and airbag system used by Ford.

16.6.6 Intelligent airbag sensing system

An 'Intelligent Airbag Sensing System' has been developed, which can determine the right reaction for a specific accident situation. The system can control a one- or two-stage airbag inflation process via a two-stage gas generator. Acting on signals from vehicle acceleration and belt buckle sensors, which vary according to the severity of the accident, the gas generator receives different control

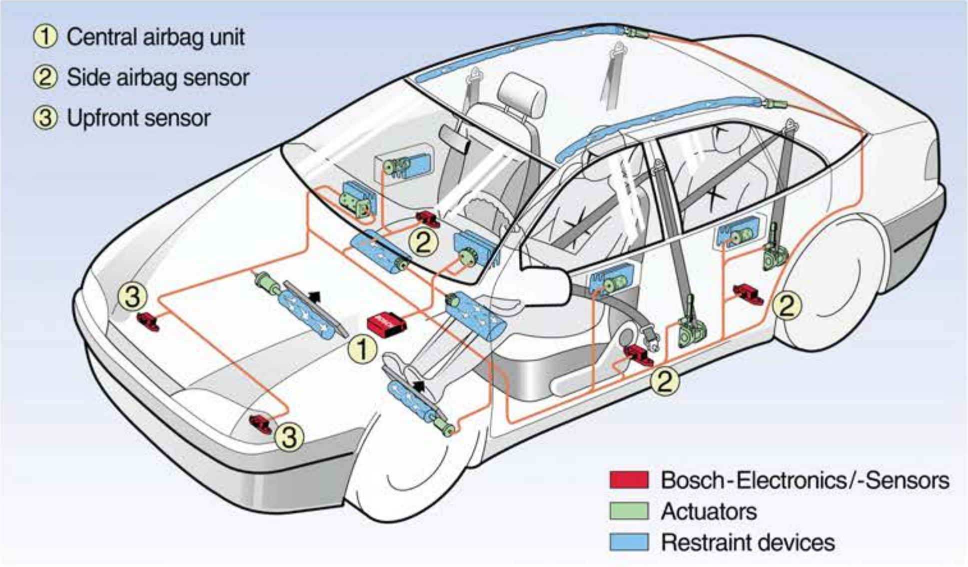


Figure 16.48 Optimized airbag control (Source: Bosch Media)

pulses, firing off one airbag stage (de-powering), both stages (full inflation), or staged inflation with a time interval.

Future developments will lead to capabilities for multistage inflation or a controllable sequence of inflation following a pattern determined by the type of accident and the position of the vehicle occupants. The introduction of an automotive occupancy sensing (AOS) unit that uses ultrasonic and infrared sensors will provide further enhancements. This additional module will detect seat and child occupancy and will be capable of assessing whether a passenger is in a particular position, such as feet on the dashboard!

The latest radar technology will also assist the design of a pre-crash sensor capable of detecting an estimated impact speed prior to collision, and activating individual restraint systems, such as seat-belt pre-tensioners. Or, when necessary, all available restraint systems will be activated.

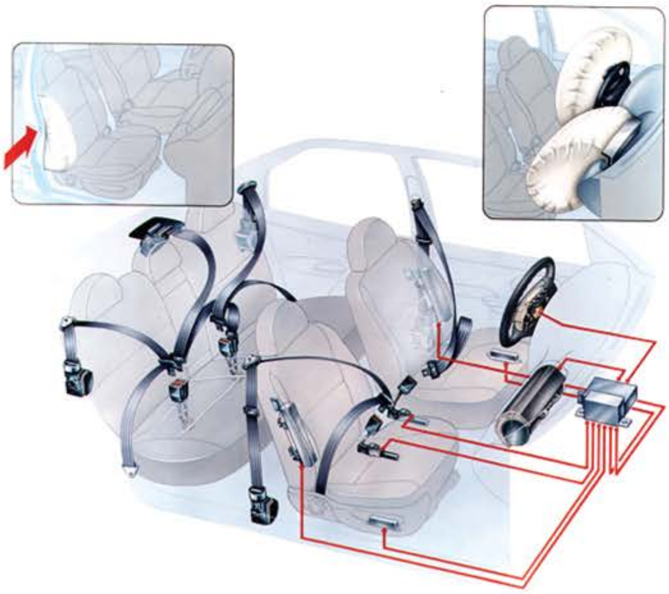


Figure 16.49 Seat-belt and airbag operation

16.7 Other safety and comfort systems

16.7.1 Obstacle avoidance radar

This system, sometimes called collision avoidance radar, can be looked at in two ways. First, an aid to reversing, which gives the driver some indication as to how much space is behind the car. Second, collision avoidance radar can be used as a 'vision' enhancement system.

The principle of radar as a reversing aid is illustrated in Figure 16.50. This technique is, in effect, a range-finding system. The output can be audio or visual, the latter being perhaps most appropriate, as the driver is likely to be looking backwards. The audible signal is a 'pip-pip-pip' type sound, the repetition frequency of which increases as the car comes nearer to the obstruction, and becomes almost continuous as impact is imminent. Many systems now also make the noise come from the appropriate speaker(s) so as to indicate direction.

A reverse sensing system is a reverse only parking aid system that uses sensors mounted in the rear bumper. Parking aid systems feature both front and rear sensors. Low-cost, high-performance ultrasonic range sensors are fitted to the vehicle. Generally, four intelligent sensors are used to form a detection zone as wide as the vehicle. A microprocessor monitors the sensors and emits audible beeps during slow reverse parking to help the driver back up or park the vehicle.

Key fact

A microprocessor monitors the sensors and emits audible beeps during slow reverse parking to help the driver back up or park the vehicle.

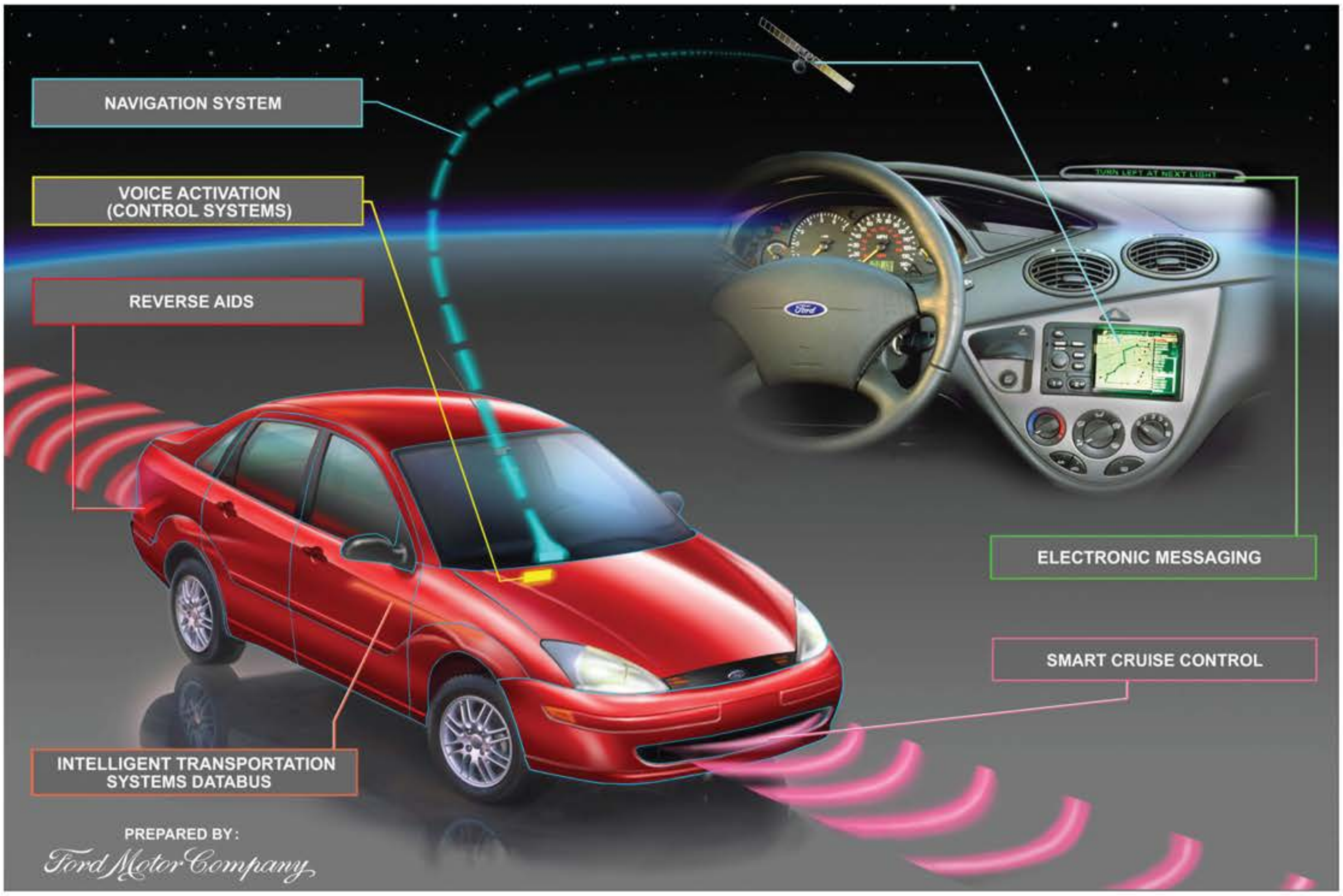


Figure 16.50 Reversing aid as part of a control system (Source: Ford)

Figure 16.51 Obstacle avoidance radar

This leads to easier and convenient reversing and parking manoeuvres, especially for vehicles where drivers have limited view at the front, rear or corners of the vehicle. The technique is relatively simple as the level of discrimination required is fairly low and the radar only has to operate over short distances.

Obstacle avoidance radar, when used as a vision enhancement system, is somewhat different. Figure 16.52 is a block diagram to demonstrate the principle of this system. This may be linked with adaptive cruise control, as discussed in an earlier section. A frequency of 94 GHz has been used for development work; this frequency is known as millimetre waves.

A short look at the history and principle of radar at this stage will help with an overall understanding. Radar was the name given during World War II to an electronic system by which radio waves were bounced off an aircraft in order to detect its presence and locate its position. The term is an acronym, made from the fuller term ‘radio detection and ranging’. A large number of researchers helped to develop the devices and techniques of radar, but the development of the earliest practical radar system is usually credited to Sir Robert Watson-Watt.

Definition

RADAR: The term is an acronym, made from the ‘radio detection and ranging’.

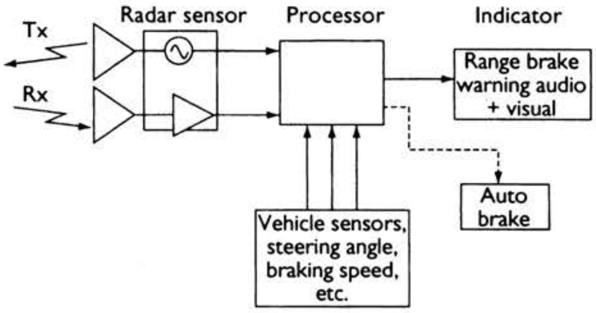


Figure 16.52 Block diagram of obstacle avoidance radar when as a vision enhancement system

Key fact

Because radio waves travel at a known constant velocity, the speed of light, which is 3 x 10^8 m/s, the range may be found by measuring the time taken for a radio wave to travel from transmitter to obstacle and back to the receiver.

The operation of a basic radar system is as follows: a radio transmitter generates radio waves, which are then radiated from an antenna, 'lighting up' the airspace with radio waves. A target, such as another vehicle that is in this space, scatters a small portion of the radio energy back to a receiving antenna. This weak signal is amplified by an electronic amplifier and displayed, often on a cathode ray tube. To determine its position, the distance (range) and bearing must be measured. Because radio waves travel at a known constant velocity, the speed of light, which is 3 x 10^8 m/s, the range may be found by measuring the time taken for a radio wave to travel from transmitter to obstacle and back to the receiver.

For example, if the range were 150 m, the time for the round trip would be:

$$t = \frac{2d}{C}$$

where: t = time, d = distance to object, and C = speed of light.

In this example:

$$t = \frac{2 \times 150}{3 \times 10^8}$$

Relative closing speed can be calculated from the current vehicle speed. The radar is actually transmitted in the form of pulses. This is done by frequency modulating the signal, maybe using a triangular wave with a frequency of the order of 100 MHz: this can also be used to trigger a display and for calculation of distance.

The bearing, if required, is given by the relative position on the display device. Radar for use in a vehicle must fulfil the following general requirements.

- Range to be at least 300 m in bad weather. This gives about 7 seconds warning at 160 k/h (100 mile/h).
- Objects greater than 0.1 m^2 must be detected.
- Data update greater than one per second.
- Beam spread of about 15° horizontal and vertical.
- The driver's display should not intrude on concentration and only act as a warning.

The type of display or output that may be used on a motor vehicle will vary from an audible warning to a warning light or series of lights and possibly a display screen.

16.7.2 Tyre pressure warning

Poorly inflated tyres cause loss of control and increase fuel consumption. The idea is to give the driver warning of reduced pressure – as rapid deflation is generally apparent to the driver!

There are three basic components to the system. Mounted in the wheel rim is a pressure operated switch, the contacts of which close when pressure falls. This is recognized by a high-frequency sender which the switch passes but does not contact as the wheel rotates. The high-frequency sender transmits an appropriate pulse to the electronic evaluator. If the pressure drops below the set value then the switch contacts open, causing the high-frequency sender to interrupt its stream of pulses to the evaluation circuit and the warning lamp comes on. The system measures the tyre pressure with an accuracy of 50 mbar. The design of the switch is such that changes in temperature of the air in the tyre will not cause false readings.

If the tyre pressure warning system is used in conjunction with wheels fitted with 'limp-home' tyres, it will provide a reminder that the limp-home mode is in use.

Another tyre pressure warning system using active analog sensors in the tyre and wireless transmission of the signal from the wheel to the body is under development. The advantage is that absolute values of pressure and temperature are measured continuously, even when the car is at rest. Values such as vehicle speed and load are also included in the calculation.

16.7.3 Noise control

The principle of adaptive noise control is that of using sound, which is identical and 180° out of phase, or in anti-phase, to cancel out the original source of noise. Figure 16.53 shows three signals, the original noise, the anti-phase cancelling waveform and the residual noise.

A microphone picks up the original noise. It is then inverted and amplified, and then replayed by a suitably positioned speaker. This effectively cancels out the noise. Whilst the theory is relatively simple, until recently it has not been particularly suitable for motor vehicle use. This is due to the wide range of noise frequencies produced, and the fast response time, which is needed to give acceptable results. Low-frequency noise (200 Hz), causes 'boom' in a vehicle, this is very difficult to reduce by conventional methods.

Much development time and money has been spent on reducing cabin noise levels. This can range from simple sound-deadening material to a special design of engine mountings, exhaust systems and using balance shafts on the engine. Even so, the demand still exists to reduce noise further and this is becoming ever more expensive.

Most vehicles today are susceptible to some low-frequency boom in the passenger compartment, even when a large amount of sound deadening is used. The trend to produce lighter vehicles using thinner grade metal further exacerbates the problem. Conventional techniques solve the problem at certain frequencies, not all across the range.

To apply the adaptive noise control system to a car required the development of high-speed digital signal processors as well as a detailed understanding of noise generation dynamics in the vehicle. A typical four-cylinder engine running between 600 and 6000 rpm has a firing frequency of about 20–200 Hz. There are several critical speeds at which the vehicle will display unpleasant boom. Low-profile tyres and harder suspension also generate considerable low-frequency noise.

Key fact

A typical four-cylinder engine running between 600 and 6000 rpm has a firing frequency of about 20–200 Hz.

One system uses several microphones embedded in the vehicle headlining to sample the noise. A digital signal processor measures the average sound

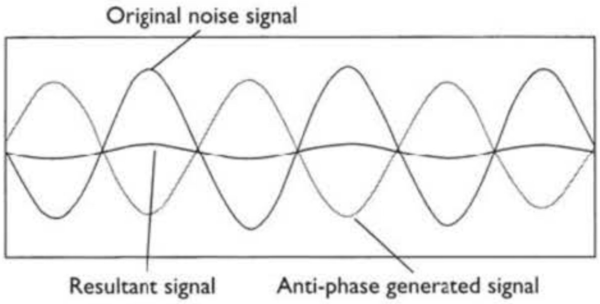


Figure 16.53 Three signals; the original noise, the anti-phase cancelling waveform and the residual noise

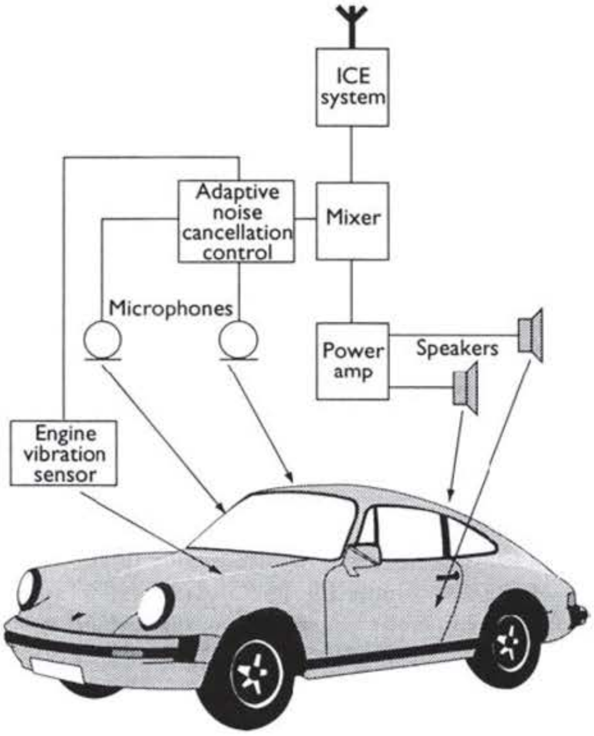


Figure 16.54 Layout of an adaptive noise control system

pressure energy across the cabin and adjusts the phase and amplitude of the anti-noise signals. These are played through the in-car speaker system until, by measuring the error signal from the microphones, a minimum noise is achieved. The maximum active noise control can be achieved in about 70 ms.

Key fact

A quality loudspeaker system is needed which must be able to produce up to 40 W RMS per channel.

A quality loudspeaker system is needed which must be able to produce up to 40 W RMS per channel. This is not uncommon on many ICE systems. Figure 16.54 shows a typical layout of an adaptive noise control system. The greatest improvements are gained in small vehicles where the perceived reduction is as much as 80%.

Other methods of noise control include active engine mountings and even exhaust monitoring.

A hydraulic engine mount, which is electronically controlled in response to the engine vibration, can significantly reduce noise. Some manufacturers, however, are now using a much simpler version, which can switch between hard and soft settings. The system is claimed to be as effective as about 45 kg of sound deadening material.

Key fact

A hydraulic engine mount, which is electronically controlled in response to the engine vibration, can significantly reduce noise.

An active muffler for reducing exhaust noise has been developed. The heart of this system is a digital processor. Two inputs are used, a microphone to measure the noise from the tailpipe and an engine speed sensor. The system calculates the correct anti-noise and delivers this by means of special speaker drivers mounted on the exhaust system. The residual noise is measured and adjustments can be made. Because the system is self-learning it will adapt to the changing noises of an ageing engine.

The active muffler allows straight gas flow from the exhaust after the catalytic converter. This allows improved engine performance that can mean less fuel is used. An average reduction in fuel consumption of 5% is possible. Future EC directives are expected relating to exhaust noise, which are currently set at

77 dB (A) in Germany. Larger mufflers will be needed to comply, which means this system may well become quite popular.

16.7.4 Auto dimming mirrors

An automatic dimming function is achieved by the use of electro-chromatic mirror glass. A control unit and forwards and backwards directed light sensors are used. The forward facing light sensor monitors the ambient light level at the front of the vehicle; the rearward facing light sensor monitors the light level coming from the rear of the vehicle. If the light falling in from the rear is higher than that from the front, the interior rear view mirror is automatically dimmed. The control for the automatic dimming mechanism is integrated in the mirror housing.

A reverse gear signal is used so that the dimming function is deactivated when reverse gear is selected. To do this the GEM supplies the reverse gear signal input to the interior rear view mirror. Figure 16.55 shows the components associated with the auto-dimming mirror.

Key fact

If the light falling in from the rear is higher than that from the front, the interior rear view mirror is automatically dimmed.

16.7.5 Automatic parking system

Valeo have produced an automatic parking system known as Park4U® (Figure 16.56). It is used by several manufacturers. The second generation has been available since spring 2009 in different Volkswagen models. A gap 55 cm to the front and rear of the vehicle are enough for the system to park, which is reduction of over 20% in comparison to the first generation. This is all made possible through a more complex geometrical calculation of the approach vector, which can now take into account multiple reversing manoeuvres. Therefore, the steering assistance is no longer ended after the first reverse manoeuvre, but aids the driver until the vehicle is finally parked, regardless of the number of manoeuvres necessary.

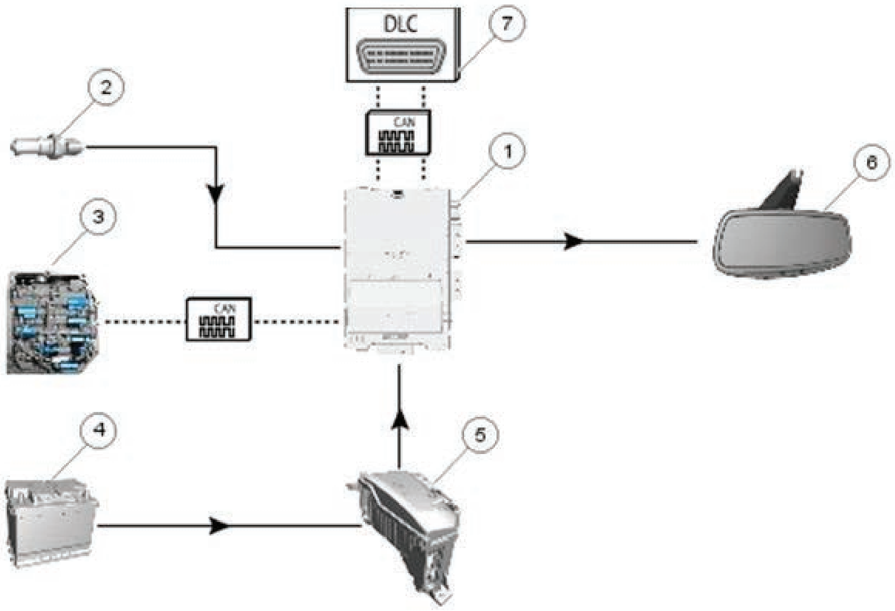


Figure 16.55 Rear view mirror system for auto-dimming: 1, GEM (generic electronic module); 2, Reversing lamp switch; 3, Transmission control module; 4, Battery; 5, Battery junction box; 6, Interior mirror; 7, Data link connector (Source: Ford Motor Company)

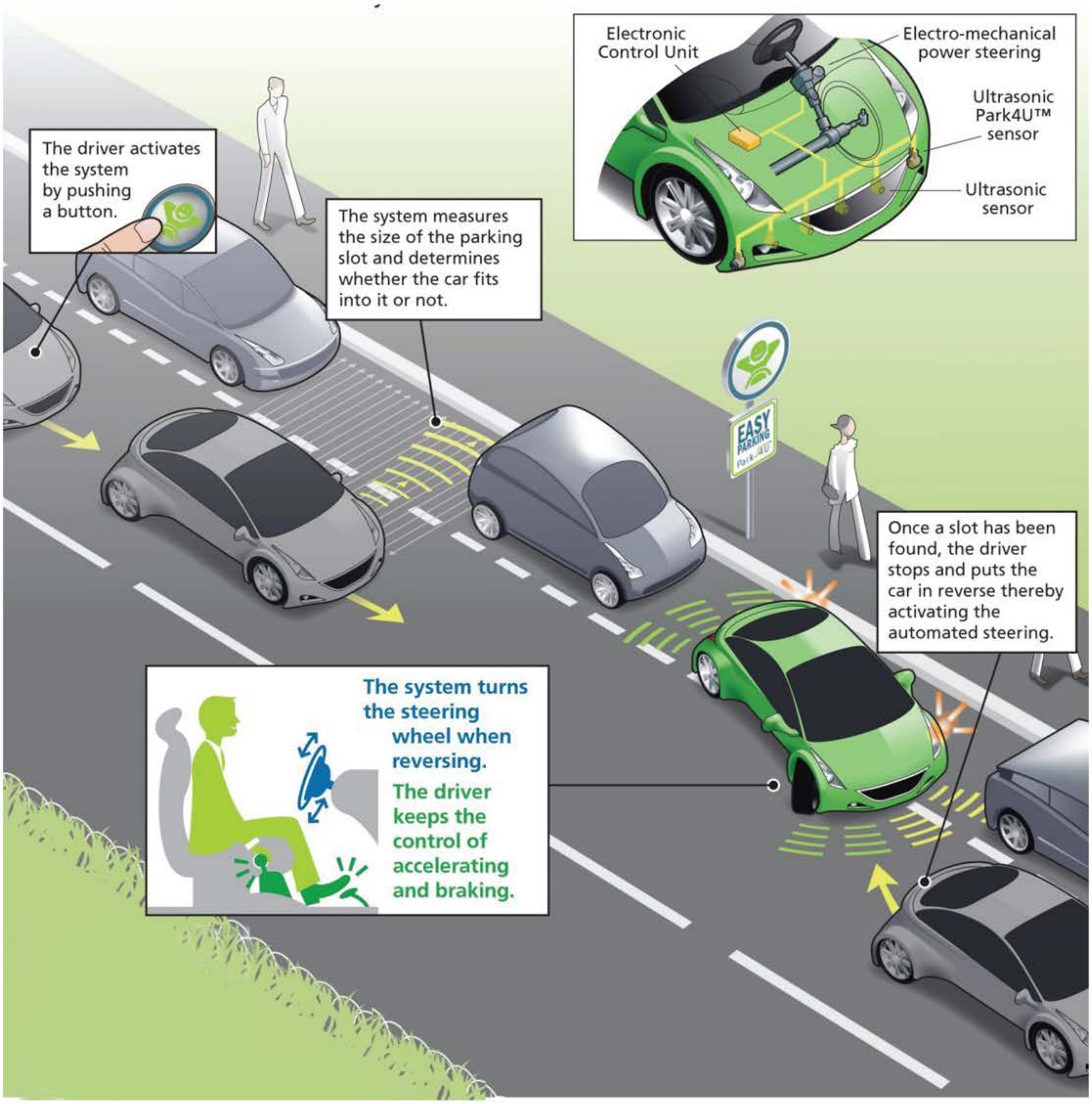


Figure 16.56 Parallel parking assistance (Source: Valeo Media)

Two ultrasonic sensors installed on either side of the car scan the edge of the road and detect suitable spaces. The parking manoeuvre itself is now hands free. As soon as the car has stopped and the reverse gear has been engaged, the system takes over the steering. The driver continues to control the speed of the vehicle with the accelerator and brake. The ultra-sonic sensors in front and to the rear give the driver additional security and help him or her to use the available space as efficient as possible. The driver can end the manoeuvre at any time by touching the steering wheel, the system automatically deactivates.

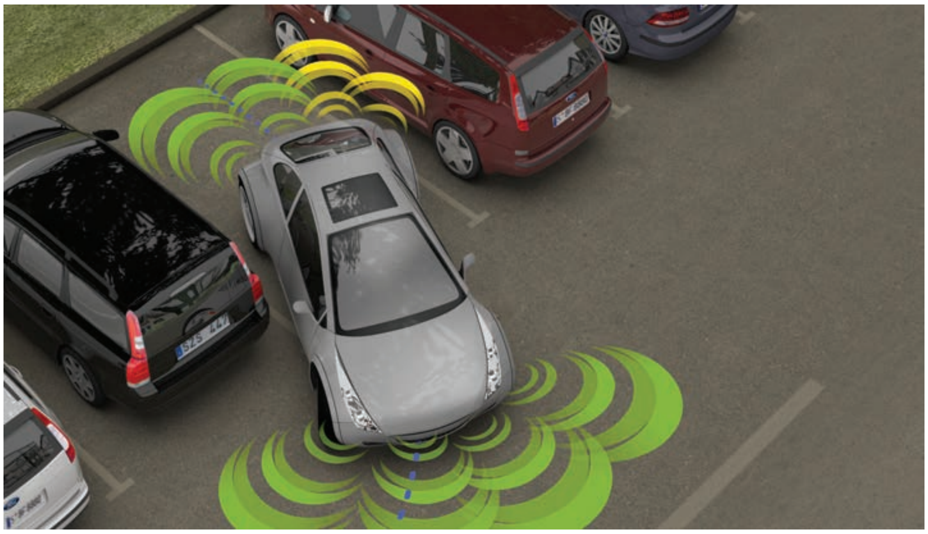


Figure 16.57 Perpendicular parking assistance (Source: Valeo Media)

The system can also help the driver with perpendicular parking, as well as in exiting a parking space. The ‘Exit Assistant’ will measure the space to the front and rear of the vehicle and determine the best strategy for exiting the parking space. While the driver controls the vehicle’s speed, the system takes over the steering, just as it does in parking. The system recognizes when the space can be exited and is automatically deactivated on leaving the space, so that the driver can merge with the traffic.

The parking assistant for perpendicular parking spaces offers the function for reversing into a perpendicular space (Figure 16.57). The space is detected and measured as the vehicle drives past, the approach vector is calculated and the vehicle is steered into the space automatically, while the driver operates the pedals.

The measurement of the parking spaces takes place on both sides of the car simultaneously. The driver chooses the side they want to park on using the indicator. The sensors to the side detect the start and end of the space by registering the contours of the bumpers on the cars which border the space. The kerb is also registered and taken into account when calculating the parking manoeuvre. If there is no kerb, the vehicle is parked in line with the car in front of it.

Information from other systems in the vehicle are used and accessed over the CAN network. In order not to inconvenience any following traffic, the vehicle can drive at speeds of up to 30 kmh while scanning for a park. The possible distance to the parked cars is a realistic 50 to 150 cm. Reliable functionality is guaranteed at parking speeds of up to 7 kmh. If this speed is exceeded, the system deactivates for safety reasons. This also occurs if the driver takes hold of the steering wheel, indicating that they wish to finish parking manually.

16.7.6 General systems diagnostic procedure

The following procedure is very generic but with a little adaptation can be applied to any electrical system. Refer to the manufacturer’s recommendations if in any doubt. The process of checking any system circuit is broadly as follows.

1. Hand and eye checks (loose wires, loose switches and other obvious faults) – all connections clean and tight.

2. Check battery (see Chapter 5) – must be 70% charged.
3. Check motor/solenoid/linkage/bulbs/unit – visual check.
4. Fuse continuity – (do not trust your eyes) voltage at both sides with a meter or a test lamp.
5. If used, does the relay click (if yes, jump to stage 8) – this means the relay has operated, but it is not necessarily making contact.
6. Supply to switch – battery volts.
7. Supply from the switch – battery volts.
8. Supplies to relay – battery volts.
9. Feed out of the relay – battery volts.
10. Voltage supply to the motor – within 0.5 V of the battery.
11. Earth circuit (continuity or voltage) – 0 Ω or 0 V.

16.8 Advanced comfort and safety systems technology

16.8.1 Cruise control and system response

Figure 16.58 shows a block diagram of a cruise control ECU. Many cruise control systems work by the proportional-integral control technique. Proportional control means that an error signal is developed via the feedback loop, which is proportional to the difference between the required and actual outputs. The final output of a cruise control system is the vehicle speed but this depends on the throttle position, which is controlled by the actuator. The system electronics must take into account the lag between throttle movement and the required change in vehicle speed.

If the system overreacts, then the vehicle speed would become too high and then an over-reaction would cause the speed to become too low and so on. In other words, the system is not damped correctly (under damped) and will oscillate, much like a suspension spring without a damper. Proportional control alone is prone to this problem because of steady-state errors in the system. To improve on this, good system design will also include integral control. Thus, the final signal will be the sum of proportional and integral control signals. An integral controller produces a signal, which is a ramp, increasing or decreasing, proportional to the original error signal.

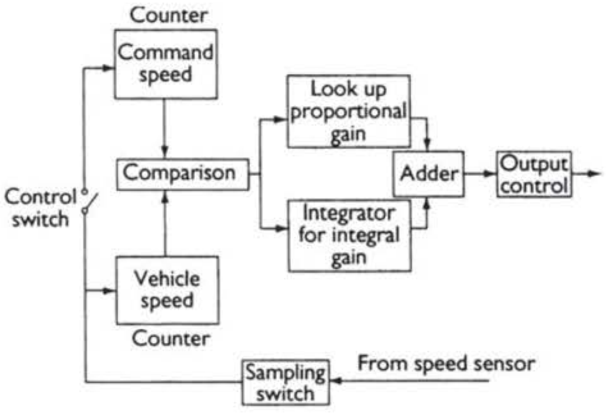


Figure 16.58 Cruise control systems – detailed block diagram

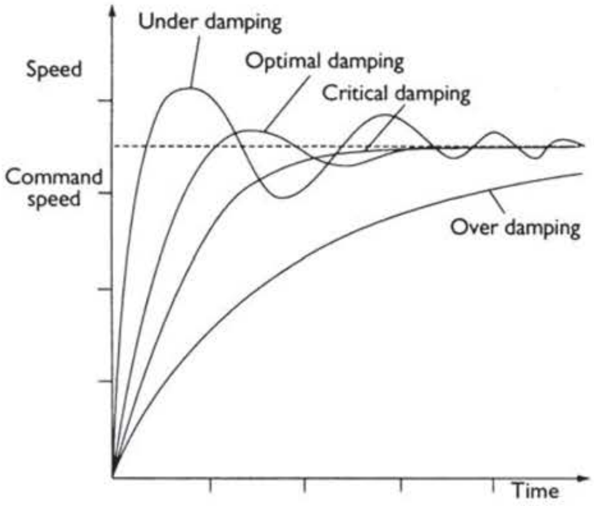


Figure 16.59 Damping factors

The use of integral control causes the final error signal to tend towards zero. The combination therefore of these two forms of control in the weighting given to each determines the damping factor of the control electronics. Figure 16.59 shows the effect on vehicle speed of different damping factors. These four responses are well known in engineering and electronics and can be modelled by mathematics to calculate the response of a system.

The above technique can be based on analogue or digital electronics. The principle is much the same in that for any system the proportional and integral control can be used. The theoretical values can be calculated prior to circuit design as follows:

$$G_i = \omega_n^2 M$$

$$G_p = (2d\omega_n M) - C$$

where: G_i *= integral gain,* G_p *= proportional gain,* ω_n *= natural frequency of the system* ($2\pi f_n$)*,* M *= mass of the vehicle,* C *= experimentally determined frictional factor (mechanical),* d *= damping coefficient.*

16.8.2 Radio suppression calculations

Capacitors and inductors are used to act as filters. This is achieved by using the changing value of 'resistance' to alternating signals as the frequency increases. The correct term for this resistance is either capacitive or inductive reactance. These can be calculated as follows:

$$X_C = \frac{1}{2\pi f C}$$

$$X_L = 2\pi f L$$

where: X_C *= capacitive reactance (Ohms),* X_L *= inductive reactance (Ohms),* C *= capacitance (Farads),* L *= inductance (Henrys),* f *= frequency of the interference (Hertz).*

Using the above formulae gives the following results with a 0.1 mF capacitor and a 300 mH inductor first at 50 Hz and then at 1 MHz.

By choosing suitable values of capacitor in parallel and or inductor in series it is possible to filter out unwanted signals of certain frequencies. To home in on a

Table 16.1 Inductive and capacitive reactance

Frequency	100 Hz	1 MHz
Capacitive reactance	15.5 k Ω	1.6 Ω
Inductive reactance	0.18 Ω	1.9 kΩ

specific or resonant frequency a combination of a capacitor and inductor can be used. The resonant frequency of this combination can be calculated:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

When the range of interference frequency is known suitable values of components can be determined to filter out its effect.

CHAPTER 17

Alternative fuel, hybrid and electric vehicles

17.1 Alternative fuels

17.1.1 Overview

The use of an alternative fuel can lessen dependence upon oil and reduce greenhouse gas emissions. There are a number of alternative fuels and each of these is outlined briefly in this section.

17.1.2 Fuels

Ethanol is an alcohol-based fuel made by fermenting and distilling starch crops, such as corn. It can also be made from plants such as trees and grasses. E10 is a blend of 10% ethanol and 90% petrol/gasoline. Almost all manufacturers approve the use of E10 in their vehicles. E85 is a blend of 85% ethanol and 15% petrol/gasoline and can be used in flexible fuel vehicles (FFVs). FFVs are specially designed to run on petrol/gasoline, E85, or any mixture of the two. These vehicles are offered by several manufacturers.

Key fact

Ethanol is an alcohol-based fuel made by fermenting and distilling starch crops, such as corn.

There is no noticeable difference in vehicle performance when E85 is used. However, FFVs operating on E85 usually experience a 20–30% drop in miles per gallon due to ethanol's lower energy content.

Figure 17.1 NASCAR flex-fuel (Source: Ford Media)

Figure 17.2 E85 vehicle (Source: Ford Media)

There are some advantages and disadvantages of using ethanol:

Advantages

- Lower emissions of air pollutants.
- More resistant to engine knock.
- Added vehicle cost is very small.

Disadvantages

- Can only be used in flex-fuel vehicles.
- Lower energy content, resulting in fewer miles per gallon.
- Limited availability.

Key fact

Biodiesel is a form of diesel fuel manufactured from vegetable oils, animal fats, or recycled restaurant oils.

Biodiesel is a form of diesel fuel manufactured from vegetable oils, animal fats, or recycled restaurant oils. It is safe, biodegradable, and produces less air pollutants than petroleum-based diesel. It can be used in its pure form (B100) or blended with petroleum diesel. Common blends include B2 (2% biodiesel), B5, and B20. B2 and B5 can be used safely in most diesel engines. However, most vehicle manufacturers do not recommend using blends greater than B5, and engine damage caused by higher blends is not covered by some manufacturer warranties.

Listed here are some advantages and disadvantages of biodiesel:

Advantages

- Can be used in most diesel engines, especially newer ones (Figure 17.3).
- Less air pollutants (other than NOx) and less greenhouse gases.
- Biodegradable.
- Non-toxic.
- Safer to handle.

Disadvantages

- Use of blends above B5 may not yet be approved by manufacturers.
- Lower fuel economy and power (10% lower for B100, 2% for B20).
- More nitrogen oxide emissions.
- B100 generally not suitable for use in low temperatures.
- Concerns about B100's impact on engine durability.

Figure 17.3 Biodiesel engine

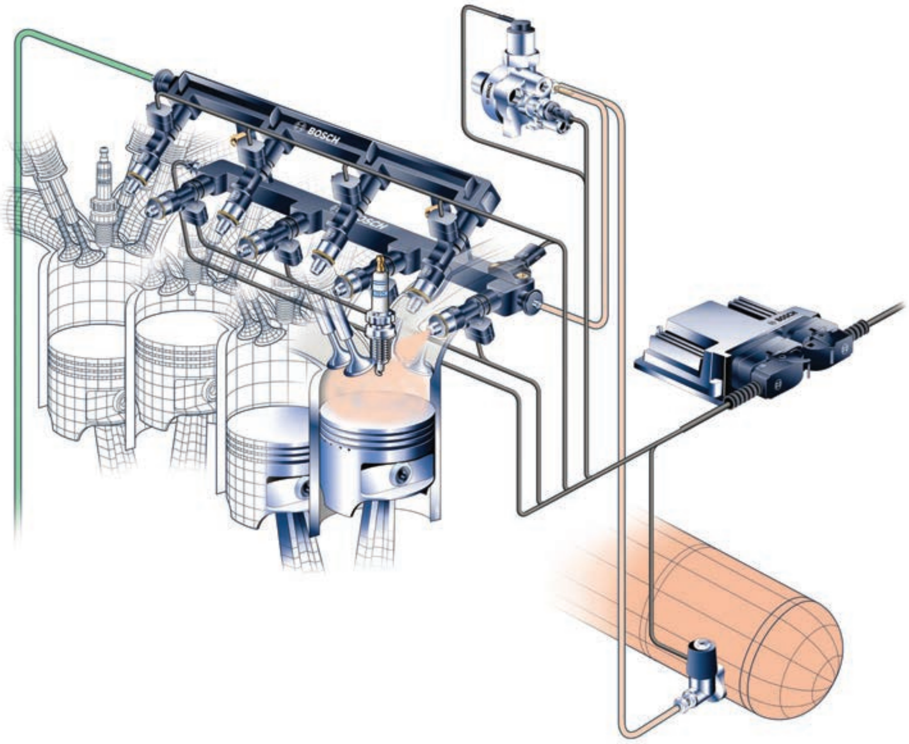


Figure 17.4 Natural gas and GDi system (Source: Bosch Media)

Natural gas is a fossil fuel made up mostly of methane. It is one of the cleanest burning alternative fuels (Figures 17.4 and 17.5). It can be used in the form of compressed natural gas (CNG) or liquefied natural gas (LNG) to fuel cars and trucks. Dedicated natural gas vehicles are designed to run on natural gas only, while dual-fuel or bi-fuel vehicles can also run on petrol/gasoline or diesel. Dual-fuel vehicles take advantage of the widespread availability of conventional fuels but use a cleaner, more economical alternative when natural gas is available. Natural gas is stored in high-pressure fuel tanks so dual-fuel vehicles require two separate fuelling systems, which take up extra space. Natural gas vehicles are not produced commercially in large numbers. However, conventional vehicles can be retrofitted for CNG (Figure 17.7).

Key fact

Natural gas is a fossil fuel made up mostly of methane.

Some advantages and disadvantages of this alternative fuel are noted here:

Advantages

- 60–90% less smog-producing pollutants.
- 30–40% less greenhouse gas emissions.
- Less expensive than petroleum fuels.

Disadvantages

- Limited vehicle availability.
- Less readily available.
- Fewer miles on a tank of fuel.

Propane or liquefied petroleum gas (LPG) is a clean-burning fossil fuel that can be used to power internal combustion engines. LPG-fuelled vehicles produce fewer toxic and smog-forming air pollutants. Petrol/gasoline and diesel vehicles can be retrofitted to run on LPG in addition to conventional fuel. The LPG is stored in high-pressure fuel tanks, so separate fuel systems are needed in vehicles powered by both LPG and a conventional fuel (Figure 17.6).

Key fact

Propane or liquefied petroleum gas (LPG) is a clean-burning fossil fuel.

Figure 17.5 The NGI2 injects CNG (compressed natural gas) into a spark-ignition engine's cylinders. Compared with petrol/gasoline, the combustion of natural gas produces 25% less carbon dioxide (CO_2)

Figure 17.6 Propane tank (Source: http://www.rasoenterprises.com)

Following are advantages and disadvantages of LPG:

Advantages

- Fewer toxic and smog-forming air pollutants.
- Less expensive than petrol/gasoline.

Disadvantages

- No new passenger cars or trucks commercially available but vehicles can be retrofitted for LPG.
- Less readily available than conventional fuels.
- Fewer miles on a tank of fuel.

Hydrogen (H_2) can be produced from fossil fuels (such as coal or natural gas), nuclear power, or renewable resources, such as hydropower. Fuel cell vehicles powered by pure hydrogen emit no harmful air pollutants. It is being aggressively explored as a fuel for passenger vehicles. It can be used in fuel cells to power electric motors or burned in internal combustion engines. Hydrogen is an environmentally friendly fuel that has the potential to dramatically reduce dependence on oil, but several significant challenges must be overcome before it can be widely used.

There are advantages and disadvantages to using hydrogen:

Advantages

- Can be produced from several sources, reducing dependence on petroleum.
- No air pollutants or greenhouse gases when used in fuel cells.
- It produces only NOx when burned in internal combustion engines.

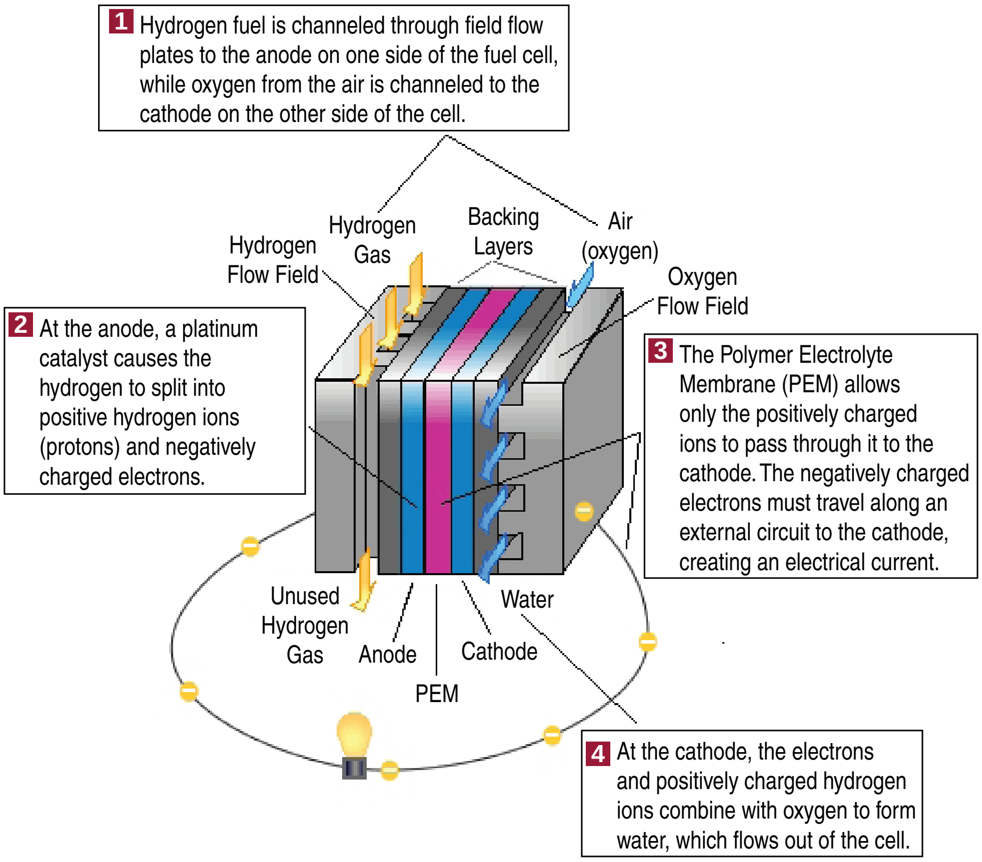
Disadvantages

- Expensive to produce and is only available at a few locations.
- Fuel cell vehicles are currently too expensive for most consumers.
- Hydrogen has a lower energy density than conventional petroleum fuels. For this reason it is difficult to store enough hydrogen on a vehicle to travel more than 200 miles.

This section has given an overview of some alternative fuels. All of them offer some significant advantages either commercially, environmentally or both. There are also some disadvantages not least of which is that the cost of production is high. This is likely to change however as their use becomes more widespread.

Key fact

Hydrogen (H_2) can be produced from fossil fuels (such as coal), nuclear power, or renewable resources, such as hydropower.



1 Hydrogen fuel is channeled through field flow plates to the anode on one side of the fuel cell, while oxygen from the air is channeled to the cathode on the other side of the cell.

2 At the anode, a platinum catalyst causes the hydrogen to split into positive hydrogen ions (protons) and negatively charged electrons.

3 The Polymer Electrolyte Membrane (PEM) allows only the positively charged ions to pass through it to the cathode. The negatively charged electrons must travel along an external circuit to the cathode, creating an electrical current.

4 At the cathode, the electrons and positively charged hydrogen ions combine with oxygen to form water, which flows out of the cell.

Figure 17.7 Fuel cell operation

Figure 17.8 First ethanol vehicle in the UK (Source: Ford Media)

17.2 Electric vehicles (EVs)

17.2.1 Introduction

EVs have come of age and there are now some really good vehicles on general sale. See the sections later in the chapter on Tesla Motors, GM, and Honda as examples.

In 1990, General Motors announced that its EV, the ‘Impact’, could accelerate to 100 km/h in just 8 s, had a top speed of 160 km/h (100 mile/h) and had a range of 240 km between charges. Running costs were about double the fossil-fuel equivalent but this was falling. The car was a totally new design with drag-reducing tyres and brakes, which when engaged, act as generators (regenerative braking). The car was powered by a 397 kg array of advanced gel electrolyte lead-acid batteries (32 at 10 V) and two small AC electric motors to drive the front wheels. The recharging time was about 2 hours but this could be reduced to 1 hour in an emergency. This was very impressive, but things have moved on further.

The following sections look at some of the EV issues in more detail, but the subject of ‘electric and hybrid vehicles’ could (and does) fill many books in its own right. This chapter is presented as an introduction to a technology that is certain to become a major part of the general motor trade.

The case studies are presented to outline the ‘state of the art’ as well as to show how developments are taking place very quickly.

17.2.2 Electric drive system

Figure 17.9 shows the general layout in block diagram form of an electric vehicle (EV). Note that the drive batteries are often a few hundred volts, so a lower 12/24 V system is still required for ‘normal’ lighting and other systems.

17.2.3 EV batteries

A number of options are available when designing the electric car but, at the risk of over-simplification, the most important choice is the type of batteries.

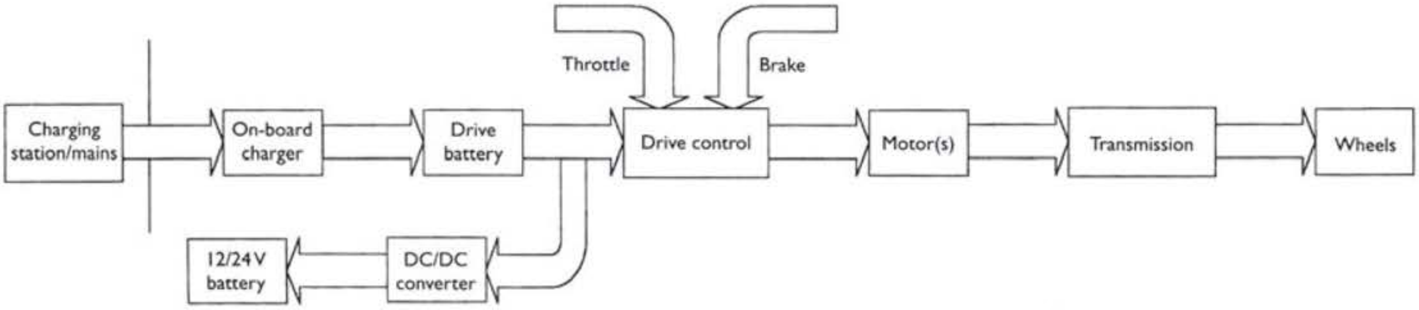


Figure 17.9 General electric vehicle (EV) layout

Table 5.6 (in Chapter 5) summarizes the current choice relating to batteries and will allow some comparisons to be made. Further details relating to some of these and other battery developments can also be found in Chapter 5.

17.2.4 Drive motors

There are several choices of the type of drive motor. The first being between an AC or DC motor. The AC motor offers many control advantages but requires the DC produced by the batteries to be converted using an inverter. A DC shunt wound motor rated at about 50 kW is a popular choice for the smaller vehicles but AC motors are now the most popular. Actually the distinction is blurred. The drive motors can be classed as AC or DC but it becomes difficult to describe the distinctions between and AC motor and a brushless DC motor.

Key fact

The AC motor offers many control advantages but requires the DC produced by the batteries to be converted using an inverter.

AC motors

In general, all AC motors work on the same principle. A three-phase winding is distributed round a laminated stator and sets up a rotating magnetic field that the rotor 'follows'. The general term is and AC induction motor. The speed of this rotating field and hence the rotor can be calculated:

$$n = 60\frac{f}{p}$$

where: n = speed in rpm; f = frequency of the supply; and p = number of pole pairs.

Asynchronous motor

The asynchronous motor is often used with a squirrel cage rotor made up of a number of pole pairs. The stator is usually three-phase and can be star or delta wound. This is shown in Figure 17.10. The rotating magnetic field in the stator induces an EMF in the rotor which, because it is a complete circuit, causes current to flow. This creates magnetism, which reacts to the original field caused by the stator, and hence the rotor rotates. The amount of slip (difference in rotor and field speed) is about 5% when the motor is at its most efficient.

Synchronous with permanent excitation

This motor has a wound rotor, known as the inductor. This winding is magnetized by a DC supply via two slip rings. The magnetism 'locks on' to the rotating magnetic field and produces a constant torque. If the speed is less than *n* (see above), fluctuating torque occurs and high current can flow. This motor needs special arrangements for starting rotation. An advantage, however, is that it makes an ideal generator. The normal vehicle alternator is very similar. Figure 17.11 shows a representation of the synchronous motor.

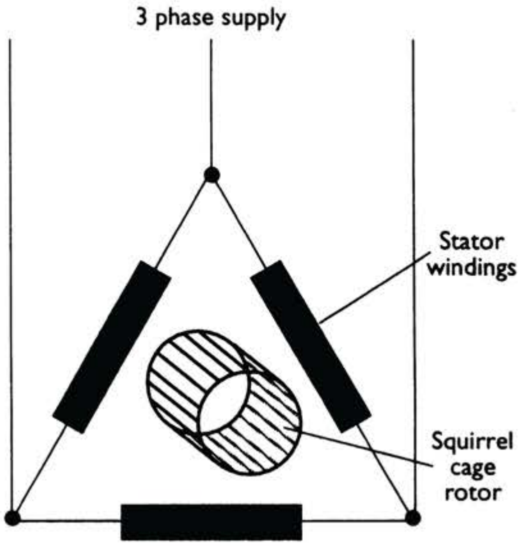


Figure 17.10 An asynchronous motor is used with a squirrel cage rotor made up of a number of pole pairs

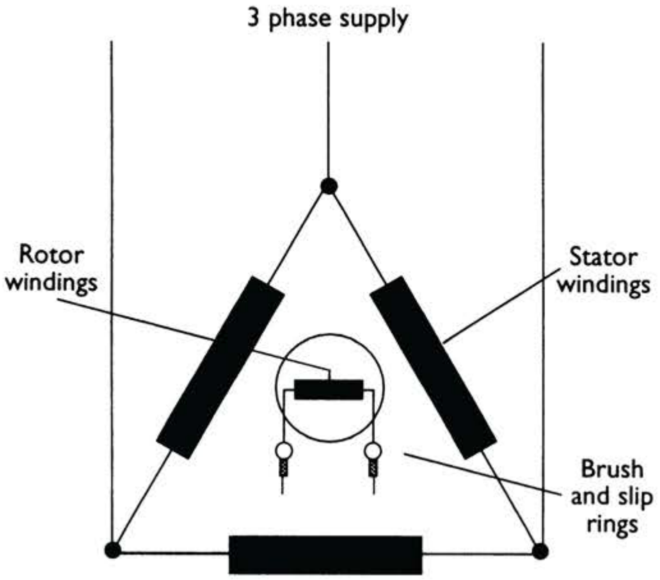


Figure 17.11 Representation of the synchronous motor

EC motors (electronically controlled)

The EC motor is, in effect, half way between an AC and a DC motor. Figure 17.12 shows a representation of this system. Its principle is very similar to the synchronous motor above except the rotor contains permanent magnets and hence no slip rings. It is sometimes known as a brushless motor. The rotor operates a sensor, which provides feedback to the control and power electronics. This control system produces a rotating field, the frequency of which determines motor speed. When used as a drive motor, a gearbox is needed to ensure sufficient speed of the motor is maintained because of its particular torque characteristics. Some schools of thought suggest that if the motor is supplied with square-wave pulses it is DC, and if supplied with sine wave pulses then it is AC. This leaves a problem describing motors supplied with trapezoidal signals!

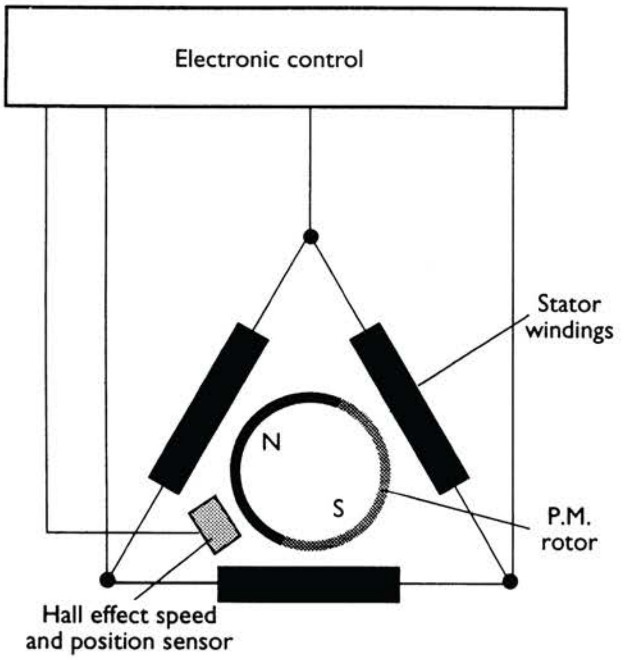


Figure 17.12 The EC motor is, in a way, halfway between an AC and a DC motor

DC motor – series wound

The DC motor is a well-proven device and has been used for many years on electric vehicles such as milk floats and fork lift trucks. Its main disadvantage is that the high current has to flow through the brushes and commutator.

The DC series wound motor has well-known properties of high torque at low speeds. Figure 17.13 shows how a series wound motor can be controlled using a thyristor and also provide simple regenerative braking.

Key fact

The DC series wound motor has well-known properties of high torque at low speeds.

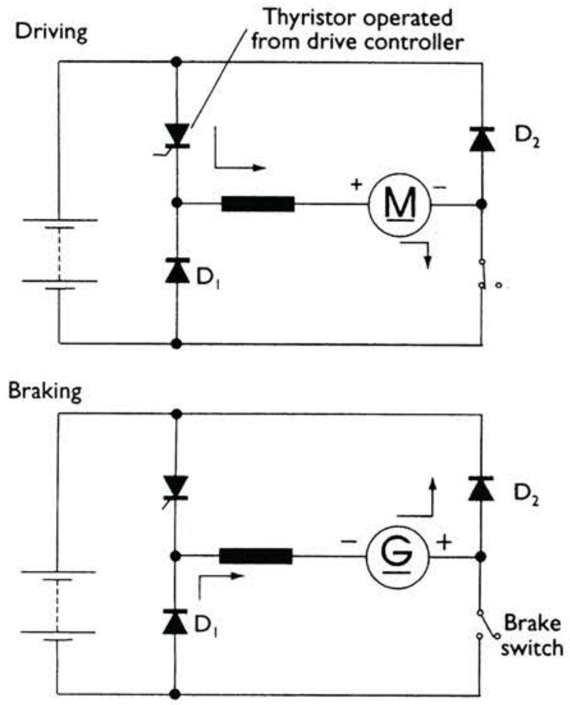


Figure 17.13 A series wound motor can be controlled by using a thyristor and can also provide simple regenerative braking

DC motor – separately excited shunt wound

The fields can be controlled either by adding a resistance or using chopper control in order to vary the speed. Start-up torque can be a problem but, with a suitable controller, can be overcome. This motor is also suitable for regenerative braking by increasing field strength at the appropriate time. Some EV drive systems only vary the field power for normal driving and this can be a problem at slow speeds due to high current.

17.2.5 General Motors EV-1

General Motors has arguably led the motor industry in electric vehicle development since the 1960s and made a major commitment of nearly half a billion dollars to its Impact and PrEView electric vehicle development programmes. As a direct result of these initiatives, GM developed the EV-1 electric car as the world's first specifically designed production electric vehicle, and it became the first to go on sale (in the USA) in 1996. The EV-1 is shown in Figure 17.14.

Key fact

GM developed the EV-1 electric car as the world's first specifically designed production electric vehicle.

The EV-1 has a drag coefficient of just 0.19 and an aluminium spaceframe chassis (40% lighter than steel) with composite body panels. Weighing just 1350 kg in total, the car has an electronically regulated top speed of 128 km/h (80 mile/h) – although a prototype EV-1 actually held the world land-speed record for electric vehicles at 293 km/h (183 mile/h)! The current record is now over 300 mile/h but in a car designed for record breaking – not for normal use.

The EV-1 can reach 96 km/h (60 mile/h) from a standing start in less than 9 s. The key to the success of the EV-1 is its electrical powertrain, based on a 103 kW (137 HP) three-phase AC induction motor with an integral, single-speed, dual-reduction gear-set driving the front wheels. The unit requires no routine maintenance for over 160 000 km (100 000 miles).

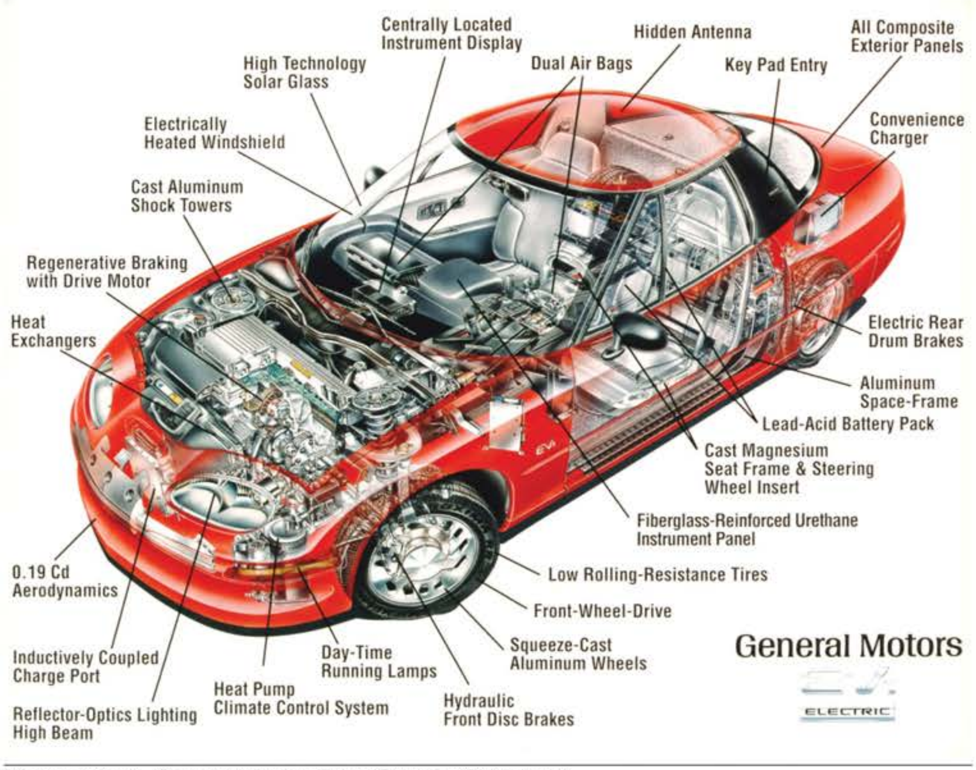


Figure 17.14 General Motors EV-1 (Source: GM Media)

The battery pack uses 26, 12 V maintenance-free lead-acid batteries, with a total voltage of 312 V. The range was 112 km (70 miles) per charge in urban conditions and 144 km (90 miles) on the open road. However, new nickel-metal-hydride (NiMH) batteries were phased into production during 1998, almost doubling the EV-1's range to 224 km (140 miles) in the city and 252 km (160 miles) on highways. An innovative regenerative braking system helps to extend that range still further by converting the energy used when braking back into electricity in order to recharge the battery pack partially.

Full recharging can be carried out safely in all weather conditions and takes 3–4 hours using a 220 V standard charger or 15 hours using the on-board 110 V convenience charger. Compared with normal fossil fuels, the lower cost of domestic electricity means operating costs are relatively low.

Regenerative braking is accomplished by using a blended combination of front hydraulic disc and rear electrically applied drum brakes and the electric propulsion motor. During braking, the electric motor generates electricity (regenerative) which is then used to partially recharge the battery pack.

The EV-1 came with traction control, cruise control, anti-lock brakes, dual airbags, power windows, door locks and outside mirrors, AM/FM, CD/cassette, tyre inflation monitor system and numerous other features.

Figure 17.15 General Motors EV (Source: GM Media)

Key fact

Regenerative braking is accomplished by using a blended combination of front hydraulic disc and rear electrically applied drum brakes and the electric propulsion motor.

17.2.6 Tesla Roadster

17.2.6.1 Introduction

I have chosen this EV as a case study because of its world-class acceleration, handling, and design. It is a cool sports car that also happens to be an electric car – which is a major step forward in perception. It is also a pure EV in that it uses rechargeable batteries (i.e. not methanol or hydrogen fuel cells).

I am grateful to Tesla Motors, for permission to use their materials.

Safety first

The batteries produce 375 V; do not work on high voltages unless trained.

17.2.6.2 Motor

The Roadster is powered by a three-phase AC induction motor. Small, but strong, the motor weighs just over 52 kg (115 lbs). The batteries produce 375 V to push up to 900 A of current into the motor to create magnetic fields. It delivers 288 peak hp and 400 Nm (295 lbs-ft) of torque at the driver's command. At top speed, the motor is spinning at 14 000 rpm.

Key fact

The Roadster is powered by a three-phase AC induction motor.

Figure 17.16 Tesla Roadster – available in several colours including racing green! (Source: Tesla Motors)

Figure 17.17 Tesla's AC induction motor (Source: Tesla Motors)

The motor is directly coupled to a single speed gearbox, above the rear axle. The simplicity of a single gear ratio reduces weight and eliminates the need for complicated shifting and clutch work. The elegant motor does not need a complicated reverse gear – the motor simply spins in the opposite direction.

The internal combustion (IC) engine is a complex, amazing machine. Unfortunately, this complexity results in wasted energy. At best, only about 30% of the energy stored in fuel is converted to forward motion. The rest is wasted as heat and noise. When the engine is not spinning, there is no torque available. In fact, the engine must turn at several hundred rpm (idle speed) before it can generate enough power to overcome its own internal losses.

An IC engine does not develop peak torque until many thousand rpm. Once peak torque is reached, it starts to drop-off quickly. To overcome this narrow torque

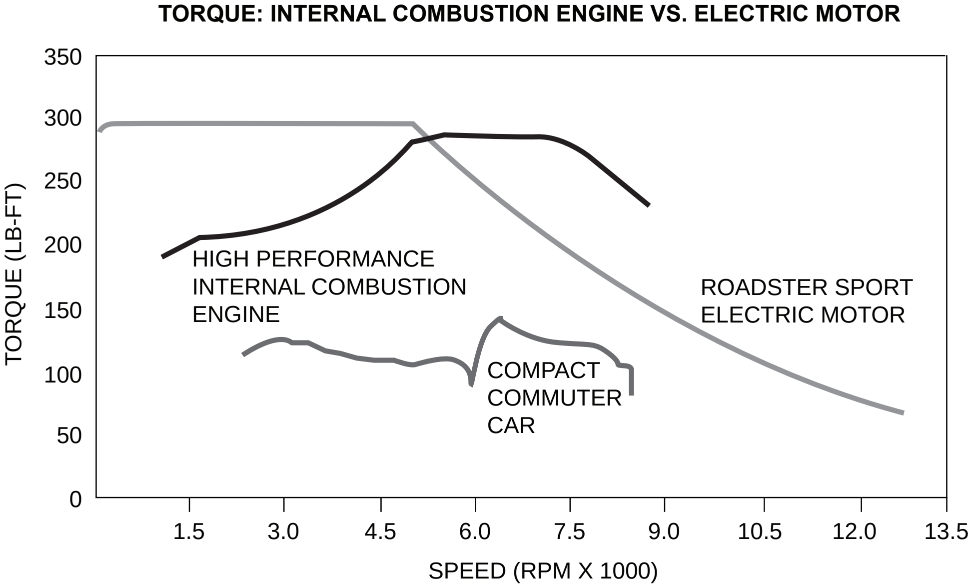


Figure 17.18 IC engine torque vs. and electric motor

range, multi-speed transmissions are employed to create gear ratios that keep the engine spinning where it is most effective.

IC engine power output can be improved with faster rotation. However, combustion engines have a limit to how fast they can spin – as engine speed exceeds 5000 or 6000 rpm it becomes challenging and costly to keep the timing of the engine on track and keep all of the parts together.

Electric motors are by comparison very simple. The motor converts electricity into mechanical power and also acts as a generator, turning mechanical power into electricity. Compared to the myriad parts in an engine, the Roadster motor has only one moving component – the rotor. The spinning rotor eliminates conversion of linear motion to rotational motion and has no mechanical timing issues to overcome.

With an electric motor, instant torque is available at any speed. The entire rotational force of the motor is available the instant the accelerator is pressed. Peak torque stays constant to almost 6000 rpm, only then does it start to slowly roll-off (Figure 17.18).

Key fact

With an electric motor, instant torque is available at any speed.

The wide torque band, particularly the torque available at low speed, eliminates the need for gears – the Roadster has only a single speed gear reduction; one gear ratio from zero to top speed. Switch two of the phases (this can be done electronically), and the motor runs in reverse. Not only is this design incredibly simple, reliable, compact, and lightweight, but it allows a unique and exhilarating driving experience. The Roadster accelerates faster than most sports cars.

Tesla's electric motor is also able to create torque efficiently. The Roadster achieves an overall driving efficiency of 88%, about three times the efficiency of a conventional car.

Key fact

The Roadster achieves an overall driving efficiency of 88%, about three times the efficiency of a conventional car.

As driving conditions permit, the motor acts as a generator to recharge the battery. When the accelerator pedal is released, the motor switches to 're-generative braking' mode and captures energy while slowing the car. The experience is similar to 'engine braking' in a conventional car.

The Tesla Roadster uses a three-phase AC induction motor. The induction motor was, appropriately, first patented by Nikola Tesla in 1888. These motors are widely used in industry for their reliability, simplicity, and efficiency.

Key fact

The Tesla Roadster uses a three-phase AC induction motor, first patented by Nikola Tesla in 1888.

The Roadster motor has two primary components: a rotor and a stator. The rotor is a shaft of steel with copper bars running through it. It rotates and, in doing so, turns the wheels. The stationary stator surrounds, but does not touch, the rotor. The stator has two functions: it creates a rotating magnetic field and it induces a current in the rotor. The current creates a second magnetic field in the rotor that chases the rotating stator field. The end result is torque. Some motors use permanent magnets, but not the Roadster motor – the magnetic field is created completely from electricity.

The stator is assembled by winding coils of copper wire through a stack of thin steel plates called laminations. The copper wire conducts the electricity fed into the motor from the power electronics module (PEM). There are three sets of wires – each wire conducts one of the three phases of electricity. The three phases are offset from each other such that combing the rises and falls of each phase creates a smooth supply of current – and therefore power. The flow of alternating current into the copper windings creates an alternating magnetic field. Because of the way the copper coils are placed within the stator, the magnetic field appears to move in a circular path around the stator.

The copper bars mentioned above are 'shorted' to each other (referred to as a 'squirrel cage') which allows current to flow with little resistance from one side of the rotor to the other. The rotor does not have a direct supply of electricity. When a conductor (one of the copper bars) is moved through a magnetic field (created by the AC in the stator), a current is induced.

Because the stator magnetic field is moving the rotor tries to catch up. The interaction of the magnetic fields creates torque. The amount of torque produced is related to the relative position of the rotor field to the rolling 'wave' of magnetism in the stator (the stator field). The further the rotor field is from the 'wave', the more torque is produced. Since the stator field is always ahead of the rotor when the accelerator is depressed, the rotor is always spinning to catch up, and it is continuously producing torque.

When the driver releases the accelerator pedal, the PEM immediately changes the position of the stator field to behind the rotor field. Now, the rotor must slow down to align its field with the stator field. The direction of current in the stator switches direction, and energy starts to flow, through the PEM, back to the battery.

17.2.6.3 Motor control

When the accelerator pedal is pressed, the power electronics module (PEM) interprets a request for torque. Flooring the pedal means a request for 100% of the available torque. Half-way is a request for partial torque and so on. Letting off the accelerator pedal means a request for re-generation. The PEM interprets the accelerator pedal input and sends the appropriate amount of alternating current to the stator. Torque is created in the motor and the car accelerates.

Key fact

The PEM supplies as much as 900 A to the stator.

The PEM supplies as much as 900 A to the stator. To handle such high current levels, the stator coils in a Tesla motor employ significantly more copper than a traditional motor of its size. The copper is tightly packed in a proprietary winding pattern to optimize efficiency and power. The copper loops are encapsulated by special polymers that facilitate heat transfer and ensure reliability under the demands of high-performance driving in extreme conditions.

High stator currents mean high rotor currents. Unlike typical induction motors which employ aluminium for its conductors, the Roadster rotor conductors are made of copper. Copper, while harder to work with, has a much lower resistance and can therefore handle higher currents. Special care is taken in the motor design to handle the high speed (14 000 rpm).

Though highly efficient, the motor still generates some heat. To keep within acceptable operating temperatures, specially engineered cooling fins have been integrated into the housing and a fan is employed to blow air across the fins to most effectively extract the heat. This helps keep the overall package light and tight.

17.2.6.4 Battery

Key fact

The pack contains 6831 lithium ion cells and is the most energy dense pack in the industry, storing 56 kWh of energy.

The battery pack in the Tesla Roadster is the result of innovative systems engineering and 20 years of advances in lithium-ion battery technology (Figures 17.19 and 17.20). The pack contains 6831 lithium ion cells and is the most energy dense pack in the industry, storing 56 kWh of energy. It weighs 990 lbs and delivers up to 215 kW of electric power. The car will charge from almost any 120 V or 240 V outlet. Most Roadster owners find they rarely use a complete charge, and charging each night means their car is ready to drive 245 miles each morning.

To achieve the required energy density, Tesla starts with thousands of lithium-ion cells and assembles them into a liquid-cooled battery pack, wrapped in a strong metal enclosure. The battery is optimized for performance, safety, longevity, and cost. With lithium-ion chemistry, there is no need to drain the battery before

Figure 17.19 Battery pack in production (Source: Tesla Motors)

recharging as there is no memory effect. Roadster owners simply top-up the charge each night.

The cells used in a Roadster battery pack are referred to as 18650 form-factor, because of their measurements: 18 mm in diameter by 65 mm length. Tesla uses versions of this form factor modified for use in EVs. The small cell size enables efficient heat transfer, allows for precise charge management, improves reliability, and extends battery pack life. Each cell is enclosed in a steel case which effectively transfers heat away from the cell. The small size makes the cell essentially isothermal, and its large surface area allows it to shed heat to the ambient environment.

Sixty-nine cells are wired in parallel to create bricks. Ninety-nine bricks are connected in series to create sheets, and 11 sheets are inserted into the pack casing. In total, this creates a pack made up of 6831 cells. Appropriate cell temperature levels are maintained by a proprietary liquid-cooling system which includes sensors within the pack monitored by the car's firmware. Liquid coolant is pumped through the pack to enable effective heat transfer to and from each cell. The cooling system is so effective that the cells on opposite sides of the battery pack stay within a few degrees of each other. This is important for maximizing battery life, optimizing performance, and guaranteeing safety.

Key fact

With lithium-ion chemistry, there is no need to drain the battery before recharging as there is no memory effect.

Definition

Isothermal: A thermodynamic process in which the temperature of the system remains constant. The heat transfer into or out of the system is at such a slow rate that thermal equilibrium is maintained.

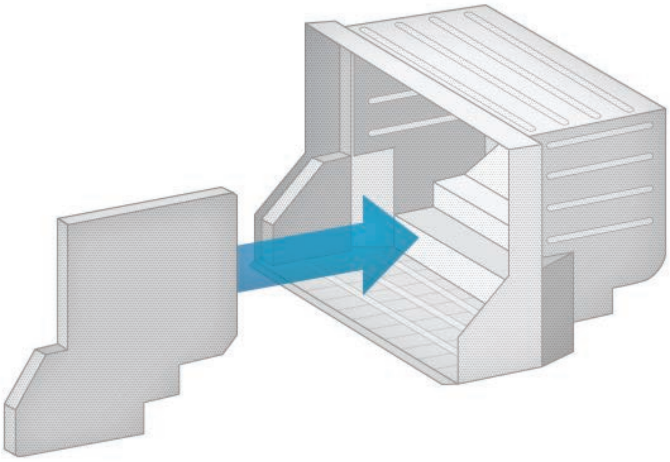


Figure 17.20 Construction of the battery pack

Key fact

Air bag deployment causes the high-voltage circuits in the vehicle to immediately shut down.

The Roadster's high-voltage systems are protected against accidental contact outside their protective enclosures and jacketed cables. Only with special tools may one gain access to the high-voltage components. In the event of significant impact or rollover, the high-voltage supply is automatically disconnected inside the pack to reduce risk of exposure to high voltage. Air bag deployment causes the high-voltage circuits in the vehicle to immediately shut down. The high-voltage systems are enclosed, labelled, and colour-coded with markings that service technicians and emergency responders are trained to recognize.

The pack enclosure is designed to withstand substantial abuse in the vehicle, while maintaining the integrity of the internal components. The pack is a stressed member of the chassis and helps provide rigidity to the rear of the car. The Roadster has been tested in standard frontal, rear and side impact crash tests. In general, lithium-ion cells cannot be charged below 0°C, which would theoretically prevent charging in cold environments. To overcome this cold weather charging obstacle, the Roadster is designed with a heater to warm the cells (when plugged in) to an appropriate charging temperature. If there were no battery pack heater, drivers living in cold environments would have difficulty charging and experience stunted driving performance.

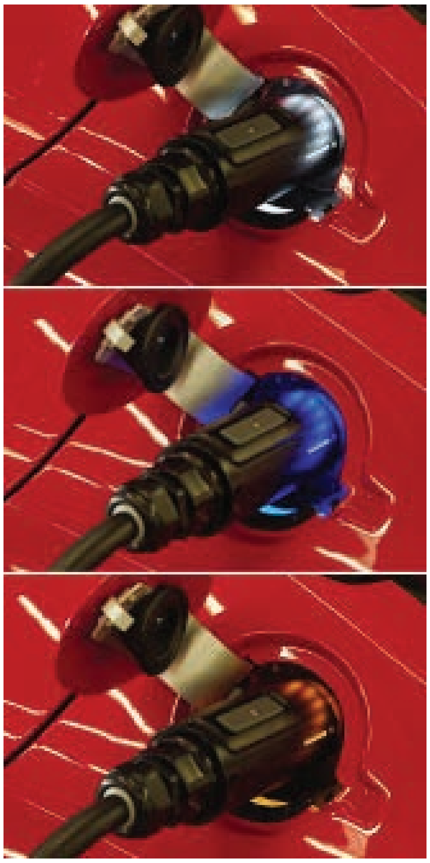


Figure 17.21 Charging port and coloured indicator

Likewise, the cells are designed to operate in high-temperature environments. High-performance driving is possible in even the hottest environments of the world. If the temperature rises above a set threshold, the air conditioning unit sends chilled coolant through the pack. Similar to the radiator fan of IC engined cars, the chilled coolant continues to circulate after the motor has been turned off to keep the temperature at an appropriate level. Cooling the pack enables a driver to quickly charge immediately after hard driving in hot climates. Without such a cooling system, recharging in hot weather would be delayed after each drive.

The battery charger is located on-board the car. This means the Roadster can be plugged into any outlet, anywhere in the world. Charge times vary based on the outlet voltage and amperage. With the Tesla high power wall connector, a Roadster charges in as little as 4 hours from empty. Most owners however simply charge overnight.

The PEM processor manages charging. When the charge port door is opened, charging systems come online and begin coordinating with the vehicle management system (VMS). The PEM is instructed to send current to the light ring in the charge port and white LEDs turn on. When the driver attaches the connector and slides the pilot switch closed, the lights turn blue. Once the connector is attached, the PEM processor detects the current available from the wall and the VMS checks if the driver had previously set charge preferences for the location.

When the car determines which level of current to use, the contactors between the motor and battery begin to close in a series of audible clicks. This function ensures that the high voltage from the wall does not flow to the battery until every connection is properly mated. The battery processor, PEM processor, and VMS work together to flash the LEDs in the light ring – if the battery is empty, the lights flash quickly. As the battery charges, the rate of flashing slows. When the battery is completely charged, the lights turn green. The battery processors help preserve battery life if the car is plugged in for long periods of time by checking the state of charge every 24 hours, topping up the battery as needed to maintain a healthy state of charge.

The PEM processor also controls the Roadster when in drive mode. The processor monitors the accelerator pedal and uses the information to control

Figure 17.22 Power electronics module (Source: Tesla Motors)

current to the motor. To ensure that generated torque is appropriate for the state of other components in the car, many other processors monitor the status of the car and send outside requirements to the PEM. For example, if the battery processor and VMS calculate that the battery is full, regenerative torque is reduced; if the PEM processor detects that the motor has exceeded an ideal temperature, current to the motor is reduced.

17.2.6.5 Power control

The power electronics module (PEM) functions as a bridge between the charge port, battery, and motor. It manages and converts current during driving and charging. As AC flows into the car from the wall (anything from 90 to 265 V between 50 and 60 Hz), the unit converts it to DC for storage in the battery. When driving, the module converts DC back to AC that the motor uses to generate torque. At many operating points, it is 97–98% efficient: less than 2% of converted energy is lost.

The voltage to the motor is varied by turning switches called IGBTs on and off very quickly. As the IGBTs allow more current from the battery to the motor, the AC waveforms grow in amplitude until peak torque is produced in the motor. In the video below, watch how the AC intensifies as the Roadster accelerates from 0 to 60 mph.

Definition

IGBT: Insulated gate bipolar transistor.

In drive mode, the PEM responds to information from the accelerator pedal, motor speed sensor, ABS speed sensors, and other powertrain sensors. It determines requested torque from the pedal position and monitors the ABS speed sensors to detect if tyres are slipping. Based on sensor feedback, it produces torque by converting the DC voltage stored in the battery to the appropriate AC voltage at the motor terminals. As the driver steps on the accelerator pedal, the PEM begins to control increasing motor current and voltage to produce the torque required to accelerate from 0 to 60 mph in just 3.7 seconds.

Key fact

The Roadster can accelerate from 0 to 60 mph in just 3.7 seconds.

Inside the PEM, there are three major systems:

1. power stages;
2. a controller;
3. a line filter.

The most complex is the power stages, called megapoles. These are large semiconductor switch arrays that connect the charge port or motor to the battery

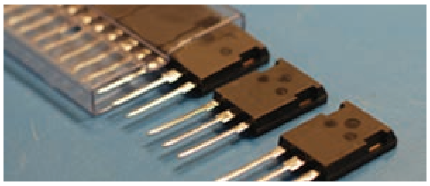


Figure 17.23 Insulated gate bipolar transistors (Source: Tesla Motors)

depending on if the car is charging or driving. Within the megapoles, there are six different switches, grouped into three pairs known as half bridges. In drive mode, each bridge forms a phase. Each phase connects to a phase of the three-phase AC induction motor. In charge mode, only two bridges are required, one for each wire in the AC line. The charge and drive modes are configured using a set of four large relays known as contactors. The contactors allow the semiconductor switches to be used to connect the battery to either the charge port or the motor. When the Roadster is turned on, a series of clicking sounds can be heard as the contactors close the connection to the motor.

Each switch is composed of 14 insulated gate bipolar transistors (IGBTs). In total, 84 IGBTs are used in the PEM. Each IGBT is less than 25 mm square and about a 6 mm thick. Inside the IGBT package is a small piece of silicon, about the thickness of a few sheets of paper and a 6 mm per side.

The second major component of the PEM is the controller board that turns the switches on and off. The switches can turn on and off up to thirty-two thousand times per second. The controller contains two processors: the main DSP and a secondary safety processor. The DSP controls torque, charge behaviours and interprets requests from the VMS. The safety processor monitors the accelerator pedal and the motor current to detect unexpected behaviours. If the safety processor measures motor current inconsistent with accelerator pedal position, it can stop the system. While this behaviour is extremely unlikely, this redundancy means a glitch in the main DSP can't produce unexpected torque.

The third major component of the PEM is the charge input filter. When the car is charging and the IGBTs are switching at 32 kHz, a large amount of electrical noise is created on the AC side of the power stages. If the noise was allowed to conduct back into the power lines, it could interfere with other appliances, radios, cell phones, etc. A group of large inductors called chokes are placed between the IGBTs and the charge port to filter out the noise and avoid unwanted interference.

Tesla power control enables a traction control system with amazing improvements over systems in internal combustion cars. Traction control systems for IC vehicles have a few options to maintain traction at prescribed levels: kill engine spark, reduce fuel supply, use electronic throttle control to actively modulate throttle requests or apply the brakes. Fundamentally, it is practically impossible to maintain near-zero output torque from an IC, whereas zero torque is simple to maintain in an electric drivetrain. In the Roadster, the motor torque can be accurately reduced either gradually or quickly – resulting in better control with less noticeable loss of power. With on-board sensors, the car predicts achievable traction when cornering before the driver can even command a change in acceleration. It's much safer to avoid loss of traction than react to it. Expert test drivers have found they are able to achieve higher performance with the Roadster traction control system than in comparable IC vehicles.

Key fact

In the Roadster, the motor torque can be accurately reduced either gradually or quickly – resulting in better traction control.

17.2.6.6 Software

The Roadster is controlled by state-of-the-art vehicle software. The code is developed in-house with an intense focus on constant innovation. The system monitors the status of components throughout the car, shares information to coordinate action, and reacts to changing external conditions (Figure 17.24).

The Roadster employs many processors to control functions the driver often takes for granted, from battery voltage management and motor control to diagnostics, door locks and touchscreen interaction. Many different operating systems and programming languages are used to optimize each processor for

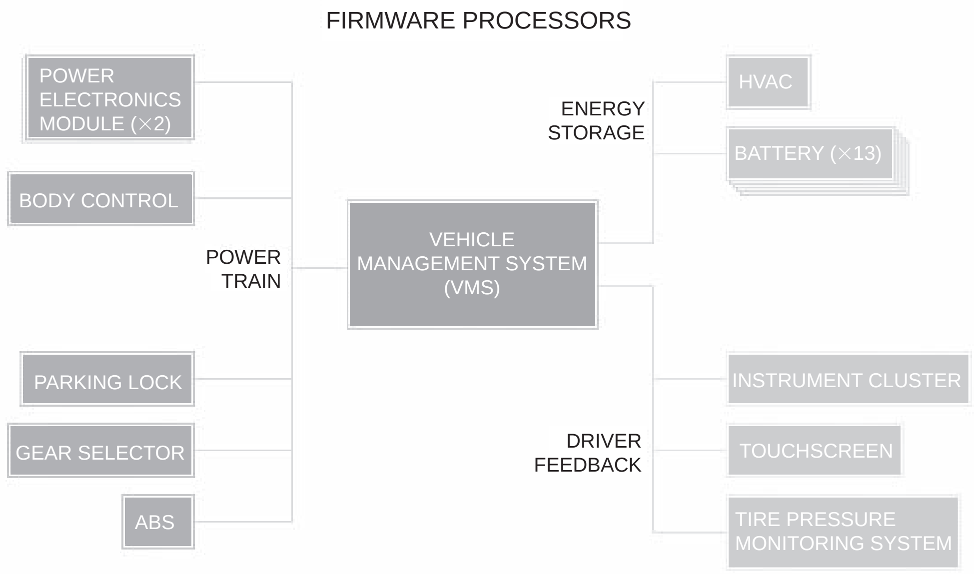


Figure 17.24 Firmware processors

completing its designated function. The processors work together to monitor the status of components throughout the car, share information to coordinate action, and react to changing external conditions.

The vehicle management system (VMS) enables functions the driver is most aware of while driving. It manages the security system, opens the doors, communicates warnings (seatbelt, door ajar, etc.), manages the owner's PIN, and initiates a valet (parking) mode.

The VMS compiles information from many of the other processors to coordinate the necessary actions for driving. When the key is inserted into the car, the VMS turns on the touchscreen. When the key is switched to the ON position, it readies the car for driving by instructing other processors to initiate their functions. It computes available range and prepares the PEM to send power to the motor from the battery. The VMS manages the driving modes (performance, standard, or range) and works with the battery processor to charge and discharge appropriately. It computes ideal and actual range using a complex algorithm that considers battery age, capacity, driving style, and energy consumption rate.

Figure 17.25 Central control touch screen (Source: Tesla Motors)

17.2.6.7 Summary

All in all, a super sports car with a range that makes it completely useable under 'normal' conditions. Even the colours are good; personally I prefer the Racing Green version. Dear Tesla Motors, please send me a free car in this colour in return for this publicity…

More information about Tesla Motors from: http://www.teslamotors.com/

17.2.7 Honda FCX Clarity – Case study

17.2.7.1 Introduction

I have picked the Honda FCX Clarity zero-emission hydrogen fuel cell electric vehicle (EV) as a case study because it has been in development for a while, and it is now at a mature technological level. Some innovative techniques are used and the result is a very useable ZEV.

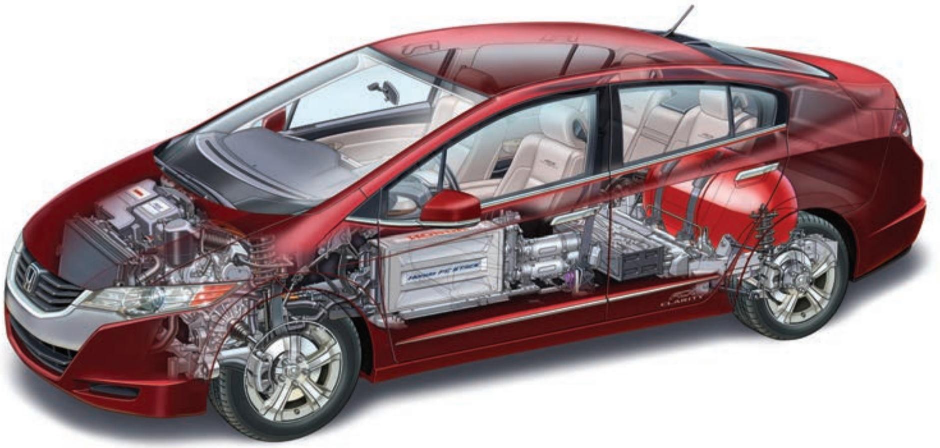


Figure 17.26 Honda FCS Clarity 2011 (Source: Honda Media)

At the time of writing it is available for leasing in the USA, but not in the UK because of a limited refuelling infrastructure. However, this is expected to change soon.

I am grateful to Honda, for permission to use their materials.

To give an overview of the vehicle, some of the main features are as follows:

- Zero harmful emissions – the only exhaust emission is water vapour ($2H_2 + O_2 = 2H_2O$).
- The car generates its electricity on-board using compressed hydrogen as an energy carrier.
- Crash tested and safety tested to the same standard as a conventional car.
- Range equivalent to a conventional petrol or diesel car.
- Improvement in fuel cell stack technology over previous versions – 180 kg lighter, 45% smaller than previous versions.
- Lithium ion battery – 40% lighter and 50% smaller than previous versions.
- Start up at –30°C possible.
- Only one hydrogen tank, reduced from two in the previous model.
- An efficiency rating of around 60% (this is about three times that of a petrol-engined car, twice that of a hybrid vehicle, and 10% better than the previous model).

Key fact

In a fuel cell, hydrogen is combined with atmospheric oxygen to generate electricity.

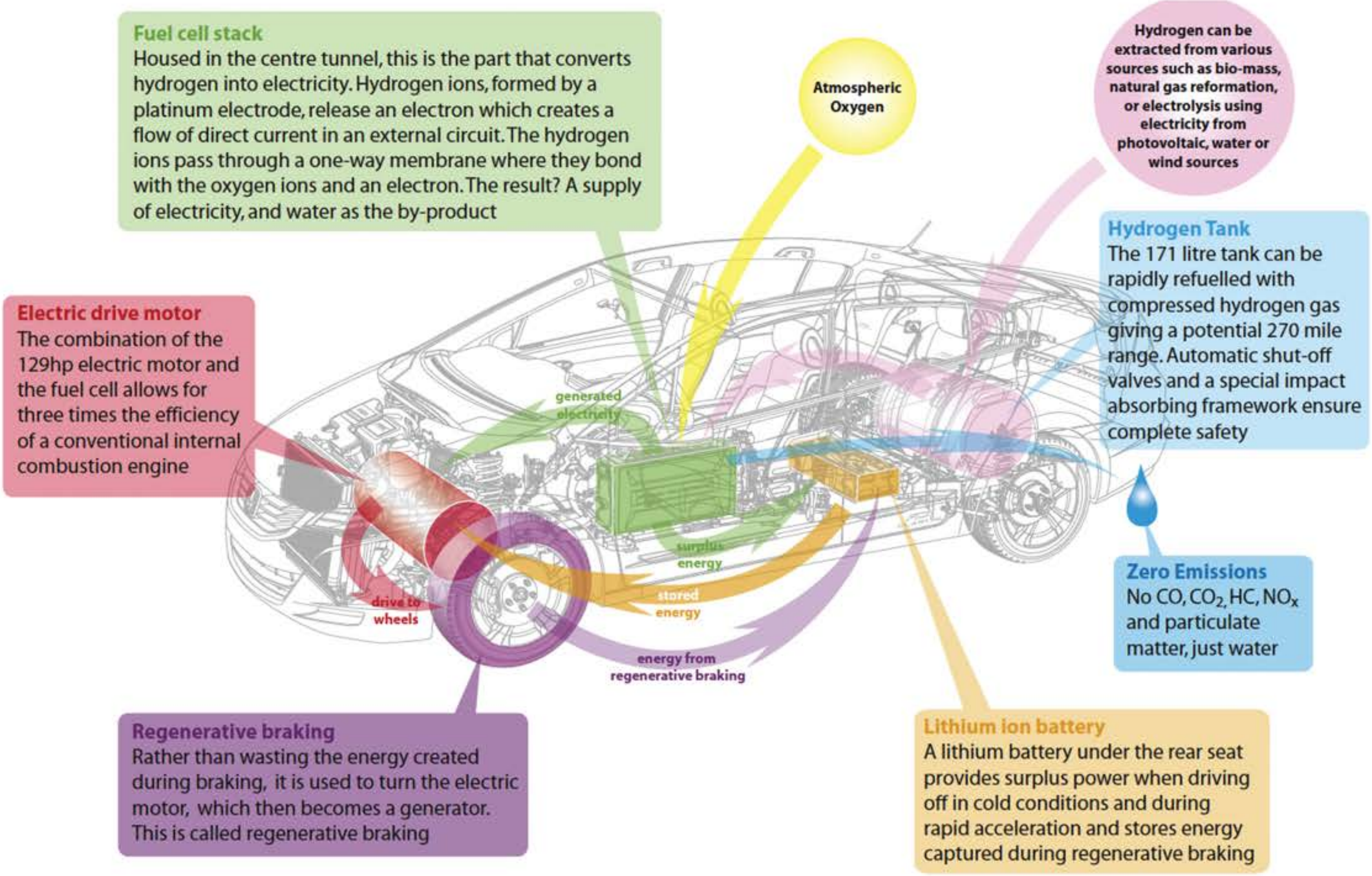
A fuel cell vehicle has a hydrogen tank instead of a petrol/gasoline tank. In the fuel cell, hydrogen is combined with atmospheric oxygen to generate electricity. The fuel cell is really a tiny electric power station, and generates its own electricity on-board rather than through a plug-in system.

Since the electricity required to power the vehicle's motor is generated on-board using hydrogen and oxygen, no CO_2 or other pollutants are emitted in this process. The only emission is the water produced as a by-product of electricity generation.

A compact and efficient lithium ion battery stores electricity generated during braking and deceleration in regenerative braking (just like a mild hybrid). The battery works with the fuel cell stack to power the vehicle.

As well emitting no harmful exhaust gases, fuel cell electric vehicles offer good driving range, short refuelling time and a flexible layout and design:

- Short refuelling time of 3–5 minutes.
- Vehicle range of 270 miles, comparable to that of a conventional car.



Fuel cell stack
Housed in the centre tunnel, this is the part that converts hydrogen into electricity. Hydrogen ions, formed by a platinum electrode, release an electron which creates a flow of direct current in an external circuit. The hydrogen ions pass through a one-way membrane where they bond with the oxygen ions and an electron. The result? A supply of electricity, and water as the by-product

Hydrogen can be extracted from various sources such as bio-mass, natural gas reformation, or electrolysis using electricity from photovoltaic, water or wind sources

Electric drive motor
The combination of the 129hp electric motor and the fuel cell allows for three times the efficiency of a conventional internal combustion engine

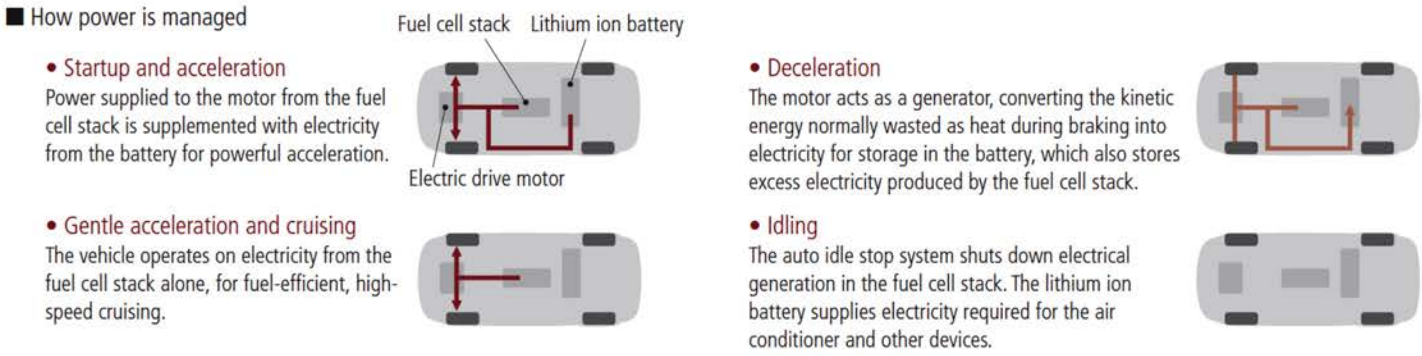
Hydrogen Tank
The 171 litre tank can be rapidly refuelled with compressed hydrogen gas giving a potential 270 mile range. Automatic shut-off valves and a special impact absorbing framework ensure complete safety

Zero Emissions
No CO, CO_2, HC, NO_x and particulate matter, just water

Regenerative braking
Rather than wasting the energy created during braking, it is used to turn the electric motor, which then becomes a generator. This is called regenerative braking

Lithium ion battery
A lithium battery under the rear seat provides surplus power when driving off in cold conditions and during rapid acceleration and stores energy captured during regenerative braking

Figure 17.27 Honda FCX Clarity features and operation (Source: Honda Media)



■ How power is managed

- Startup and acceleration
  Power supplied to the motor from the fuel cell stack is supplemented with electricity from the battery for powerful acceleration.
- Gentle acceleration and cruising
  The vehicle operates on electricity from the fuel cell stack alone, for fuel-efficient, high-speed cruising.
- Deceleration
  The motor acts as a generator, converting the kinetic energy normally wasted as heat during braking into electricity for storage in the battery, which also stores excess electricity produced by the fuel cell stack.
- Idling
  The auto idle stop system shuts down electrical generation in the fuel cell stack. The lithium ion battery supplies electricity required for the air conditioner and other devices.

Figure 17.28 Power management (Source: Honda Media)

- Performance similar to a current mid-size car.
- Zero harmful emissions or pollutants.

17.2.7.2 Hydrogen

Hydrogen can be produced from renewable sources such as solar, wind or hydroelectric power (using electrolysis to extract hydrogen from water). Certain production methods are better suited in different areas of the world, but

Table 17.1 Fuel Cell Electric Vehicle (FCEV), Battery Electric Vehicle (BEV) and Internal Combustion Engine (ICE) compared

	Fuel Cell Electric Vehicle (FCEV)	Battery Electric Vehicle (BEV)	Internal Combustion Engine (ICE)
Time required to refuel	Short	Long	Short
Vehicle range	Long	Short	Long
CO_2 emissions while driving	No emissions	No emissions	Emissions
Renewability of energy source	Renewable	Renewable	Non-renewable
Current common energy source	Steam reforming of natural gas	Coal-fired power stations	Oil

nevertheless it is possible to achieve a stable supply of hydrogen from renewable energy sources.

Currently, the most common way of producing hydrogen is steam reforming from natural gas. There is an environmental cost of extracting hydrogen in this way, but it is the most widely available approach. However, the same issue applies to battery electric vehicles (BEV). There is clearly an environmental cost of a BEV running on electricity made from a coal- or gas-fired power station.

Key fact

Hydrogen is the most abundant element in the universe.

Hydrogen is the most abundant element in the universe, and it is extremely efficient as an energy carrier. These are the key reasons why it is such a suitable fuel for fuel cell cars and clean motoring. Hydrogen can also be produced sustainably using electricity generated from renewable energy sources such as solar, wind and hydroelectric power.

Another advantage of using hydrogen is that it can be compressed or liquefied for delivery via a pipeline or for storage in tanks. Hydrogen can even be manufactured at conventional filling stations.

Honda uses hydrogen as a compressed gas because, in simple terms, more gas will fit in the tank that way. However, the tanks have to be able to cope with the pressure and it does require energy to compress the gas in the first place. Some critics say this compression process reduces the margin on zero-emission driving. However, Honda has made a number of developments in this area to ensure the car is still as efficient as possible.

The high-capacity hydrogen tanks use a newly-developed absorption material to increase the amount of hydrogen they can store. This means it is not necessary to compress it to such a high degree to fit it in the tank, again saving energy at

Table 17.2 Hydrogen production methods

Potential energy sources used in hydrogen production	Production method	Amount of CO_2 released during production	Renewability
Oil/Coal	Gasification and reforming	Large	None
Electricity produced from coal fire power stations	Burning	Large	None
Natural gas	Steam reforming	Medium to small	Limited
Electricity produced from nuclear energy	Water electrolysis	None	None
Electricity produced from solar, wind, hydro	Water electrolysis	None	100%

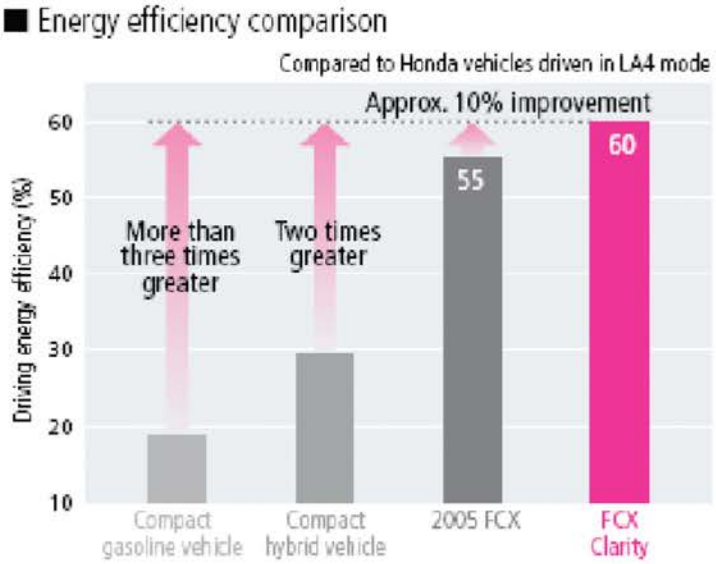


Figure 17.29 Energy efficiency comparison (Source: Honda Media)

the compression stage. The high-capacity hydrogen tanks are so effective that the hydrogen can be compressed to 350 bar, compared to other fuel cell cars that use hydrogen compressed to 750 bar.

17.2.7.3 Energy efficiency and the environment

Because the FCX Clarity has an efficient powerplant and energy management it has an efficiency rating of around 60%. Figure 17.30 shows a comparison between different types of car (Figure 17.29).

Hydrogen is not found on its own in nature, but exists as a component within many different materials from which it can be extracted – like H_2O for instance.

The ideal hydrogen cycle uses renewable energy sources, such as solar, wind or hydro, to extract hydrogen from water via electrolysis. The water produced as a by-product of the fuel cell process would then return to the rivers and oceans before once again being converted into hydrogen via electrolysis.

In a fuel cell, hydrogen is converted into electricity on demand, so just the right amount of electricity is produced. Using hydrogen to create electricity removes the challenge of storing it in large quantities in batteries.

There are two key measures when looking at the environmental cost of producing hydrogen, electricity or indeed any other fuel used for mobility:

Definition

Electrolysis: A method of using a direct electric current (DC) to drive an otherwise non-spontaneous chemical reaction.

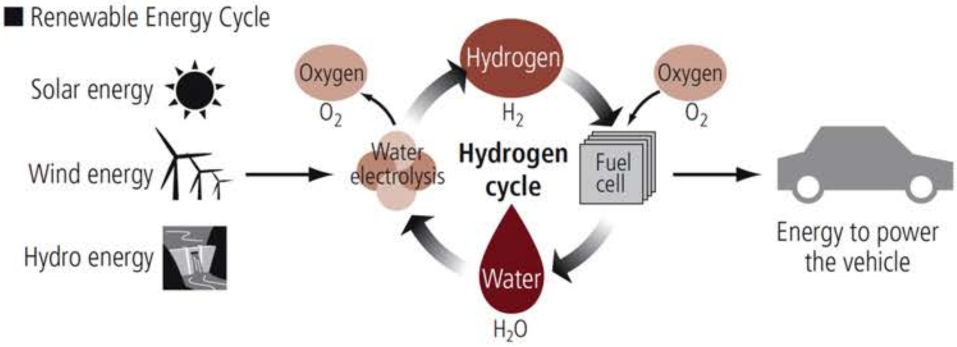


Figure 17.30 Renewable energy cycle (Source: Honda Media)

- Well-to-tank (getting it out of the ground).
- Tank-to-wheel (using the fuel to provide mobility).

In a fuel cell car or a battery-powered electric car there are zero emissions tank-to-wheel, regardless of how the hydrogen or electricity is produced. However, the key question becomes what is the environmental cost of producing the fuel?

Key fact

The most common process for producing hydrogen at present is to steam it out of natural gas.

The most common process for producing hydrogen at present is to steam it out of natural gas. This process is more readily available, and while there is an environmental cost, it is limited. To put it into perspective, if a customer were to run an FCX Clarity (fuelled on hydrogen generated from natural gas) there would still be a 60% reduction in greenhouse gas emissions compared with driving a conventional car. That's a huge benefit over the car technology in widespread use at present.

In addition, the manufacturing of chlorine (for industrial use) generates hydrogen as a by-product. This can be stored in tanks and piped directly to a public fuel station. In Germany alone there's enough hydrogen produced as a by-product of chemical processes to fuel half a million cars.

Hydrogen can also be produced in a variety of other ways, including using renewable energy sources such as solar, wind or hydroelectric power. For example, Honda already produces ultra-efficient thin film solar cells, which can produce electricity in a sustainable way. That electricity can be used to electrolyse water and extract hydrogen via their solar-powered refuelling station, which reduces the well-to-tank environmental cost even further.

It is the combination of environmental benefit and practicality of a fuel cell vehicle that means it excels at meeting customer demands and requirements – and those benefits are because it uses hydrogen as a fuel.

17.2.7.4 Core technologies

The previous FCX had two hydrogen tanks, but the FCX Clarity has only one. This creates more space for the rear seats and boot. The shut-off valve, regulator, pressure sensor and other components in the refuelling and supply system are integrated into a single in-tank module, reducing the number of parts considerably (Figure 17.31).

The main components of the vehicle's powerplant are the fuel cell stack, the hydrogen tank, the lithium ion battery, the electric drive motor and the power drive unit (PDU), which governs the flow of electricity.

Key fact

The fuel cell stack is a device that uses an electrochemical reaction between hydrogen (H_2) and oxygen (O_2) to convert chemical energy into electrical energy.

At the heart of the system is the fuel cell stack – a device that uses an electrochemical reaction between hydrogen (H_2) and oxygen (O_2) to convert chemical energy into electrical energy. In effect, this is the reverse of the principle of electrolysis, in which an electrical current is used to separate water (H_2O) into hydrogen and oxygen. When supplied with hydrogen and oxygen, the FC stack simultaneously generates electricity and water, with no CO_2 or other harmful emissions.

Definition

PEMFC: Proton exchange membrane fuel cell.

The Honda V Flow FC Stack uses a proton exchange membrane fuel cell (PEMFC) electrical generation system that directly converts chemical energy produced in hydrogen-oxygen reactions into electrical energy. The extremely thin proton exchange membrane (electrolytic membrane) is sandwiched between pairs of electrode layers and diffusion layers (the hydrogen and oxygen electrodes) to form a membrane electrode assembly (MEA). The MEA is enclosed between two separators to form a cell – a single electrical generation unit. Several hundred cells are stacked together to form a fuel cell stack. As with batteries, these individual cells are connected in a series to produce a high voltage.

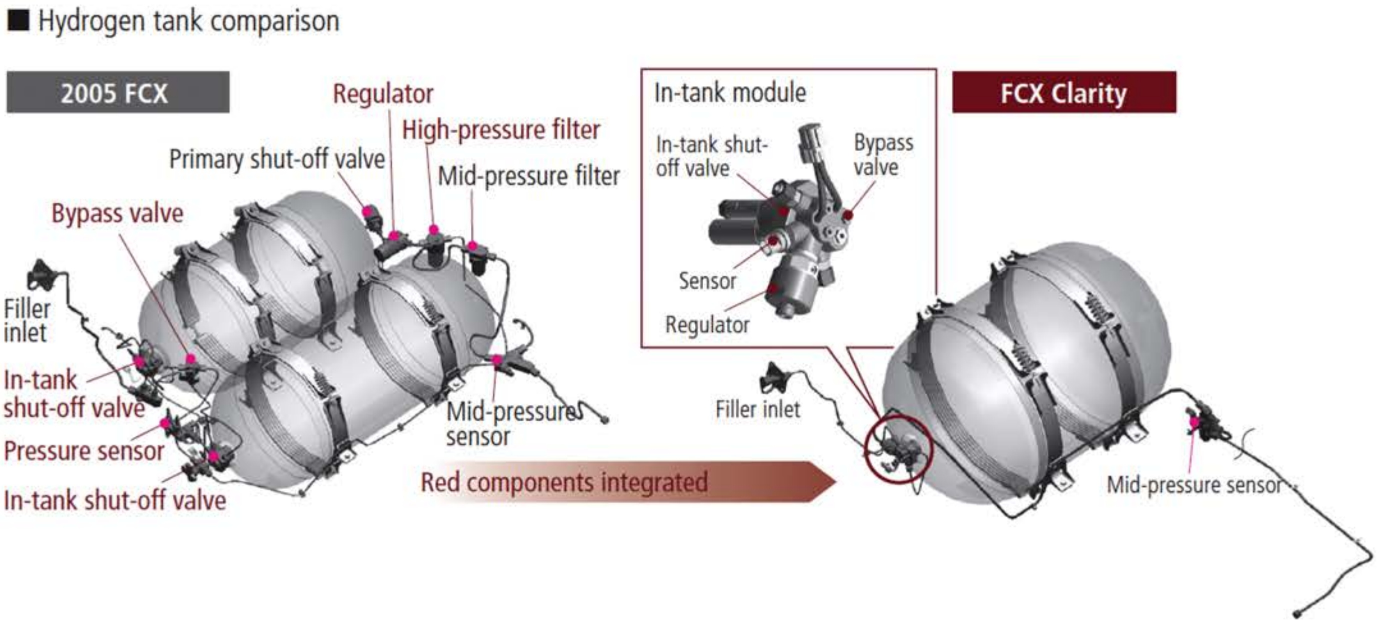


Figure 17.31 hydrogen tank developments (Source: Honda Media)

Hydrogen gas is passed over the hydrogen electrode. Each hydrogen atom is converted into a hydrogen ion in a catalytic reaction with the platinum in the electrode, releasing an electron. Having given up its electron, the hydrogen ion passes through the electrolytic membrane, where it joins with oxygen from the oxygen electrode and an electron arriving via an external circuit.

Figure 17.32 Fuel cell stack (Source: Honda Media)

The released electrons create a flow of direct current in the external circuit. The reaction at the oxygen electrode produces water as a by-product. Because the electrolytic membrane must be kept continually damp, it is necessary to humidify the supply of hydrogen and oxygen. The water by-product is recycled for this purpose. Unneeded water and air are released via the exhaust.

Until more recently, hydrogen and air flowed horizontally through the cells of Honda fuel cell stacks. The new V Flow FC Stack introduces a cell structure in which hydrogen and air flow vertically, and gravity is used to facilitate more efficient drainage of the water by-product from the electricity-generating layer (Figure 17.33).

The result is greater stability in power generation. The new structure also allows for a thinner flow channel and reduction in the stack's size and weight. The innovative and original wave flow-channel separators provide a more even and efficient supply of hydrogen, air and coolant to the electricity-generating layer. The results are higher generating performance, optimal cooling characteristics and major reductions in size and weight (Figure 17.34).

Key fact

The wave flow-channel separators provide a more even and efficient supply of hydrogen, air and coolant to the electricity-generating layer.

Improved water drainage due to the V Flow cell structure helps to achieve better output immediately after start-up. The reduced coolant volume and single-box design made possible by the wave flow-channel separators result in heat mass 40% lower than previous stacks. As a result, the amount of time required to achieve 50% output after start-up at –20°C is only one-quarter that of the previous stack. Start-up is now possible at temperatures as low as –30°C.

The lithium ion battery is 40% lighter and 50% smaller than the ultra-capacitor of the previous FCX, allowing it to be stowed under the rear seat. This gives the

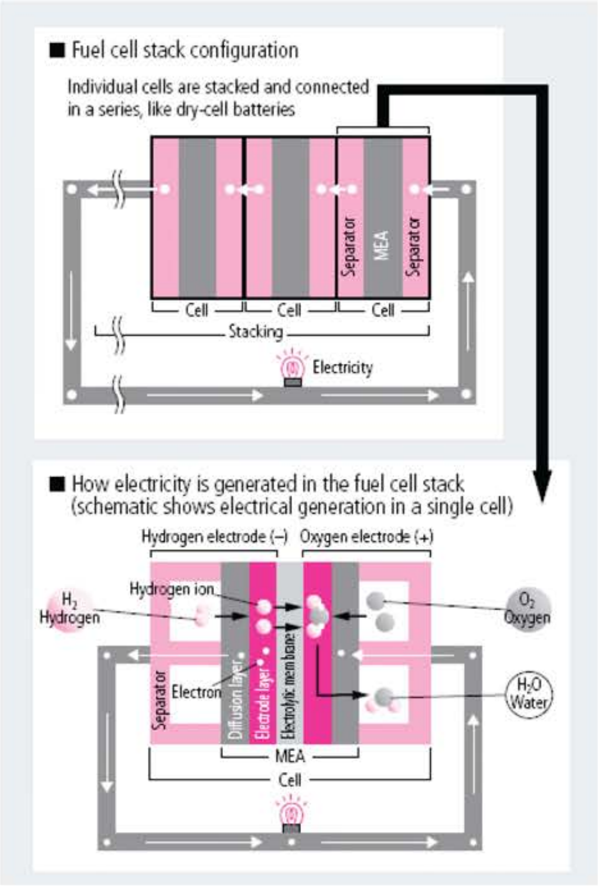


Figure 17.33 Generation of electricity in the fuel cell stack (Source: Honda Media)

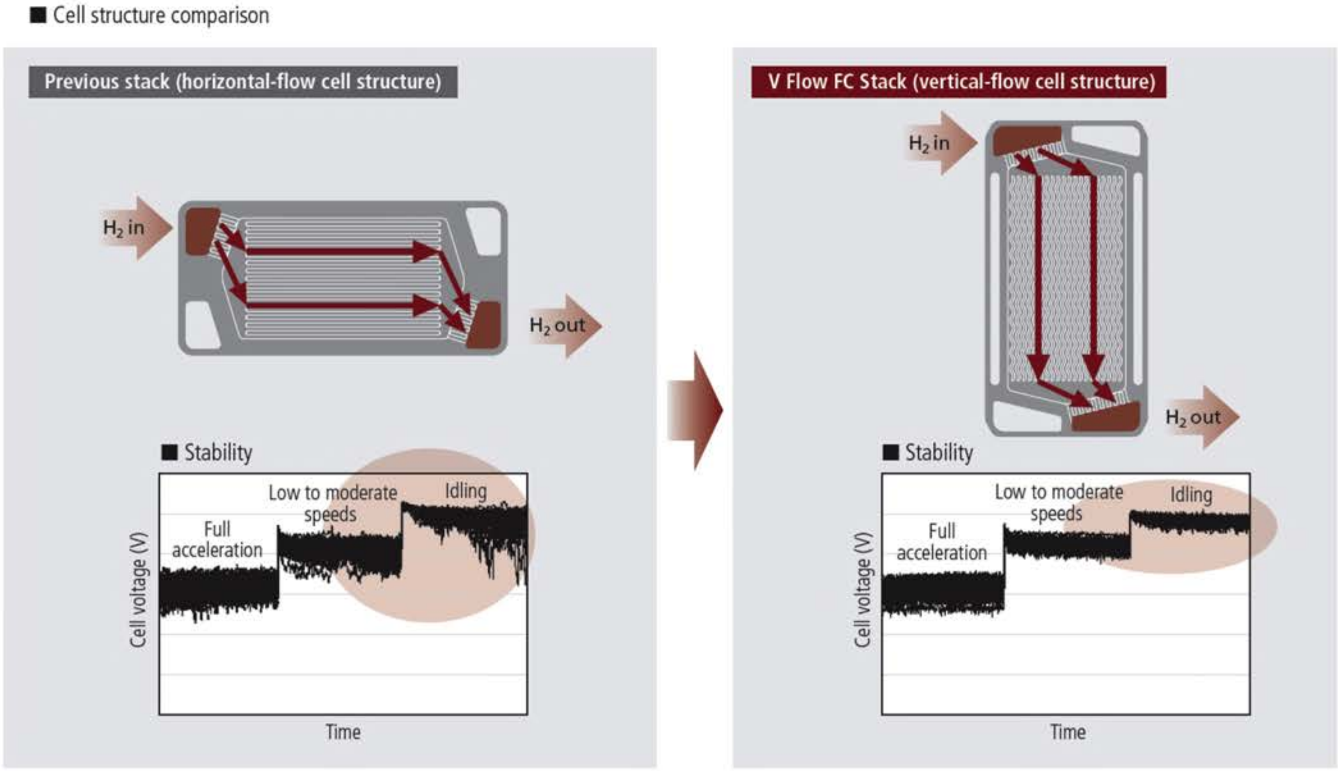


Figure 17.34 Cell structure development (Source: Honda Media)

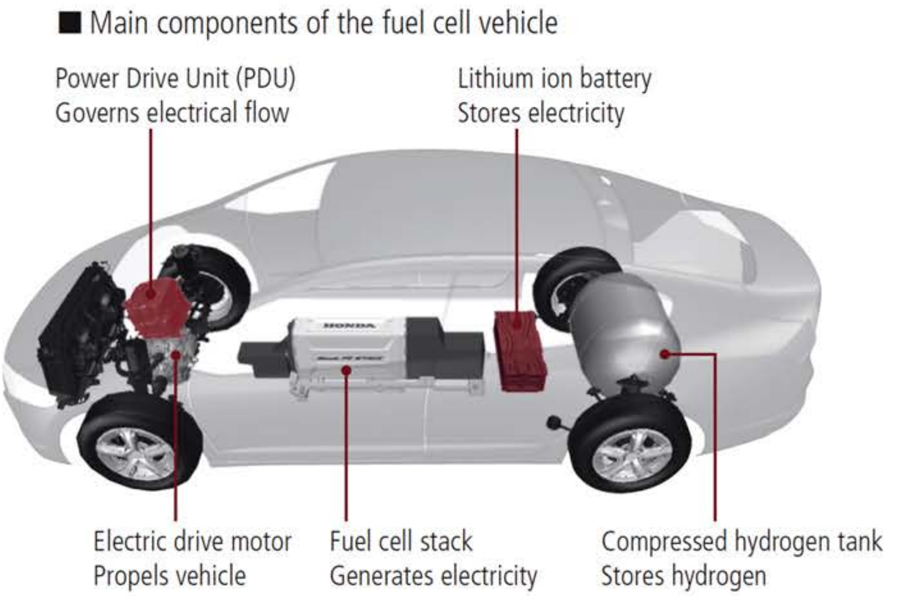


Figure 17.35 Battery and other main components (Source: Honda Media)

car more passenger and boot space. The advanced battery provides a powerful supplement to the fuel cell stack's output, improving the motor torque, for better acceleration. In addition to increasing the total energy capacity, the battery efficiently stores energy generated by the intelligent regenerative braking system, capturing 11% more kinetic energy than the ultra-capacitor used in the 2005 FCX. 57% of the energy of deceleration is regenerated with the new system.

Key fact

The battery provides a supplement to the fuel cell stack's output, improving the motor torque, for better acceleration. It also stores energy from the regenerative braking system.

The drive motor configuration delivers powerful acceleration and a high top speed, along with a quieter, more luxurious ride. The new rotor and stator (stationary permanent magnets) feature a combined reluctance torque, low-loss magnetic circuit and full-range, full-digital vector control to achieve high efficiency and high output over a wide speed range (Figure 17.36).

The innovative shape and layout of the magnets in the rotor result in high-output, high-torque, high-rpm performance. These innovations deliver a maximum

Figure 17.36 Induction drive motor, differential and final drive components (Source: Honda Media)

Key fact

The motor can deliver a maximum output of 100 kW and high torque.

Key fact

The motor rotor uses an interior permanent magnet (IPM) to lower inductance. This improves reluctance torque.

output of 100 kW along with impressive torque and power output density. At the same time, resonance points in the high-frequency range have been eliminated for quieter operation.

A newly-designed rotor features an interior permanent magnet (IPM) to lower inductance, improving reluctance torque for high-torque performance. The magnet's high-energy characteristics also contribute to high torque and a more compact design. These innovations result in 50% higher output density and 20% higher torque density. The number of poles has also been reduced and the magnet widened to better withstand stress, allowing the yoke to wrap around the outside of the IPM. A centre rib has been installed for greater rigidity. This more robust construction allows for operation at higher rpm.

The stator features a low iron-loss electrical steel sheet and higher density windings that decrease resistance and contribute to high torque and higher output. The number of magnetic poles in the rotor has been reduced from 12 to 8, eliminating resonance points within the operating rpm range.

17.2.7.5 Driving dynamics

The electric motor-driven FCX Clarity delivers a completely different driving sensation from a conventional car powered by an internal combustion engine.

There are no gear changes to interrupt power delivery and the torque characteristics are smooth, making acceleration seamless. There is none of the vibration that comes from reciprocating pistons. Acceleration times are around the same as a 2.4-litre petrol or diesel-engined car of a similar size (Figure 17.37).

The vehicle's fixed gear ratio allows for simple operation: there's an easy-to-use shift control for forward, reverse and park that has a light touch and a short stroke. The compact shift unit features electronic control, allowing the shift lever to be installed on the dashboard. The shifter, start switch and parking switch are all easy to operate.

Along with a new brushless motor with increased output, the front double-wishbone suspension helps facilitate tight cornering and delivers a 5.4 m turning radius; very tight given the vehicle's long wheelbase. The low inertia of the motor and minimal friction of the suspension when turning contribute to smoother steering. And a tilt-and-telescopic steering wheel provides an optimal steering position for a wide range of drivers. The FCX Clarity has Adaptive Cruise Control as standard.

The FCX Clarity features an integrated braking, traction control and electric-controlled steering system that works together to help the driver maintain control of the vehicle in emergency situations and in varying road conditions.

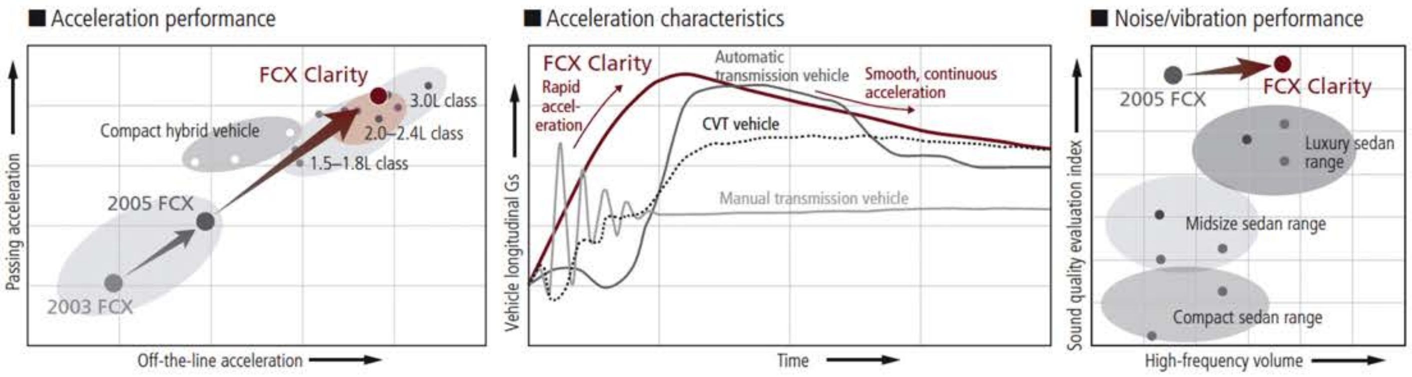


Figure 17.37 Performance and characteristics (Source: Honda Media)

Figure 17.38 Dashboard and controls

Working in conjunction with the vehicle's anti-lock brakes, traction control system (TCS) with slideslip control and vehicle stability assist (VSA), the electric power steering (EPS) enhances steering force for even better handling.

In controlling understeer, EPS provides supplementary steering force to prevent the steering wheel from being turned too far as motor torque is reduced and braking force is applied to the inner rear wheel by the VSA. In controlling oversteer, the EPS provides steering force to help the driver counter the spin-generating moment as braking is applied to the outer front wheel to stabilize the vehicle. When road conditions under the left and right tyres are different, torque and steering force are supplemented to help the driver maintain stability.

As a result of increased energy storage capacity and a broader range of regeneration control, it has been possible to implement a system that regulates acceleration and reduces the need for pedal operation in downhill driving.

Assessing incline and vehicle speed, the system regulates acceleration when the driver first releases the accelerator pedal, minimizing the need for frequent braking. The system simultaneously adjusts the amount of regenerative braking to help maintain constant vehicle speed after brake pedal inputs. The function is similar to engine braking in a conventionally powered vehicle, but more intelligent, smoother and easier to use.

Key fact

The system adjusts the amount of regenerative braking to help maintain constant vehicle speed after brake pedal inputs under downhill conditions.

Figure 17.39 Smooth performance (Source: Honda Media)

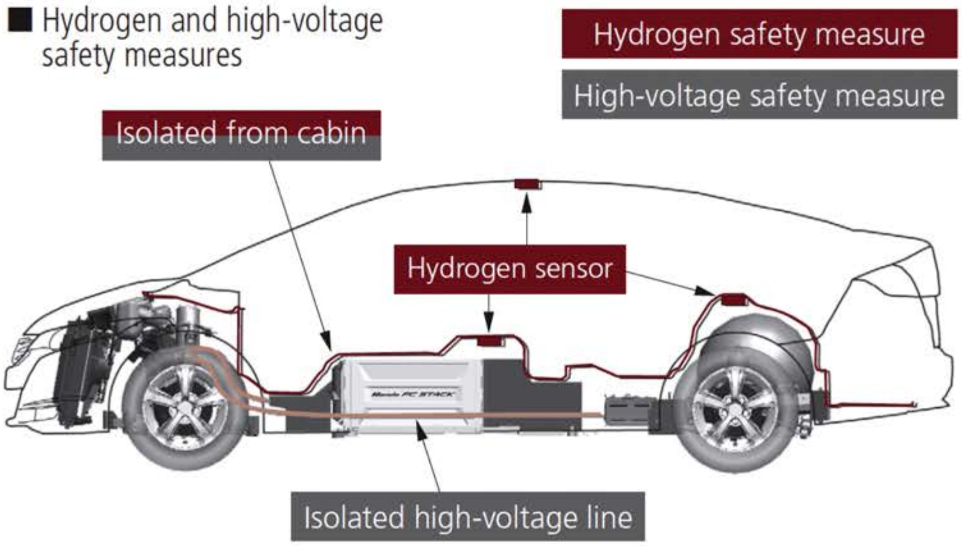


Figure 17.40 Hydrogen and high-voltage safety measures (Source: Honda Media)

17.2.7.6 Safety: hydrogen and high voltage

Sensors are located throughout the vehicle to provide a warning in the unlikely event of a hydrogen leak. If a leak occurs, a ventilation system is activated and an automatic system closes the main cut-off valves on the hydrogen tanks or supply lines (Figure 17.40).

The high-voltage lines are electrically isolated and sensors provide a warning in case of grounding. In the event of a collision, high-voltage contactors shut down the source power line. Repeated flood and fire testing have confirmed a very high level of safety and reliability. Orange covers are used on all high-voltage cables.

During refuelling, to prevent reverse flow from the tank, the hydrogen filler inlet has an integrated check valve. The fuel intake mechanism is also designed to prevent contamination by other gases or the connection of nozzles designed for hydrogen at incompatible pressure levels.

Safety first

In the event of a collision, high-voltage contactors shut down the source power line.

Figure 17.41 Fuelling pipe connection (Source: Honda Media)

Table 17.3 Honda FCX Clarity specifications

Model type	Honda fuel cell electric car	
Dimensions, weight and occupancy	Overall length (mm)	4833
	Overall width (mm)	1846
	Overall height (mm)	1468
	Wheelbase (mm)	2799
	Track (front/rear) (mm)	1579/1595
	Weight (kg)	1625
	Number of occupants	4
Performance	Max speed (mph)	100
	Range (miles)	270
Drive method		Front-wheel drive
Motor	Type	AC synchronous electric motor (permanent magnet)
	Max. output (kW)	100
	Max. torque (Nm)	256
Fuel cell stack	Type	PEMFC (Proton Exchange Membrane Fuel Cell)
	Max. output (kW)	100
Lithium ion battery	Voltage (V)	288
Fuel	Type	Compressed hydrogen gas
	Storage	High-pressure hydrogen tank
	Tank capacity (litres)	171
	Max. pressure when full (MPa)	35

17.2.8 EV summary

The concept of the electric vehicle is not new, the essential battery technology was developed in the late 19th century and many such cars were being manufactured by the year 1900. Although some models achieved high speeds at that time, the electric car was generally slow and expensive to operate. Its range was also limited by its dependence on facilities to recharge the battery. Many of these problems have been overcome, but not all of them. Cost is still an issue, but ‘cost’ is a relative value and when the consequences of pollution are considered the ‘cost’ may not be as high as it appears.

Advances in battery, motor and fuel cell technology have increased the range of the EV significantly. In addition, the electric car is expected to be mechanically more dependable and durable than its fossil-fuelled equivalent. There is some debate as to the environmental costs of building EVs.

However, the EV is here, and only time will tell about the 'real' costs and reliability…

17.3 Hybrid electric vehicles (HEVs)

17.3.1 Introduction

A hybrid vehicle uses a combination of electric drive and a more traditional IC engine. A light hybrid only used the electric motor for 'assistance' and regenerative braking. A full hybrid can drive the vehicle on the electric motor or the IC engine, or both together.

17.3.2 Honda light hybrids

17.3.2.1 Safety

Integrated motor assist (IMA) hybrid vehicles use high-voltage batteries so that energy can be delivered to a drive motor or returned to a battery pack in a very short time. The Honda Insight system, for example, uses a 144 V battery module to store re-generated energy. This energy is then used to drive the IMA motor. This decreases the load on the fuel engine, resulting in reduced emissions and increased efficiency.

The Toyota Prius originally used a 273.6 V battery pack but this was changed in 2004 to a 201.6 V pack, which reduced weight by 26%. Clearly, there are safety issues when working with hybrid vehicles.

Hybrid vehicle batteries and motors have high electrical and magnetic potential that can severely injure or kill if not handled correctly. It is essential that you take note of all the warnings and recommended safety measures outlined by manufacturers and in this resource. Any person with a heart pacemaker or any other electronic medical devices should not work on an integrated motor assist (IMA) system since the magnetic effects could be dangerous. It is essential that you take note of all the warnings and recommended safety measures outlined by manufacturers and in this resource.

Figure 17.42 At the time of writing, the world land speed record for a Prius hybrid was 130.794 mph! (Source: Toyota Media)

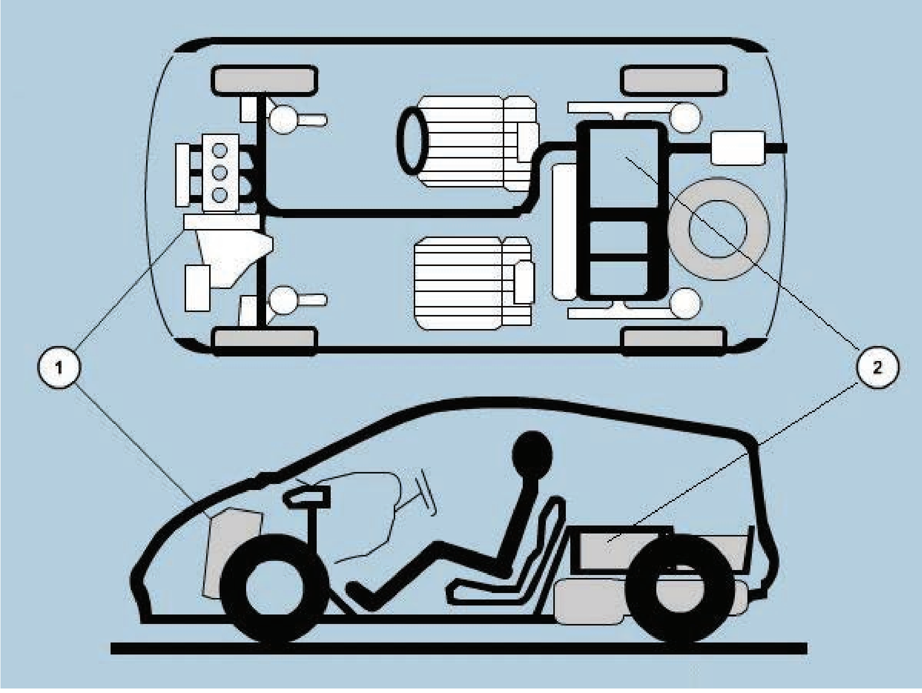


Figure 17.43 Motor and power pack locations: 1-Power pack, 2-IMA (motor)

Most of the hybrid components are combined in the power unit (or integrated power unit, IPU). This is located behind the rear seats or under the luggage compartment floor (Figures 17.43 and 17.44). The unit is a metal box that is completely closed with bolts. A battery module switch is usually located under a small secure cover on the power unit. The electric motor is located between the engine and the transmission or as part of the transmission.

All high voltage components (except the motor) are located in the power unit. The electrical energy is conducted to or from the motor via three thick orange wires. If these wires have to be disconnected, SWITCH OFF the battery module switch. This will prevent the risk of electric shock or short circuit of the high-voltage system. High voltage wires are always orange (Figure 17.45).

Key fact

Most of the hybrid components are combined in the power unit (or integrated power unit, IPU).

Safety first

Whenever high-voltage (orange) wires have to be disconnected: SWITCH OFF the battery module switch.

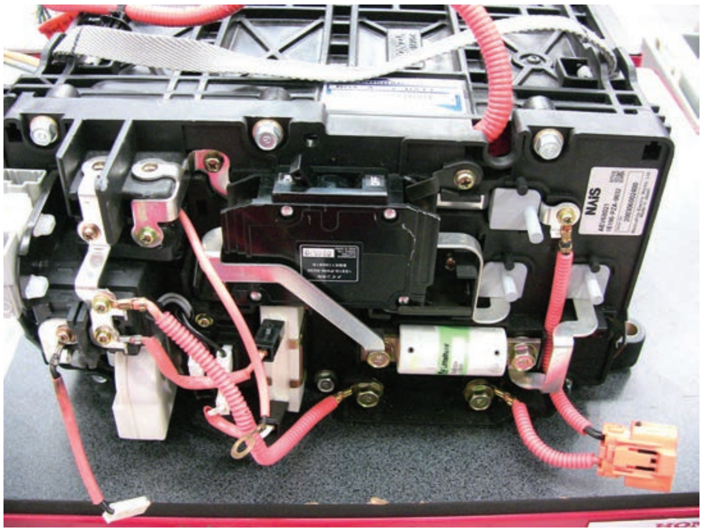


Figure 17.44 Honda battery pack (integrated power unit)

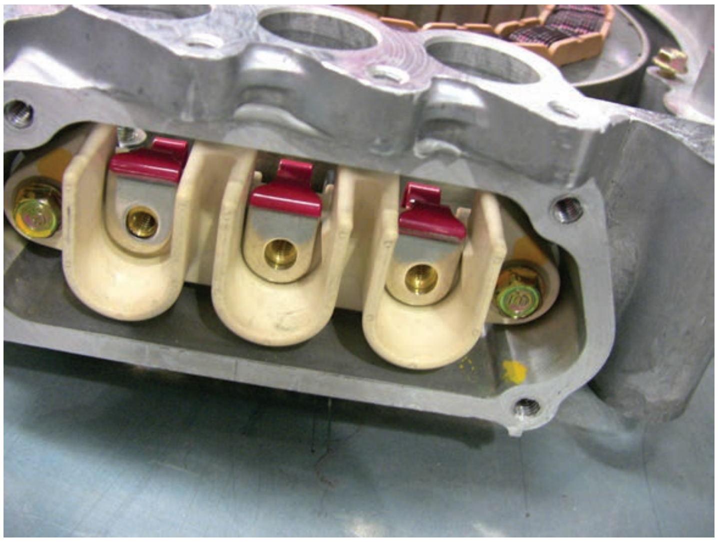


Figure 17.45 Motor power connections

Safety first

Any person with a heart pacemaker or any other electronic medical devices should not work on the IMA system because of the high strength magnetic fields.

Any person with a heart pacemaker or any other electronic medical devices should not work on the IMA system. The magnetic fields present can affect these devices and is therefore a very significant danger (Figure 17.46). The use of any magnetic storage media near the IMA system should be avoided. In the presence of the system's strong magnetic field, data could be partially or totally erased. A mechanical or electronic wristwatch would also be damaged.

Before maintenance

- turn OFF the ignition switch and remove the key;
- switch OFF the Battery Module switch;
- wait for 5 minutes before performing any maintenance procedures on the system. This allows the large storage capacitors to be discharged.

Make sure that the junction board terminal voltage is nearly 0 V.

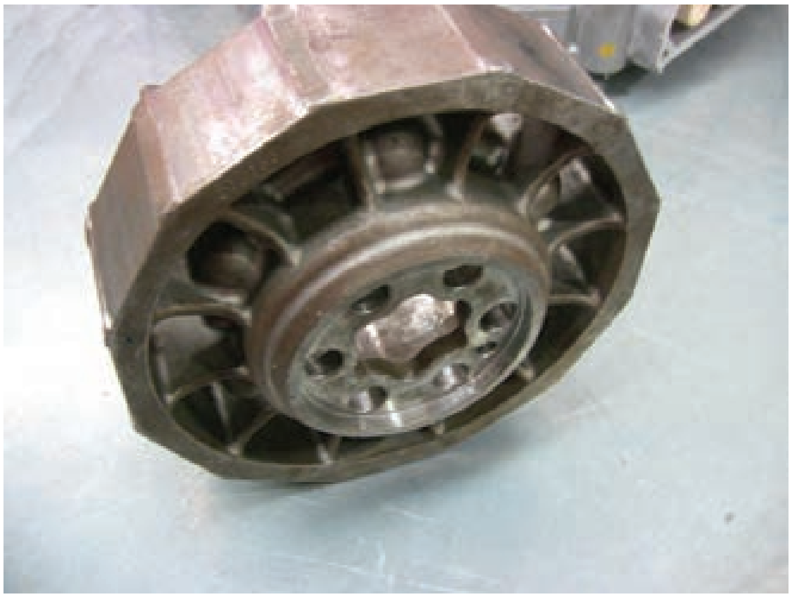


Figure 17.46 The core or rotor is made of very strong rare earth metal permanent magnets

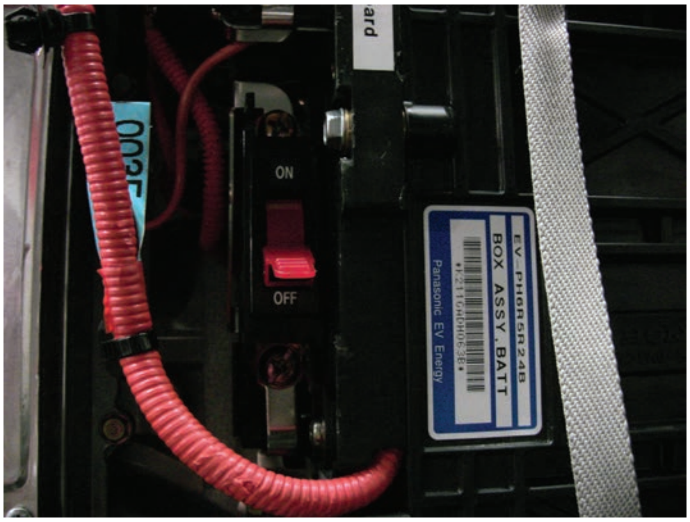


Figure 17.47 High-voltage battery power switch

During maintenance

- always wear insulating gloves.

Always use insulated tools when performing service procedures to the high-voltage system. This precaution will prevent accidental short-circuits.

When maintenance procedures have to be interrupted while some high-voltage components are uncovered or disassembled, make sure that:

- the ignition is turned off and the key is removed (Figure 17.47);
- the Battery Module switch is switched off.

No untrained persons have access to that area and prevent any unintended touching of the components.

Before switching on the battery module switch after repairs have been completed, make sure that:

- all terminals have been tightened to the specified torque;
- no high-voltage wires or terminals have been damaged or shorted to the body.

The insulation resistance between each high-voltage terminal of the part you disassembled and the vehicle's body has been checked.

Safety first

Electrical safety gloves are NOT the same as general working gloves.

Figure 17.48 Insulated gloves

Figure 17.49 High-voltage cables are always orange

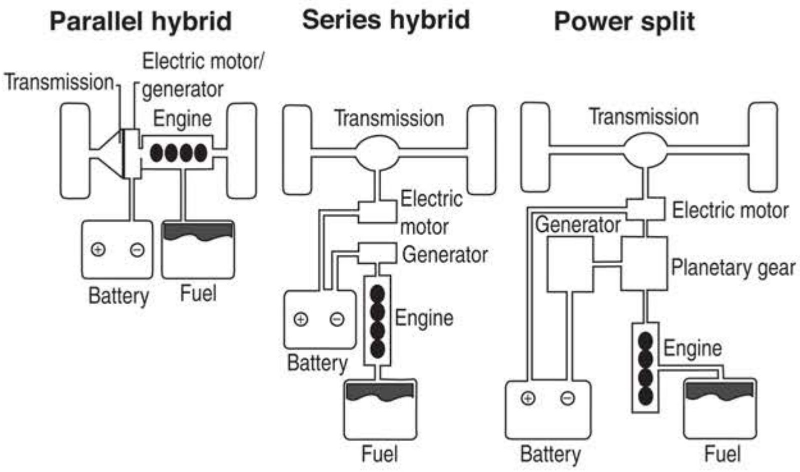


Figure 17.50 Three types of hybrid vehicles

Key fact

A hybrid power system for an automobile can have a series, parallel or power split configuration.

Key fact

The IMA system is a parallel hybrid system.

Working on hybrid vehicles is not dangerous **IF** the previous guidelines and manufacturers procedures are followed. Before starting work, check the latest information – DON'T take chances. Dying from an electrical shock is not funny.

17.3.2.2 Overview

A hybrid power system for an automobile can have a series, parallel or power split configuration (Figure 17.50). With a series system, an engine drives a generator, which in turn powers a motor. The motor propels the vehicle. With a parallel system, the engine and motor can both be used to propel the vehicle. Most hybrids in current use employ a parallel system known as Integrated Motor Assist (IMA). The power split has additional advantages but is also more complex.

The IMA method is a technologically advanced parallel hybrid power system. By employing techniques such as brake-energy regeneration to maximize the efficiency with which energy is used, it combines low-pollution, low-cost operation with high levels of safety and running performance.

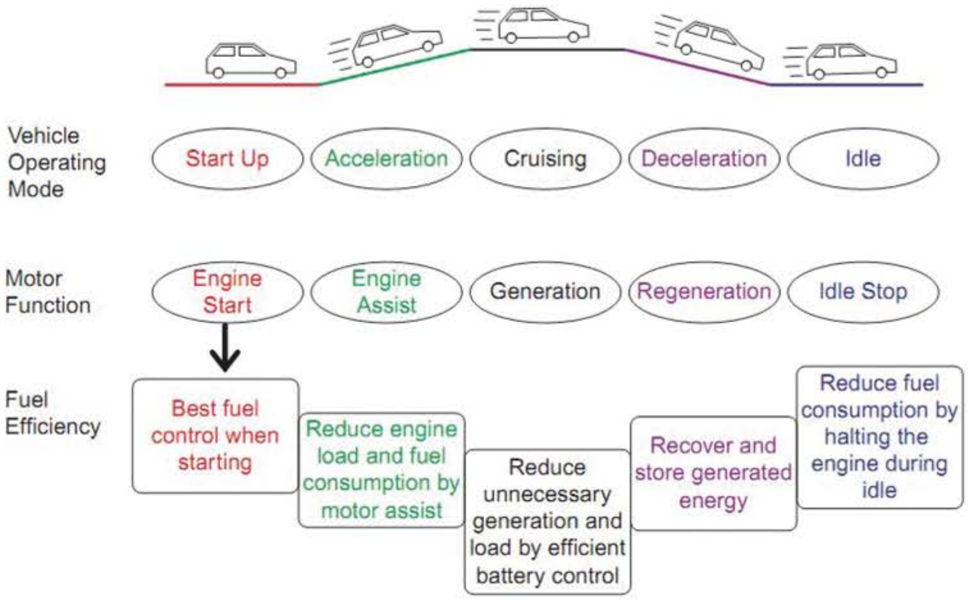


Figure 17.51 Operating conditions

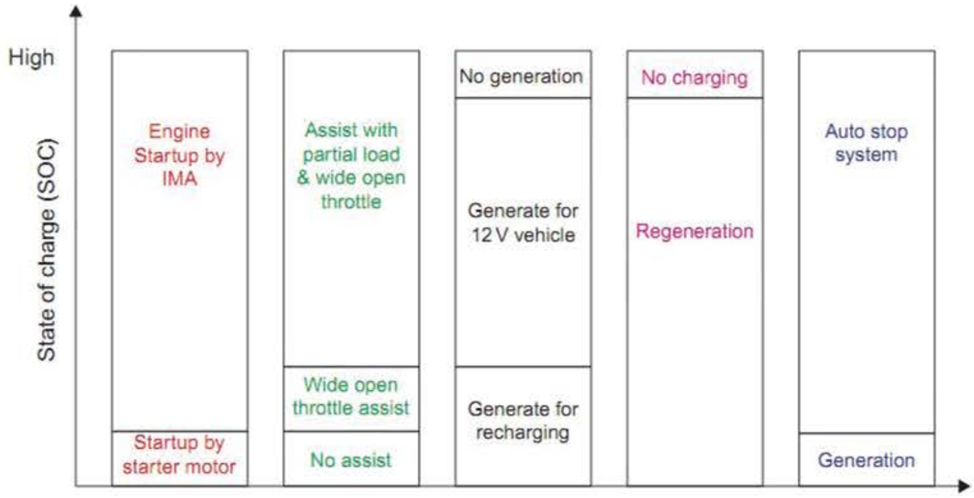


Figure 17.52 IMA operating details

The main components of the system are:

- IMI motor.
- Battery module.
- Power drive unit (PDU).
- Motor control module (MCM).
- DC-DC converter.

There are five main IMA operating modes (Figures 17.51 and 17.52 give an overview of each mode).

Table 17.4 explains these operating modes in more detail.

The IMA technique used by most hybrid cars can be thought of as a kinetic energy recovery system (KERS). This is because instead of wasting heat energy

Figure 17.53 Hybrid vehicles still need exhaust extraction!

Table 17.4 IMA operating modes

Mode	Details
Start-up	Under normal conditions, the IMA Motor will immediately start the engine at a speed of 1000 rpm. When the state of charge (SOC) of the high-voltage battery module is too low, when the temperature is too low, or if there is a failure of the IMA system, the engine will be cranked by the normal 12 V starter motor.
Acceleration	During acceleration, current from the battery module is converted to AC by the power drive unit (PDU) and supplied to the IMA motor, which functions as a motor. The IMA motor output is used to supplement the engine output so that power available for acceleration is maximized. Current from the battery module is also converted to 12 V DC for supply to the vehicle electrical system. This reduces the load that would have been caused by a normal alternator and so improves acceleration. When the remaining battery module state of charge is too low, but not at the minimum level, assist will only be available during wide open throttle (WOT) acceleration. When the remaining state of charge is reduced to the minimum level, no assist will be provided. The IMA system will generate energy only to supply the vehicle's 12 V system.
Cruising	When the vehicle is cruising and the battery module requires charging, the engine drives the IMA Motor, which now acts as a generator. The resulting output current is used to charge the battery module and is converted to 12 V DC to supply the vehicle electrical system. When the vehicle is cruising and the high-voltage battery is sufficiently charged, the engine drives the IMA motor. The generated current is converted to 12 V DC and only used to supply the vehicle electrical system.
Deceleration	During deceleration (during fuel cut), the IMA motor is driven by the wheels such that regeneration takes place. The generated AC is converted by the power drive unit (PDU) into DC and used to charge the battery module. The DC output of the PDU is also applied to the DC-DC converter which reduces the voltage to 12 V, which is supplied to the vehicle electrical system. It is further used to charge the 12 V battery as necessary. During braking (brake switch on), a higher amount of regeneration will be allowed. This will increase the deceleration force so the driver will automatically adjust the force on the brake pedal. In this mode, more charge is sent to the battery module. If the ABS system is controlling the locking of the wheels, an 'ABS-busy' signal is sent to the motor control module. This will immediately stop generation to prevent interference with the ABS system. When the high-voltage battery is fully charged, there will only be generation for the vehicle's 12 V system.
Idling	During idling, the flow of energy is similar to that for cruising. If the state of charge of the battery module is very low, the motor control module (MCM) will signal the engine control module (ECM) to raise the idle speed to approximately 1100 rpm.

from the brakes as the vehicle is slowed down, some is converted to electrical energy and stored in the battery as chemical energy. This is then used to drive the wheels so saving chemical energy from the fuel!

17.3.2.3 IMA battery

The Honda battery module uses nickel metal hydride (Ni-MH) technology for high energy density and long service life. The batteries are constructed in a modular form with a terminal voltage of 100.8 V to 144 V (or more on the Toyota Prius) and a rated capacity of about 6.5 Ah.

If servicing is required to the battery module, refer to manufacturers' instructions as serious injury or even death can occur if the safety precautions are not observed. The batteries typically only weigh about 22 kg. Its operating range is from –30 to +50°C.

The high-voltage batteries are fitted either behind the rear passenger seats or in some cases under the floor of the luggage compartment.

The battery module is used to supply high voltage to the electric motor during the assist mode. The battery module is also used to store the regenerated power while cruising, deceleration and braking. Current from the battery module is also converted to 12 V DC, which is supplied to the vehicle electrical system. A conventional 12 V battery located within the engine compartment is used for the vehicle's 12 V system.

Key fact

The IMA system is a kinetic energy recovery system (KERS).

Safety first

If servicing is required to the battery module, refer to manufacturers' instructions as serious injury or even death can occur if the safety precautions are not observed.

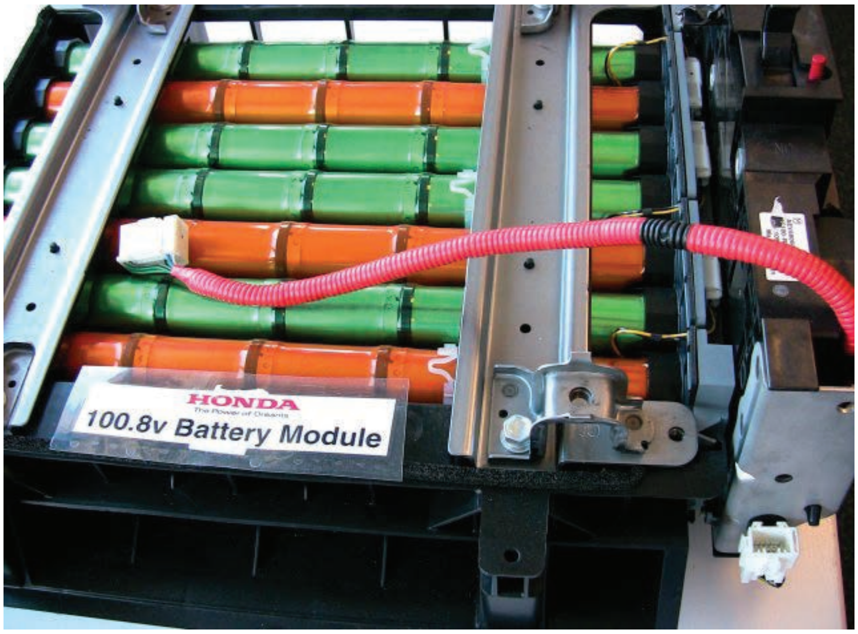


Figure 17.54 Battery module

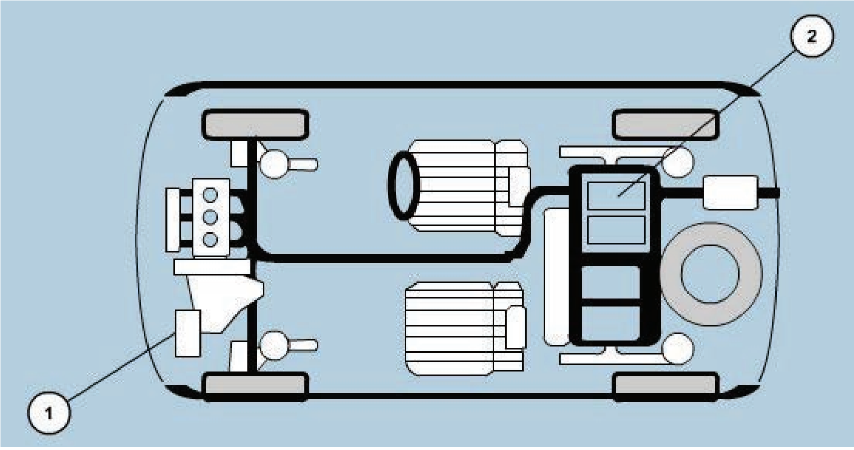


Figure 17.55 High-voltage battery location

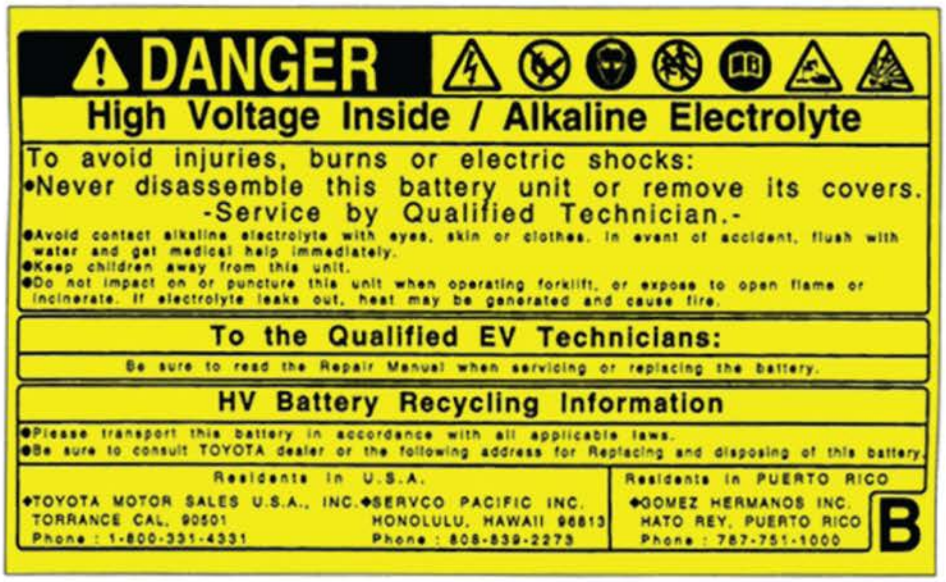
A battery module typically consists of:

- Voltage sensors.
- Temperature sensors (thermistors).
- Battery cell groups. Each cell group consists of 6 cells, 1 cell equals 1.2 V.
- Cooling fan.
- Terminal plate.

The battery cell groups are connected in series by the terminal plates, located on both sides of the battery module.

Charge and discharge are caused by movements of hydrogen when a chemical reaction takes place in the cells. The general construction of a cell is similar to that of a conventional battery but the positive electrode is made of nickel hydroxide. The negative electrode is made of metal hydride (a hydrogen absorbing alloy) and the electrolyte is potassium hydroxide, a strongly alkaline solution.

Figure 17.56 Battery cells and groups



DANGER

High Voltage Inside / Alkaline Electrolyte

To avoid injuries, burns or electric shocks:

- Never disassemble this battery unit or remove its covers.

-Service by Qualified Technician.-

- Avoid contact alkaline electrolyte with eyes, skin or clothes. In event of accident, flush with water and get medical help immediately.
- Keep children away from this unit.
- Do not impact on or puncture this unit when operating forklift, or expose to open flame or incinerate. If electrolyte leaks out, heat may be generated and cause fire.

To the Qualified EV Technicians:

Be sure to read the Repair Manual when servicing or replacing the battery.

HV Battery Recycling Information

- Please transport this battery in accordance with all applicable laws.
- Be sure to consult TOYOTA dealer or the following address for Replacing and disposing of this battery.

Residents in U.S.A.

◆TOYOTA MOTOR SALES U.S.A., INC.
TORRANCE CAL. 90501
Phone : 1-800-331-4331

◆SERVCO PACIFIC INC.
HONOLULU, HAWAII 96813
Phone : 808-839-2273

Residents in PUERTO RICO

◆GOMEZ HERMANOS INC.
HATO REY, PUERTO RICO
Phone : 787-751-1000

B

Figure 17.57 Information plate from a Toyota

Follow safety procedures at all times – a strong alkali is just as dangerous as a strong acid:

- Wear protective clothing:
  - Safety shoes
  - Safety glasses
  - Suitable rubber, latex or nitrile gloves.
- Neutralize electrolyte.

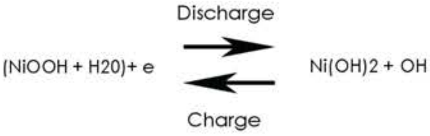


Figure 17.58 Positive plate reaction

In a discharged state the surface of the positive electrode will contain nickel di-hydroxide - $Ni(OH)_2$ and there will be hydroxide ions OH- in the electrolyte. As the battery is charged the positive electrode loses a hydrogen atom and becomes nickel hydroxide - NiOOH. The freed hydrogen atom joins with the hydroxide ion to form water H_2O, and a free electron is released (Figure 17.58).

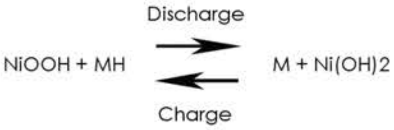


Figure 17.59 Negative plate reaction

In a discharged state the negative electrode consists of the metal alloy, surrounded by H_2O and free electrons. As the battery is charged a hydrogen atom is dislodged from the water and is absorbed by the metal alloy to make metal hydride (MH). This leaves hydroxide ions OH- in the surrounding electrolyte (Figure 17.59).

Nickel metal hydride (Ni-MH) batteries are used because they are robust, long lasting, charge or discharge quickly and have a high energy density.

The energy store of the future may be the lithium-ion battery. Bosch and Samsung are working together to further develop this technology for automotive applications. The main aim is to improve the energy density of this battery threefold, and to cut costs by two-thirds.

Key fact

The energy store of the future may be the lithium-ion battery.

17.3.2.4 IMA motor

The thin design Honda IMA motor is located between the engine and the transmission. It is a permanent magnet type, brushless DC machine, which operates as a motor or a generator.

Key fact

An IMA motor is usually located between the engine and the transmission.

The functions of the IMA motor are:

- To assist the engine under certain conditions determined by the motor control module (MCM) for improving fuel economy, low emission and drivability.

Figure 17.60 Ni-MH battery cells

Figure 17.61 Lithium battery pack (Source: Bosch Media)

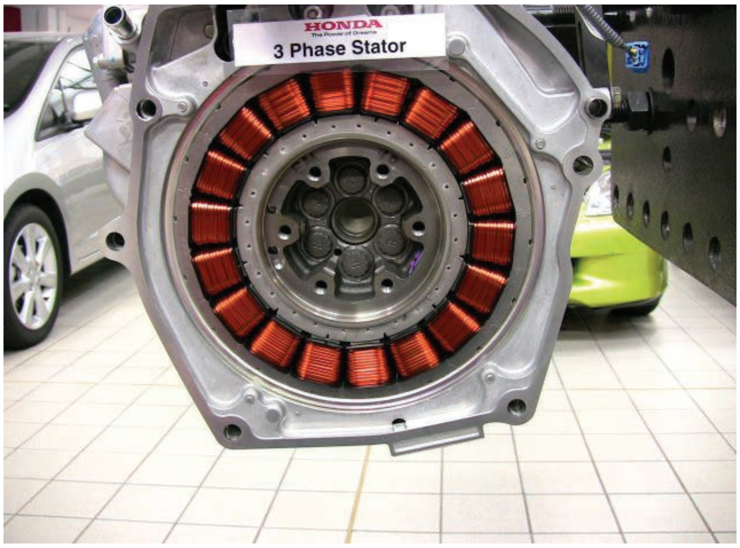


Figure 17.62 Motor stator in position on a Honda engine

Figure 17.63 Power cable connected to the motor (remember, orange cables are high voltage)

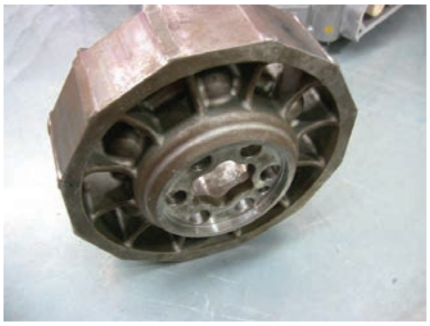


Figure 17.64 Permanent magnet rotor

Key fact

In a brushless motor the conventional DC motor is turned inside out.

- To regenerate power under certain conditions to charge the high-voltage battery module and the normal 12 V battery.
- To start the engine when the state of charge is sufficient.

The motor is located between the engine and transmission gearbox.

The specifications for the Honda IMA motor shown here are as follows:

- Type DC brushless.
- Rated voltage 144 V.
- Power 10 kW/3000 rpm.
- Torque 49 Nm/1000 rpm.

The rated voltage will vary between about 100 V and almost 300 V. These figures for power and torque are typical but other motors will vary.

In a conventional DC motor, the housing contains field magnets, the rotor is made up of coils wound in slots in an iron core and is connected to a commutator. However, in a brushless motor the conventional DC motor is turned inside out. The rotor becomes a permanent magnet and the stator becomes the wound iron core. The advantages are:

- better cooling;
- no brushes to wear out;
- no maintenance.

The disadvantages, however, are:

- More complex motor control circuits.
- Expensive rare earth magnets have to be used because conventional iron magnets demagnetize when a large current is applied to them.

Key fact

A brushless DC motor is supplied with AC.

The stator is constructed of 18 coils surrounding a rotor containing 12 poles. Although the motor is a DC type, note that the current supplied to it must be AC. This is because in normal DC motors (brush type) the current is reversed by the brushes and commutator. If the current is not reversed the rotor would come to a stop. The force that rotates the rotor is the interaction of the two magnetic fields produced by the stator coils and the rotor. These fields must remain constant in magnitude and relative orientation to produce a constant torque (Figures 17.62 and 17.65).

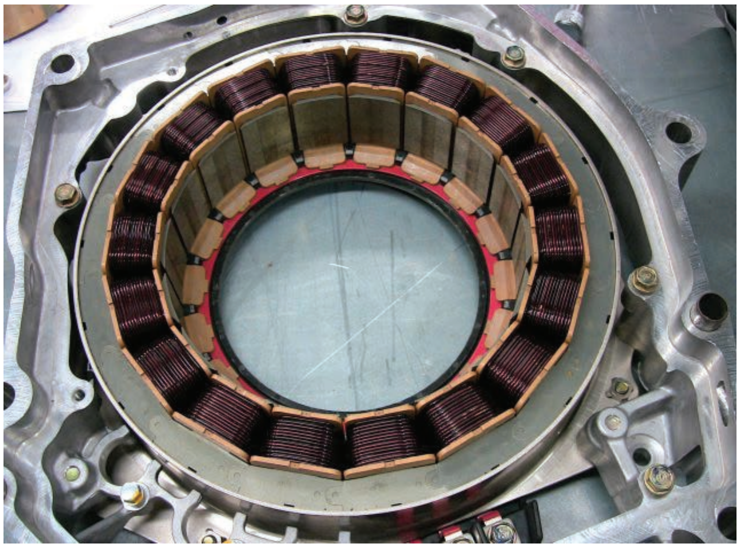


Figure 17.65 IMA motor stator removed from the vehicle

To maintain a constant field the stator coils are divided in to six groups of three. Each group has three electrical phases in each coil designated as: U, V and W. Switching of current for each phase of these coils takes place in a power driver unit (PDU) when the motor is acting as a motor or as a generator. Maximum torque is produced when the rotor and the stator fields are at 90 degrees to each other.

To control the stator coil fields correctly the relative position of the rotor must be known. A sensor disk is therefore attached to the rotor and is divided into 12 partitions – 6 high and 6 low. These are detected by three commutation sensors.

There are three commutation sensors (Figure 17.67). They act in a similar way to an ABS sensor where metal teeth passing on a sensor wheel induce a signal current in the sensor. Each sensor is composed of two small magnetic reluctance elements that detect the presence of a high partition or low partition passing the sensor. The two variable magnetic reluctance elements transform their signal from a variable to a high (1) or low (0) signal.

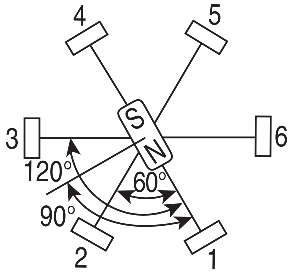


Figure 17.66 Switching sequence

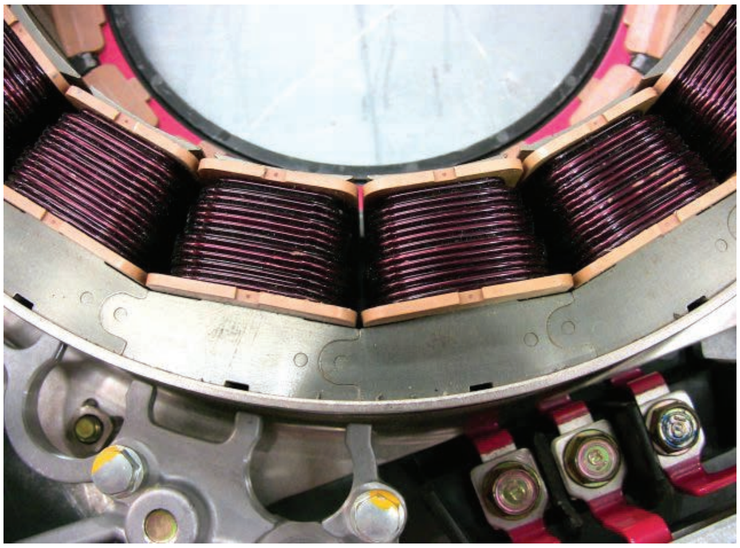


Figure 17.67 The sensor ring (just behind the stator coils) has high and low sections that are detected by the commutation sensors

Figure 17.68 Commutation sensor connection

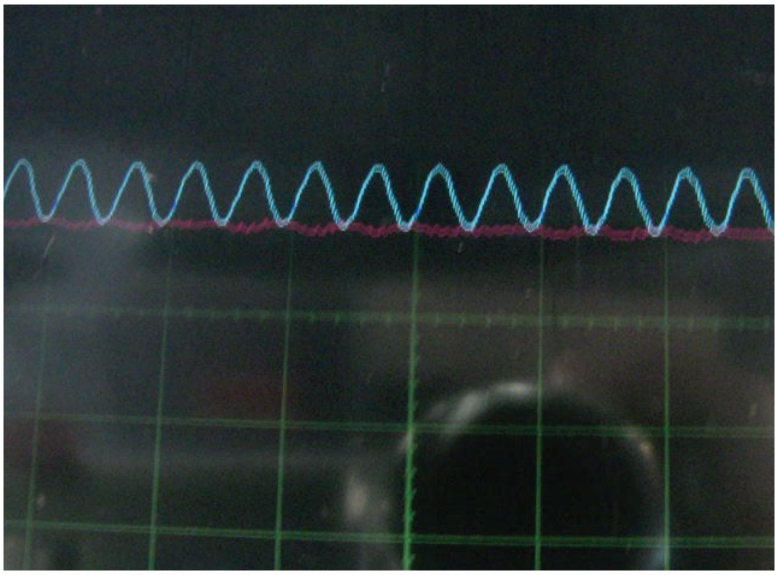


Figure 17.69 Sensor signal

Push on or ring and screw type terminals are used on the motor. Orange coloured cables are attached, and make the connection between the IMA motor and the power drive unit (PDU) at the rear of the vehicle.

Developments are ongoing in the hybrid motor field. However, the technology of stationary coils rotating permanent magnets seems to be well developed. A range of switching and control methods are used but in simple terms the stator coils are energized in sequence to drive the rotor. When the rotor is driven by the wheels (on deceleration or braking) it induces electrical energy in the coils and this is used to charge the battery via suitable rectification and voltage controls.

This section has outlined a system used by Honda. An alternative Bosch hybrid transmission and the associated IMA system are shown below.

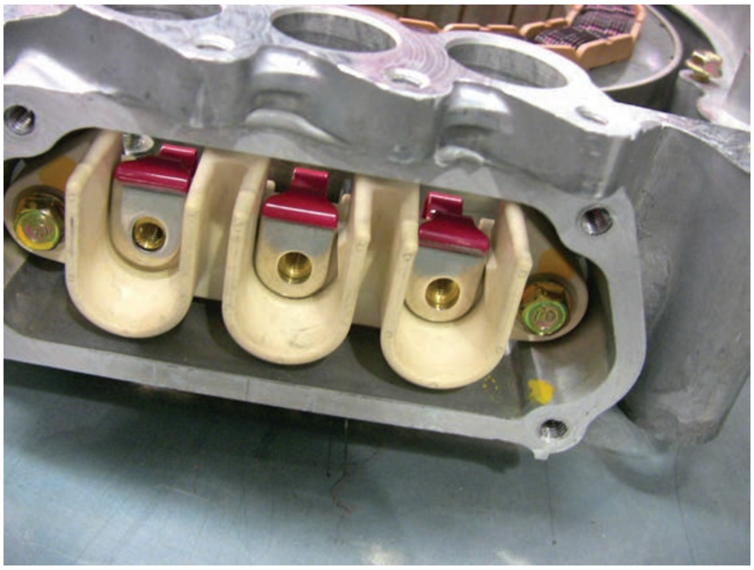


Figure 17.70 Motor terminals for the high-voltage cables (which always have an orange covering)

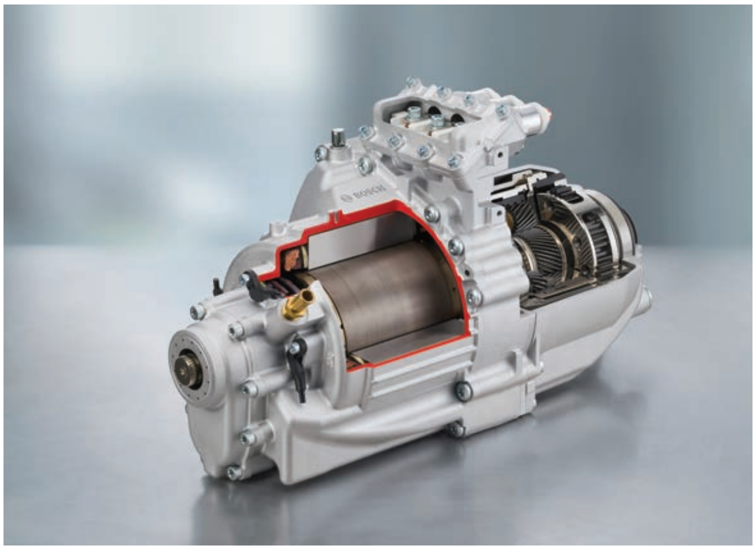


Figure 17.71 Transmission for a hybrid (Source: Bosch Media)

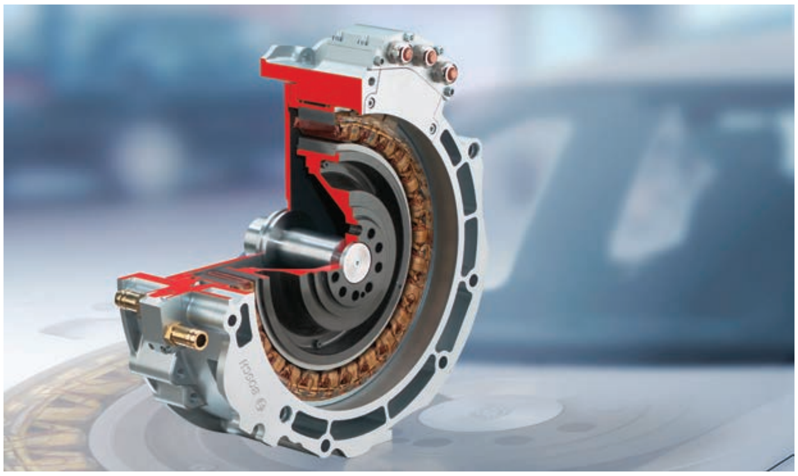


Figure 17.72 IMA motor/generator (Source: Bosch Media)

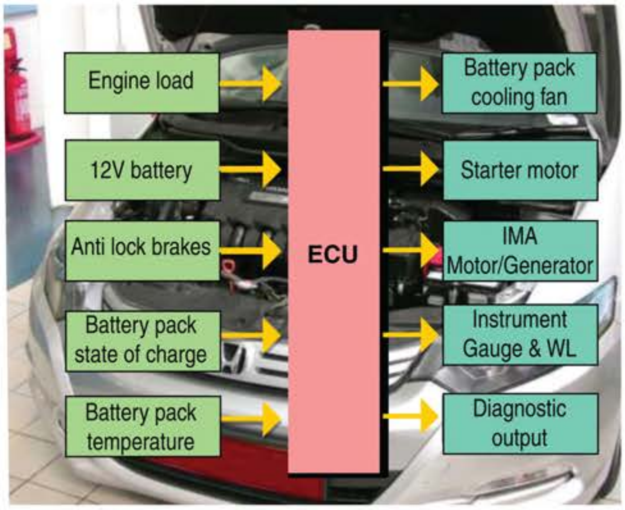


Figure 17.73 Basic input-control-output block diagram

17.3.2.5 Hybrid IMA control system

Like any other complex control system, the control of the hybrid IMA system can be represented as a block diagram showing inputs and outputs (Figure 17.73). The IMA system can seem more complex because the motor changes to become a generator and back to a motor depending on road conditions. However, thinking of the system as shown here will help with your understanding of the operation.

The following diagram expands the basic block diagram. In this case the main component locations are shown.

Signals from the three commutation sensors are sent to the motor control module (MCM). The MCM is connected to the power drive unit making it possible for the battery module and the IMA motor to interact.

Key fact

Signals from the three commutation sensors are sent to the motor control module (MCM).

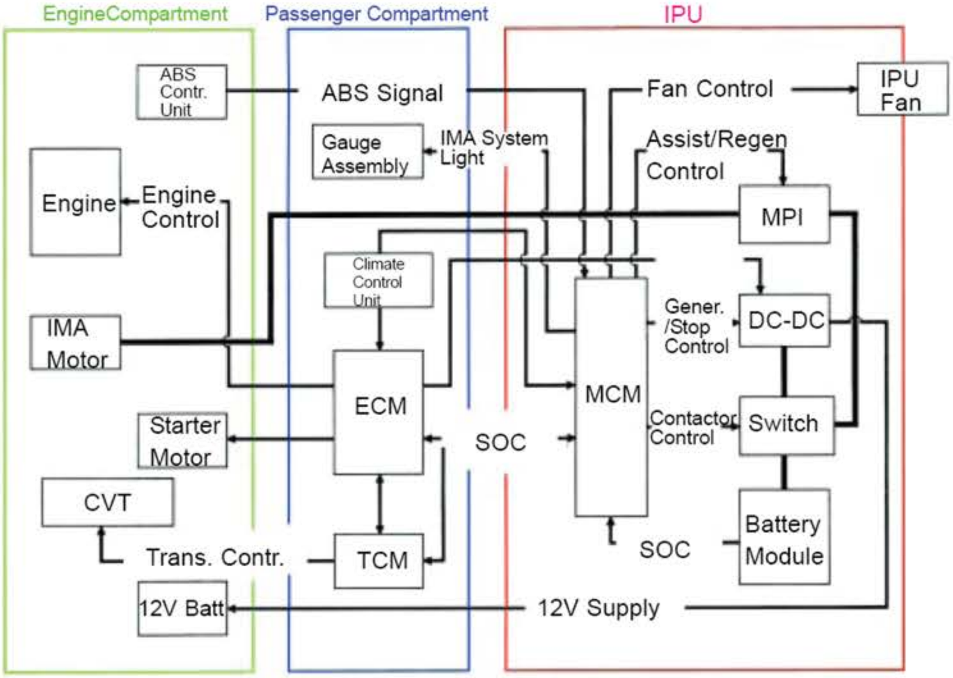


Figure 17.74 Block diagram showing all components and their locations

Figure 17.75 Motor control module

The three signals coming from the commutation sensor on the IMA motor are sent to the MCM and transferred by the module into high and low signals for the stator's coil phases U, V, and W. According to these signals the circuits from the battery module to motor, or from motor to battery, are made by the power drive unit.

The power drive unit (PDU) consists of six power switches with a gate drive circuit. The switches are insulated gate bipolar transistors (IGBT), which are able to control very large amount of power with a very small signal.

There are six motor commutation steps and as each step is made, another signal is generated. All six steps are different and none has a position, where U, V, and W are all low or all high. Two IGBTs in the same line never switch on together. It is always one IGBT in the upper side, and one in the lower side. This is very similar to how a stepper motor driver circuit works. Figures 17.77 and 17.78 show example switched paths that are the same when driving the motor or charging from the generator – except that current flow is reversed.

Definition

IGBT: The insulated gate bipolar transistor is a three-terminal power semiconductor device, noted for high efficiency and fast switching.

Figure 17.76 Inside a PDU

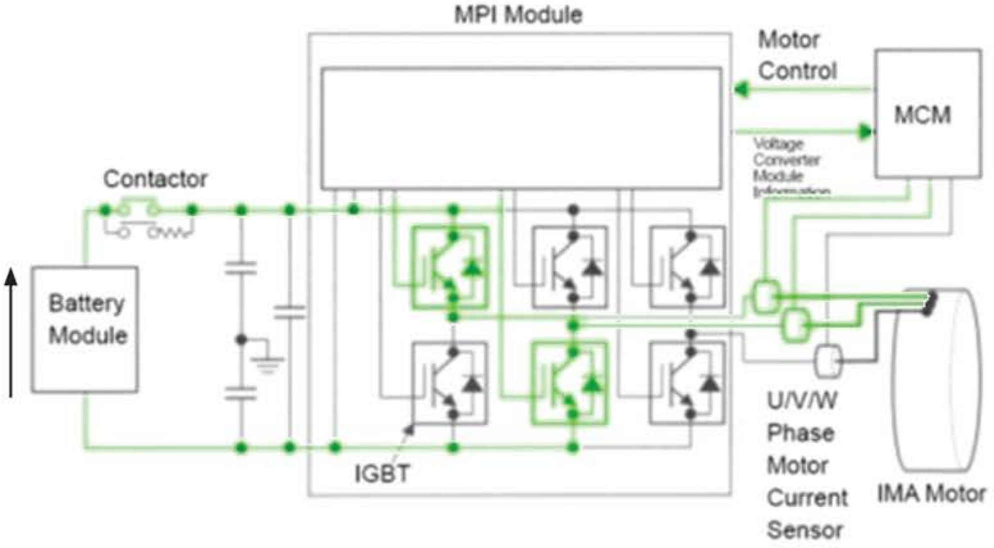


Figure 17.77 Motor circuit operation (example phase highlighted)

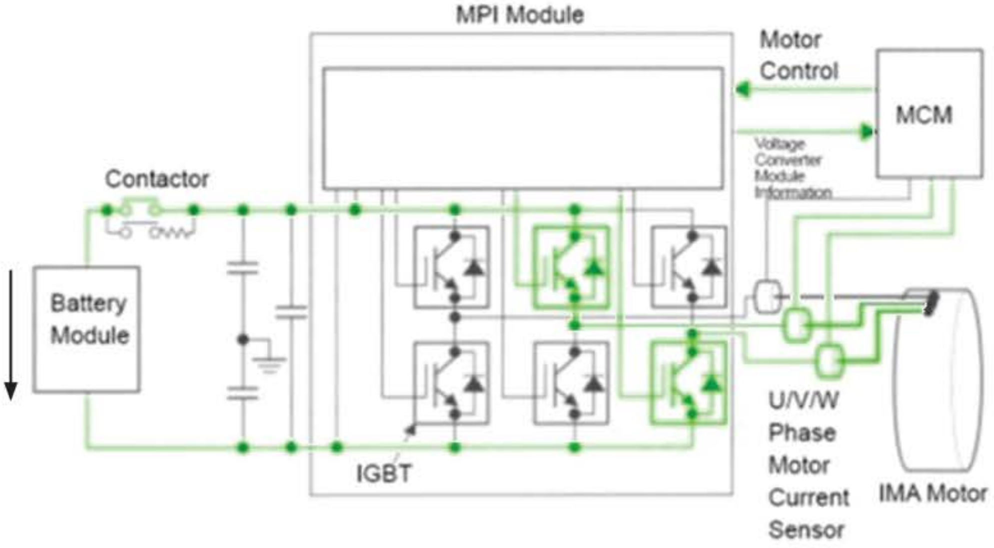


Figure 17.78 Generator circuit operation (example phase highlighted)

When the motor is acting as a generator power is transferred from the stator through the PDU diodes controlled by signals from the commutation sensors. The PDU works in a similar way to a normal alternator rectifier.

Key fact

A DC-DC converter converts the high voltage from the main battery to charge the 12 V battery and run the system.

The DC-DC converter takes the high battery module voltage and converts it to charge the 12 V battery and run the system. Charging the battery and running the low-voltage system from the high-voltage system is more efficient than using a standard alternator.

The other key features used by many hybrid vehicles to improve efficiency are:

- Idle stop/start – to save fuel the engine is stopped, at traffic lights for example, and restarted almost instantly by the IMA.
- Braking control – the most important aspect of a hybrid is collecting energy normally lost on braking. If the normal brake operation is also electronically controlled so that more regenerative effect is used, efficiency is improved further.

- Engine valve control – to further enhance the regenerative effect, the braking effect of the internal combustion engine is reduced by preventing the valves operating.
- AC control – on some systems the AC is run by an electric motor so that it continues to work in stop/start conditions.
- Instrumentation feedback – it is well known that on any vehicle, a significant effect on economy is driving style. Drivers who have opted for a hybrid tend to be looking for economy so are willing to change their style even further based on feedback. Some instruments show images such as growing green trees to indicate driving performance improvement!

The efficiency of the hybrid car has now improved significantly. Sophisticated control systems and highly developed and efficient component designs are the reason for this. However, remember that as with any complex system, it can be thought of and inputs and outputs – and this makes it much easier to understand.

Figure 17.79 DC-DC converter

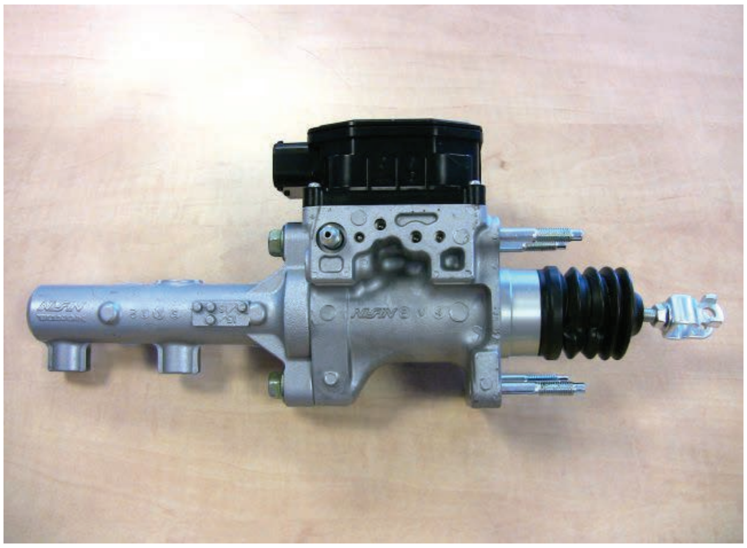


Figure 17.80 Electronically controlled brake master cylinder

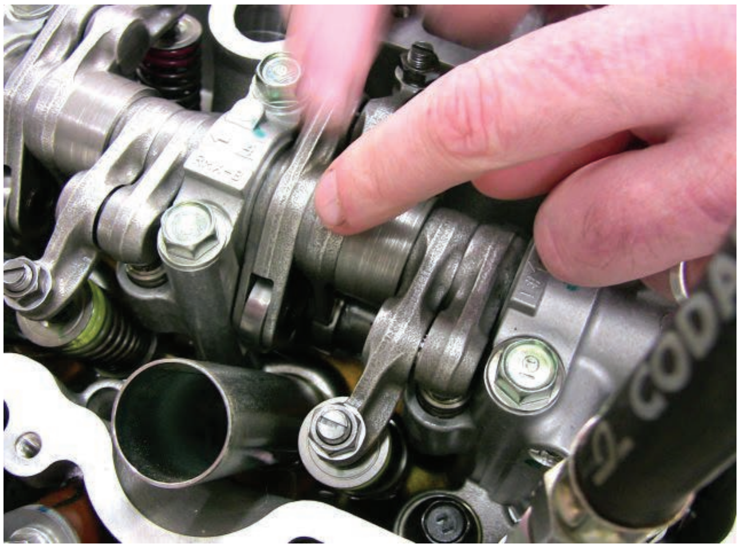


Figure 17.81 Valve control allows more braking to be done by the motor/generator

Figure 17.82 Feedback on economy/performance helps to change driving style and improve efficiency even more

17.3.3 Bosch parallel full-hybrid technology

17.3.3.1 Introduction

The hybrid variants of the Volkswagen Touareg and Porsche Cayenne S, which recently went into production, feature hybrid technology supplied by Bosch. This is the first time that either of these models has been available as a parallel full hybrid. As well as key components such as the power electronics and electric motor, Bosch is also providing the ‘brain’ of the vehicles in the form of the Motronic control unit for hybrid vehicles, which governs when the electric motor, internal-combustion engine, or a combination of the two kick into action.

Volkswagen and Porsche both chose to equip their hybrid vehicles with a 3.0-litre V6 supercharged direct-injection engine and an eight-speed automatic transmission. The six-cylinder V-engine delivers 245 kW (333 hp) and a maximum torque of 440 Nm starting from 3000 rpm. The vehicle also features an integrated motor generator (IMG) developed by Bosch. The water-cooled electric motor includes a separate clutch. The hybrid module is positioned between the IC engine and the transmission; it has a diameter of 30 cm and a length of just 145 mm. The IMG delivers 34 kW and a maximum torque of 300 Nm. That means the cars can cruise at a maximum of 50–60 km/h running on electric power alone, as long as the nickel metal hydride (NiMH) battery has enough charge.

Key fact

The IMG delivers 34 kW and a maximum torque of 300 Nm.

The battery has an energy capacity of 1.7 kWh with a peak of 288 V. During braking, the electric motor, now operating as a generator, recovers kinetic energy, which is then stored in the high-voltage battery.

Lifting off the throttle at any speed up to around 160 km/h activates what is referred to as ‘sailing’ mode: the IC engine automatically shuts down and the vehicle coasts along without consuming fuel – obviously without sacrificing any of the functionality of the systems required for a safe and comfortable drive. Braking is also a fully automatic process. The hybrid control unit monitors the pressure on the brake pedal to determine what brake torque should be electrically set by the IMG. This does not affect safety systems such as ABS and ESP®, which take precedence whatever the situation.

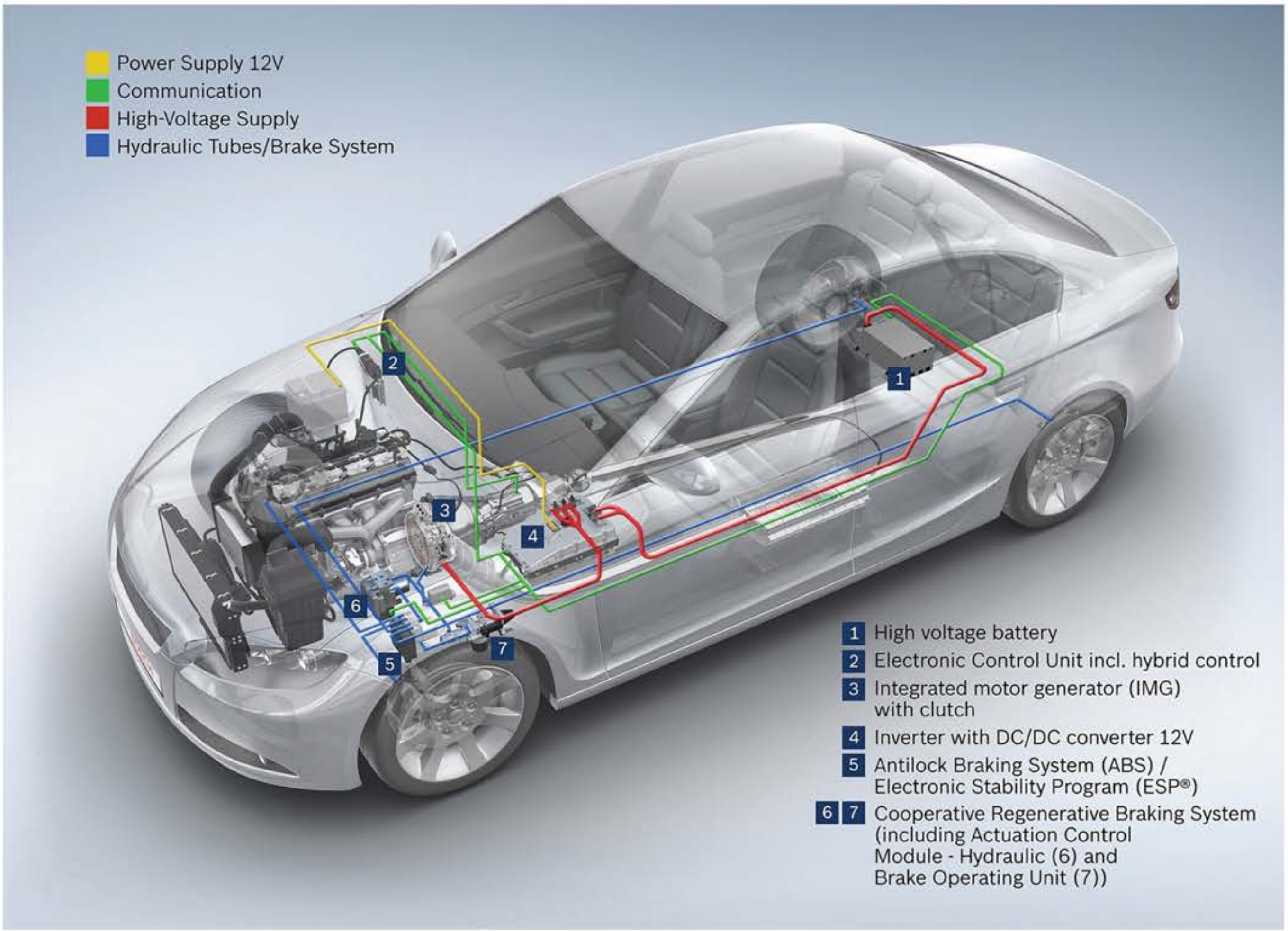


Figure 17.83 Hybrid components and supply systems (Source: Bosch Media)

Figure 17.84 Integrated motor generator (Source: Bosch Media)

Figure 17.85 Braking control components (Source: Bosch Media)

Key fact

This ‘power boost’ function increases the vehicle’s performance to 279 kW (380 hp), offering the driver a maximum torque of 580 Nm.

17.3.3.2 Power boost

For drivers in a hurry, the electric motor and the combustion engine can also work in tandem, allowing the cars to sprint from 0 to 100 kilometres per hour in 6.5 seconds. This ‘power boost’ function increases the vehicle’s performance to 279 kW (380 hp), offering the driver a maximum torque of 580 Nm. Compared to the first-generation V8 vehicles, these hybrid vehicles cut fuel consumption by up to 40%. EU cycle fuel consumption falls to 8.2 L per 100 km, equivalent to CO_2 emissions of 193 grams per km. Both vehicles also comply with the Euro 5 standard and the U.S. emissions standard ULEV 2.

17.3.3.3 Control system

The fact that the internal combustion engine and the electric motor work together so seamlessly stems from the perfectly tuned interaction between modern management and control technology and optimized hybrid components. Bosch can draw on many years of experience in this field thanks to its work on

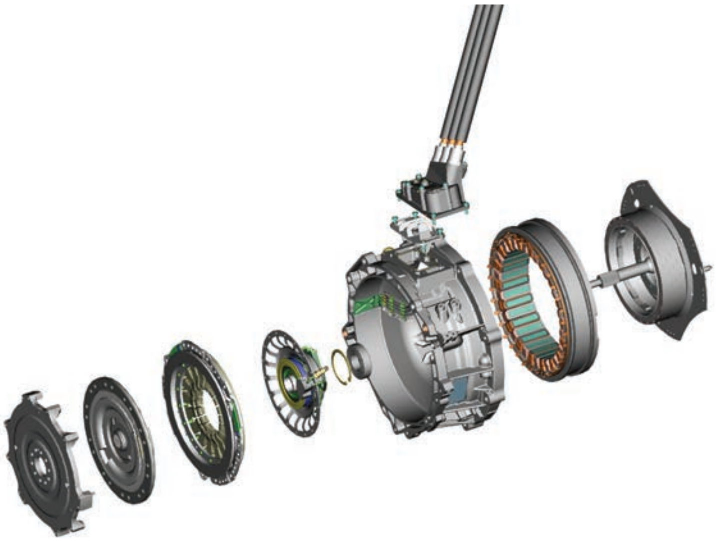


Figure 17.86 Motor and clutch assembly (Source: Bosch Media)

developing gasoline injection systems. The control unit is based on Motronic, which has already proved its worth in so many direct injection gasoline vehicles. The additional functions needed for hybrid operation were integrated. The control unit, for example, ensures that the electric motor and engine are turning at exactly the same speed when transferring the torque between them.

Key fact

The control unit ensures that the electric motor and engine are turning at exactly the same speed when transferring the torque between them.

17.3.3.4 Hybrid and GDi engines

The supercharged V6 engine is a key part of the overall concept. The Motronic control unit manages the combustion engine with tremendous precision, right down to the rate of individual injections. It employs an additional CAN bus interface to exchange all relevant data with the hybrid components, power electronics, and battery, and the efficient direct injection system also reduces exhaust emissions. The combustion engine and electric motor complement each other perfectly, enabling parallel hybrids to offer a whole series of new features to improve driving comfort.

17.3.3.5 Optimized components

Parallel full hybrid technology can be implemented as a more cost-effective solution in comparison to other hybrid concepts. For example, it requires just one electric motor, which operates as both a motor and a generator. The power electronics are a core component, providing an interface between the high-voltage electric drive and the vehicle’s 12 V electrical system, and featuring an inverter that converts the direct current from the battery into three-phase alternating current for the electric motor, and vice versa. All components are optimized for space and performance.

Key fact

All components are optimized for space and performance.

17.3.4 Nissan hybrid case study

Nissan’s 2012 hybrid system aims for 1.8 L efficiency with a 3.5 L V6 (Figure 17.89). The hybrid drive system packages within the current seven-speed automatic architecture and the use of a single e-motor is said to balance performance and cost.

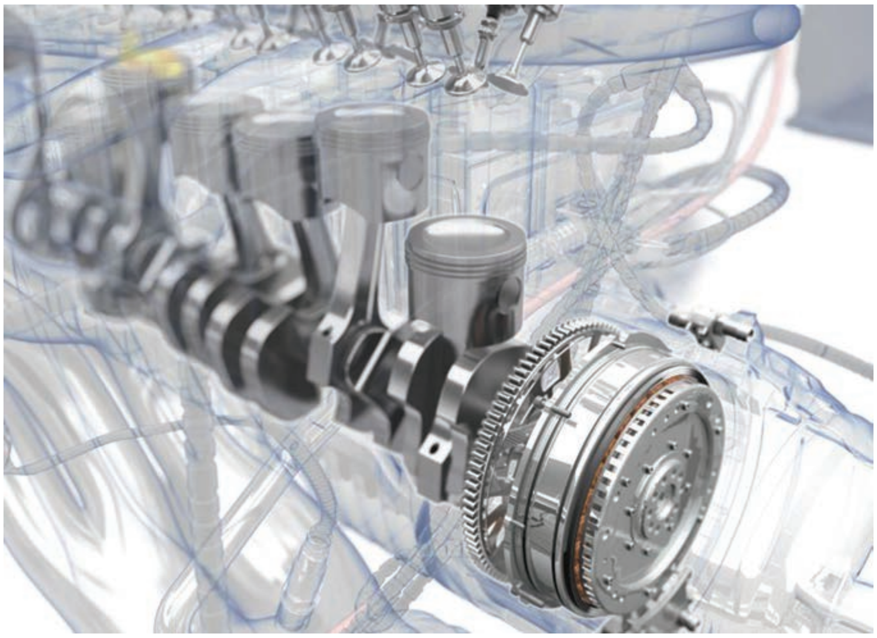


Figure 17.87 IMG in position (Source: Bosch Media)

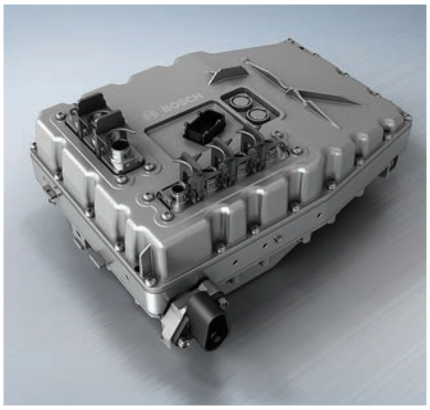


Figure 17.88 Inverter and DC-DC converter (Source: Bosch Media)

Figure 17.89 Hybrid powertrain (Source: Nissan Media)

Key fact

The V6 engine is set up to run on the Atkinson cycle to optimize efficiency with the hybrid drive.

Definition

Atkinson cycle: A type of internal combustion engine invented by James Atkinson in 1882. The Atkinson cycle is designed to provide efficiency at the expense of power density because the stroke capacities can vary.

Safety first

Caution, high voltage, this system uses a 345 V lithium-ion battery pack.

The M35 is the first hybrid vehicle for Nissan's upscale brand. It will launch approximately one year after the conventionally powered 2011 M sedan series, which includes a high-performance diesel V6 for the European market.

According to Nissan, the 1M2C drive system will provide the petrol/gasoline V6 sedan with the fuel economy of a much smaller displacement engine. The target is that of a 1.8 L vehicle. The current Nissan Fuga, on which the new M35 is based, has a kerb weight of 1780 kg (3900 lb). It is based on Nissan's current seven-speed planetary automatic but replaces the torque converter with the pair of clutches and electric motor.

The first (dry) clutch is positioned directly behind the V6 engine, which is set up to run on the Atkinson cycle to optimize efficiency with the hybrid drive. In line behind the first clutch is a 50 kW (67 hp) interior permanent-magnet e-motor which handles traction and regeneration duties. A wet clutch is fitted on the back end behind the e-motor and provides some 'slip' to help start the engine, among other duties.

In slow-speed urban driving cycles, the combustion engine shuts off and the front clutch disengages, allowing the e-motor to propel the vehicle with zero emissions. The vehicle will be capable of EV-only operation under light acceleration up to approximately 100 km/h (62 mph). When hard acceleration is called for, both clutches engage and the e-motor and combustion engine work in blended mode.

In steady-state cruise mode at high road velocities, the e-motor shuts down, both clutches are engaged, and the V6 provides the propulsion. The e-motor also serves as a generator during deceleration. In the fourth (regeneration) mode, the combustion engine shuts off, the front clutch disengages, and power generated by rear-wheel braking is sent via the rear clutch and electric motor to the 345 V lithium-ion battery pack.

Implementation of the system requires an integrated control system to simultaneously regulate battery, inverter, motor, clutches, engine, and transmission.

The battery pack contains 96 Lithium ion cells. The cells are a laminated design that improves thermal performance and enables the modules to be thin and compact. The pack's thermal stability also is enhanced by the use of manganese spinel cathodes.

Key fact

The battery pack contains 96 Lithium ion cells.

17.4 Wireless EV charging

17.4.1 Introduction

This section looks at an innovative technology, that avid readers of my books will note, I proposed as a possibility in the first edition of this book (1995, p. 302). Note to self, patent the next good idea! I referred to it then as inductive charging.

Recently however, the concept has developed to a high-level by a company called haloIPT and is known as inductive power transfer (IPT). This section is adapted from material they supplied. The company also have interesting plans to embed the technology in roads so that a vehicle can charge as it is travelling.

Definition

Patent: Exclusive rights granted by a state (national government) to an inventor or their assignee for a limited period of time in exchange for a public disclosure of an invention.

17.4.2 Inductive power transfer

Inductive power transfer (IPT) is an innovative system for wirelessly charging the batteries in electric vehicles. Electric vehicles simply park over an induction pad and charging commences automatically. IPT requires no charging poles or associated cabling. It can accommodate differing rates of charge from a single on-board unit and the rate of charge or required tariff can be set from within the vehicle. It has no visible wires or connections and only requires a charging pad buried in the pavement and a pad integrated onto the vehicle.

The system works in a range of adverse environments including extremes of temperature, whilst submerged in water or covered in ice and snow. It will operate under asphalt or embedded in concrete and is also unaffected by dust or harsh chemicals. IPT systems can be configured to power all road-based vehicles from small city cars to heavy goods vehicles and buses. Figure 17.90 shows an overview of the system.

Definition

IPT: Inductive power transfer.

Key fact

Electric vehicles simply park over an induction pad and charging commences automatically.

17.4.3 Technology overview

IPT is a technique where power at a frequency, usually in the range 20–100 kHz, can be magnetically coupled across IPT pads, which are galvanically isolated

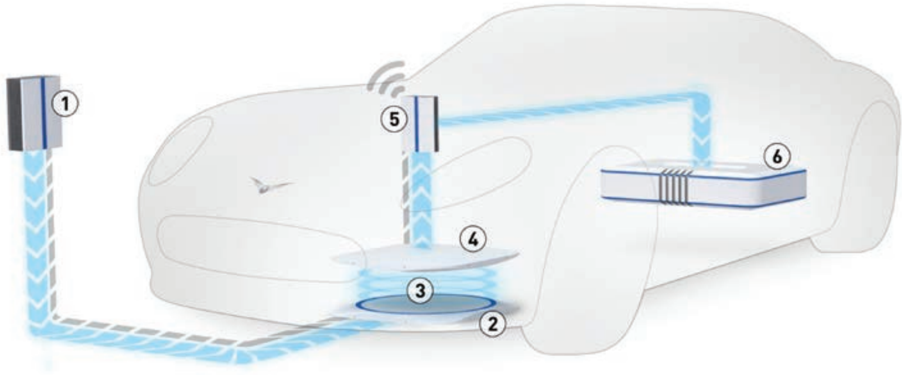


Figure 17.90 An inductive wireless charging system for statically charging an EV: 1, Power supply; 2, Transmitter pad; 3, Wireless electricity and data transfer; 4, Receiver pad; 5, System controller; 6, Battery (Source: haloIPT)

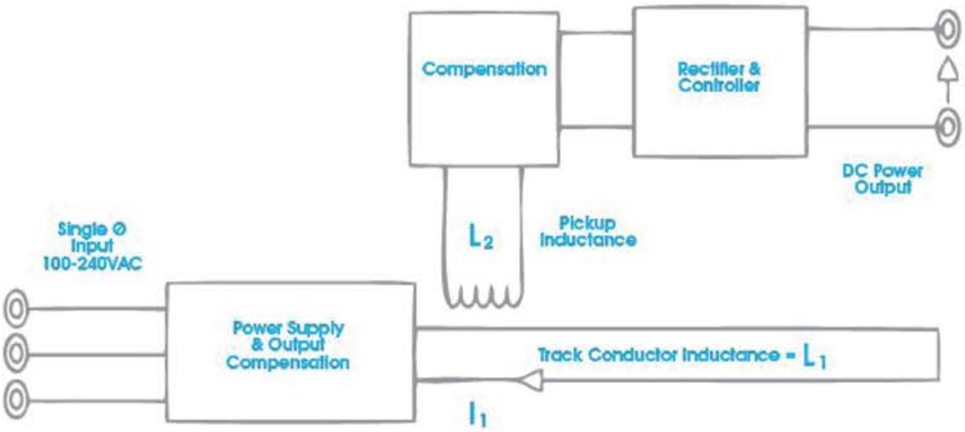


Figure 17.91 Conceptual wireless IPT charging system (Source: haloIPT)

from the original source of power. A conceptual IPT system is shown in Figure 17.91. This comprises two separate elements. A primary-side power supply, with track and a secondary-side pick-up pad, with controller.

The power supply takes electrical power from the mains supply and energizes a lumped coil, with a current typically in the range 5–125 A. Since the coil is inductive, compensation using series or parallel capacitors may be required to reduce the working voltages and currents in the supply circuitry. These capacitors also ensure an appropriate power factor.

Pick-up coils are magnetically coupled to the primary coil. Power transfer is achieved by tuning the pick-up coil to the operating frequency of the primary coil with a series or parallel capacitor. The power transfer is controllable with a switch-mode controller.

Definition

Power factor: This is defined as the ratio of the real power flowing to the load to the apparent power in the circuit. It is expressed as a percentage or a number between 0 and 1.

Real power is the capacity of the circuit for performing work in a particular time.

Apparent power is the product of the current and voltage of the circuit. The apparent power will be greater than the real power. This is because of the phase difference between voltage and current in some AC circuits. A load with a low power factor draws more current than a load with a high power factor for the same amount of useful power transferred. If capacitors are added to inductive circuits they counteract this effect and increase the power factor.

17.4.4 IPT system

A block diagram for a single-phase wireless charger is shown as Figure 17.92. The mains supply is rectified with a full-bridge rectifier followed by a small DC capacitor. Keeping this capacitor small helps the overall power factor and allows the system to have a fast start up with a minimal current surge. The inverter consists of an H-bridge to energize the tuned primary pad with current at 20 kHz. The 20 kHz current also has a 100 Hz/120 Hz envelope as a result of the small DC bus capacitor. Power is coupled to the secondary tuned pad. This is then rectified and controlled to a DC output voltage appropriate to the vehicle and its batteries. The conversion from AC to DC and back to AC, in the power supply side, is necessary so the frequency can be changed.

Definition

Inverter: An electrical device that converts direct current into alternating current.

The wireless IPT system includes three distinct hardware components:

1. High-frequency generator or power supply.
2. Magnetic coupling system or transmitter/receiver pads.
3. Pick-up controller/compensation.

The high-frequency generator takes a mains voltage input (240 V AC at 50/60 Hz) and produces high-frequency current (>20 kHz). The output current is controlled and the generator may be operated without a load. The efficiency of the generator is high at over 94% at 2 kW.

The generator comprises the following:

- mains filter (to reduce EMI);
- rectifier;

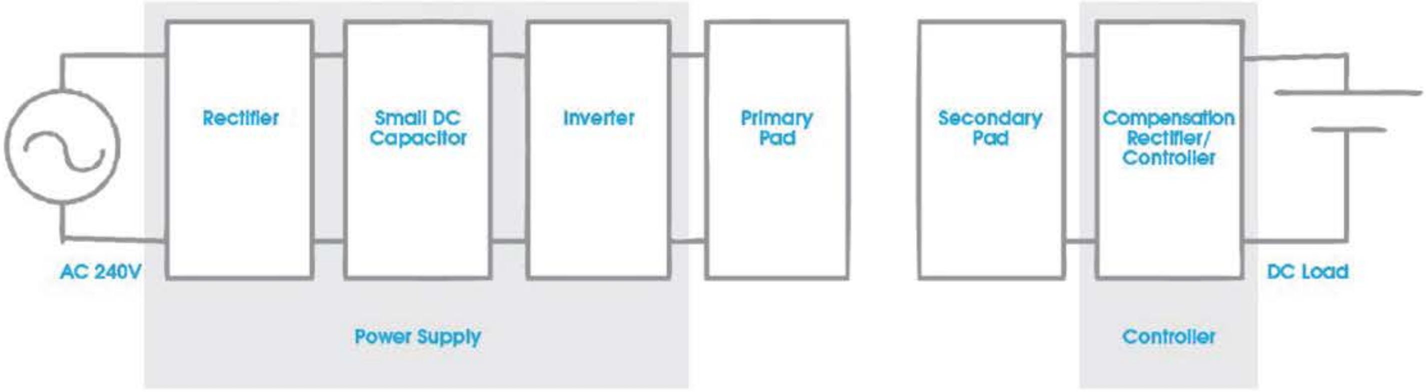


Figure 17.92 IPT system components (Source: haloIPT)

- bridge (MOSFETs) converting DC to high frequency;
- combined isolating transformer/AC inductor;
- tuning capacitors (specified for frequency and output current);
- control electronics (microcontroller, digital logic, feedback and protection circuits).

The design and construction of the transmitter and receiver pads gives important improvements over older pad topologies. This results in better coupling, lower weight and a smaller footprint, for a given power transfer and separation. The pads can couple power over gaps of up to 400 mm. The coupling circuits are tuned through the addition of compensation capacitors.

The pad construction provides shielding of magnetic fields to prevent EMI within the vehicle and ensures levels of MF exposure are within suggested international guidelines (ICNIRP). In addition, greater vertical separation and lateral tolerance, a result of superior magnetic coupling, means tolerance to misalignment is improved. This means that inch-perfect precision in parking the vehicle is not required.

A pickup controller takes power from the receiver pad and provides a controlled output to the batteries, typically ranging from 250 V to 400 V DC. The controller is required to provide an output that remains independent of the load and the separation between pads. Without a controller, the voltage would rise as the gap decreased and fall as the load current increased. The main components in the pickup controller are:

- tuning capacitors;
- transistors (MOSFETs) for power control and protection;
- rectifier;
- DC inductor;
- capacitors to smooth output voltage;
- control circuit, including sensors for voltage and current.

17.4.5 Detailed schematic

A single-phase wireless charger schematic is shown as Figure 17.93. Mains voltage input is filtered prior to rectification, using a full-bridge rectifier. This H-bridge ($S_{b1\text{-}b4}$) drives the primary side of a transformer via a DC blocking capacitor C_b, which prevents DC current flow in the transformer primary. The output of the transformer is tuned with capacitor C_1 driving the primary pad, L_1, which is magnetically coupled to the secondary pad, L_2, with a mutual inductance M.

Definition

EMI: Electromagnetic Interference.

Key fact

The coupling circuits are tuned through the addition of compensation capacitors.

Definition

ICNIRP: The International Commission on Non-Ionizing Radiation Protection issue guidelines for limiting exposure to electromagnetic, and magnetic fields (MF).

Safety first

All high voltages are completely isolated but note that safe working methods are necessary when dealing with voltages in the range 250 V to 400 V DC.

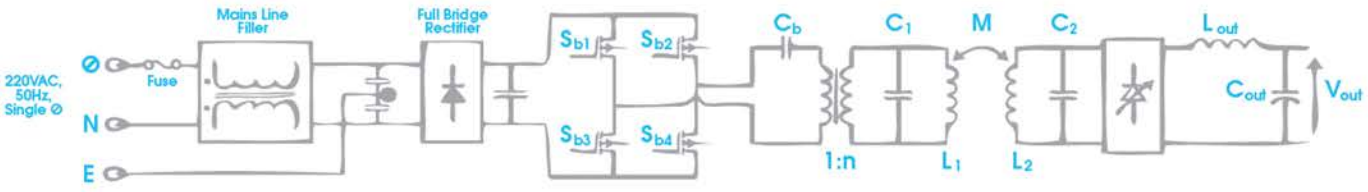


Figure 17.93 Single phase wireless EV charger schematic diagram (Source: haloIPT)

Key fact

A CANBus interface is used for control and feedback.

The secondary pad L_2 is tuned with capacitor C_2 and its output drives a controllable rectifier producing DC output. The rectifier's output is filtered using inductor L_{out} and capacitor C_{out} to obtain the output voltage V_{out} used to charge the battery. Sensors are used to measure the battery voltage and current as required for the charging process. A CANBus interface is connected to the battery management system and ECU for control and feedback purposes.

17.4.6 Battery management

The system is not designed with a specific battery management process built in. Instead the pick-up controller interfaces directly with the proprietary battery management system on the vehicle. This allows the use of precise battery management algorithms and temperature and state-of-charge sensing, to ensure battery longevity is maximized.

17.4.7 System parameters

A typical system would be defined by the following high-level parameters:

Table 17.5 High-level parameters

Parameter	Example Value
Input voltage	240 V AC
Input supply	Single Phase
Output voltage	300 V DC
Output power	3 kW
Physical separation	180 mm +/– 30 mm
Lateral tolerance	+/–150 mm
Efficiency	>85%
Power factor	>0.95
Physical size	800 mm x 400 mm x 30 mm
Output modes	CC/CV/Trickle
Output specification	Voltage variable from 250 V to 300 V in 10 V steps, power limited to 3 kW
Nominal resonant frequency	20 kHz
Communications	CANBus
Interfaces	Battery Mgt Sys ECU

17.4.8 Summary

In addition to static charging over fixed pads, HaloIPT is also developing a technology to allow power transfer to moving vehicles. This dynamic charging system utilizes transmitter and receiver systems specifically designed to transfer power to moving vehicles from modified IPT lanes. This has the potential to solve the range issues of some vehicles.

HaloIPT's vision is for an EV that can charge at home, at work and in public at differing power levels and then use the dynamic charging IPT-highway for a long intercity trip, all using the same on-car hardware. This sounds good to me.

Visit: www.haloipt.com for more information.

17.5 Advanced electric vehicle technology

17.5.1 Motor torque and power characteristics

The torque and power characteristics of four types of drive motors are represented in Figure 17.94. The four graphs show torque and power as functions of rotational speed.

A significant part of the choice when designing an EV is the drive motor(s), and how this will perform in conjunction with the batteries and the mass of the vehicle.

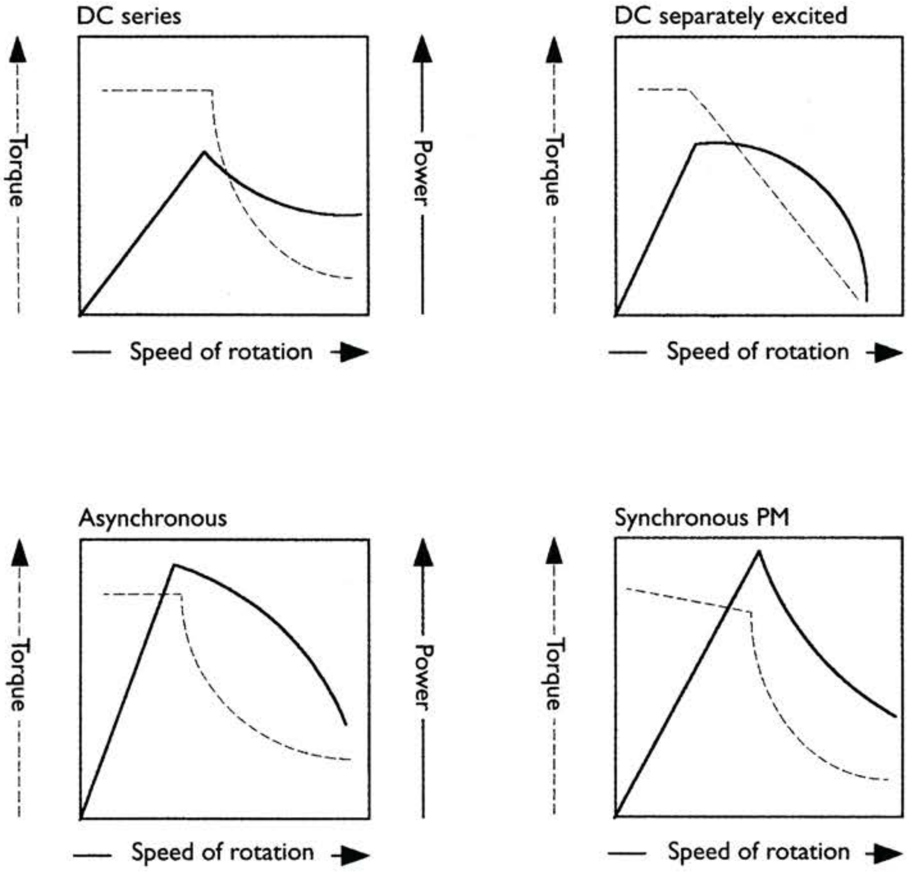


Figure 17.94 Motor torque and power characteristics

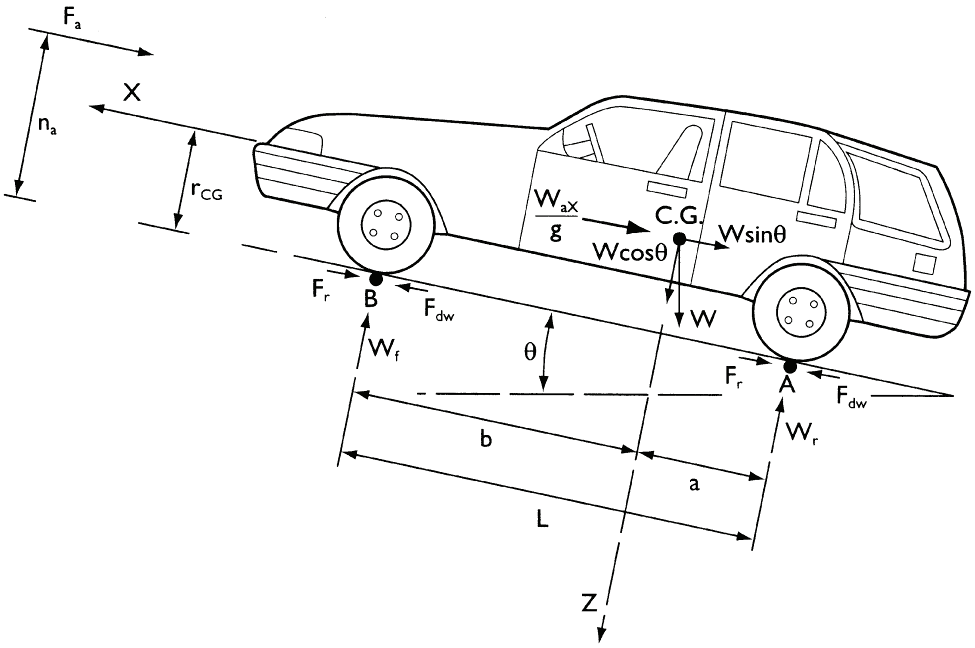


Figure 17.95 Mathematical modelling – values used for calculations

17.5.2 Optimization techniques – mathematical modelling

The effects of design parameters on the performance of an EV can be modelled mathematically. This section presents some of the basic techniques. Refer to Figure 17.95 and Table 17.6 for an explanation of the symbols.

Aerodynamic drag force:

$$F_a = \frac{\rho C_d A_f}{2}(V_v \pm V_{wind})^2$$

Rolling resistive force:

$$F_r = \mu_r mg\cos(\theta)$$

Climbing resistive force:

$$F_c = mg\sin(\theta)$$

Therefore the total resistive force is:

$$F_{resistive} = F_a + F_r + F_C$$

Force developed at the wheels:

$$F_{dw} = F_{motor}\eta_e\eta_m$$

The tractive effort therefore is:

$$F_{tractive} = F_{dw} - F_{resistive}$$

The maximum tractive force that can be developed:

$$F_{dw\ max} = \frac{\alpha\mu_a W/_L}{1+\mu_a h_{CG}/_L}$$

The effective mass of a vehicle is:

$$m_{eff} = m + \frac{J_{eff}}{r^2}$$

Acceleration time can now be shown to be:

$$t = m_{eff} \int_{v1}^{v2} \frac{dV}{F_{tractive}}$$

Power required to hold the vehicle at a constant speed:

$$Power = \frac{V_v F_{resistive}}{\eta_e \eta_m}$$

Power density of the batteries:

$$y = \frac{Power}{M_s}$$

The correlation between energy density as a function of power density can be calculated:

$$x_i = ay^5 + by^4 + cy^3 + dy^2 + ey + f$$

Range of the vehicle from fully charged batteries can be calculated from:

$$Hours = \frac{x_i}{y}$$

where $Range = V_v \times Hours$

Table 17.6 Factors and symbols used in the equations

F_a	*Aerodynamic drag force*
$\rho\tau$	*Density of air*
C_d	*Coefficient of drag e.g. 0.3 to 0.4*
A_f	*Area of the vehicle front*
V_v	*Velocity of the vehicle*
V_{wind}	*Velocity of the wind*
F_r	*Rolling resistive force*
μ_r	*Road coefficient of friction*
μ_{const}	*Tyre rolling coefficient of friction*
F_c	*Climbing resistive force*
m	*Mass of the vehicle (total)*
g	*Acceleration due to gravity*
θ	*Angle of the hill*
$F_{resistive}$	*Total resistive force*
F_{dw}	*Force developed at the driving wheels*

(Continued)

Table 17.6 *(Continued)*

η_e	*Efficiency of the electric motor*
η_m	*Efficiency of the mechanical transmission*
a	*C of G position within the wheel base*
μ_a	*Coefficient of road adhesion*
W	*Weight of the vehicle (mg)*
L	*Length of the wheel base*
h_{cg}	*Height of the vehicle's centre of gravity*
J_{eff}	*Total effective inertia of the vehicle*
M_B	*Mass of the battery*
Y	*Power density of the battery (see table 1)*
x_i	*Correlation between energy density as a function of power density*

Further calculations are possible to allow modelling – a subject which if grasped can save an enormous amount of time and money during development. The information presented here is extracted from an excellent research paper (Frantzeskakis, Krepec et al. 1994).

CHAPTER 18

Learning activities

18.1 Introduction

This section contains information, activities and ideas to help you learn more about automotive electrical and electronic systems. The best place to start is of course to read the content of the book. However, do remember when you are doing this to be active, in other words, make some notes, underline or highlight things and don't expect to understand something straight away – work at it!

I have created lots of useful online material that you can use. If you are at a school or college that is licensed to use the ATT blended eLearning package you already have access to everything. However, if not at a licensed school or college, I have created a special area for you to use free of charge.

Just go to: www.automotive-technology.co.uk and follow the links from there to find:

- Multimedia (that includes some amazing animations).
- Practical activities.
- Multiple-choice questions.
- Short answer questions.
- Glossaries.
- Virtual toolboxes.
- And more…

Figure 18.1 Learning is like combustion – once started it gets easier!

In this chapter there are some specific assessment and learning activities for each topic. Doing an assignment after each section of the book is a good way to learn more and the questions will help you check your progress.

18.2 Check your knowledge and learn more

To use these questions, you could first try to answer them without help but if necessary, refer back to the content of the chapter. Use notes, lists and sketches as appropriate to answer them. It is not necessary to write pages and pages of text! The activities here are more about learning than assessment.

18.2.1 Development of the automobile electrical system

18.2.1.1 Questions

1. State who invented the spark plug.
2. What significant event occurred in 1800?
3. Make a simple sketch to show the circuit of a magneto.
4. Who did Frederick Simms work for?
5. Which car was first fitted with a starter motor?
6. Charles F. Kettering played a vital role in the early development of the automobile. What was his main contribution and which company did he work for at that time?
7. Describe briefly why legislation has a considerable effect on the development of automotive systems.
8. Pick four significant events from the chronology and describe why they were so important.

18.2.1.2 Project

Write a short article about driving a car in the year 2020.

18.2.2 Electrical and electronic principles

18.2.2.1 Questions

1. Describe briefly the difference between ‘electron flow’ and ‘conventional flow’.
2. Sketch the symbols for 10 basic electronic components.
3. Explain what is meant by the ‘frequency response’ of an operational amplifier.
4. Draw the circuit to show how a resistor and capacitor can be connected to work as a timer. Include values and calculate the time constant for your circuit.
5. State four sources of error in a measurement system.
6. Describe how a knock sensor operates to produce a signal.
7. Make a sketch to show how a rotary idle speed actuator works and describe how it can vary idle speed when only able to take up a closed or open position.
8. Draw a graph showing the output signal of a Hall sensor.
9. Describe the operation of a permanent magnet stepper.
10. Outline the six-stage process generally required to produce a program for a computer.

18.2.2.2 Project

Discuss the developments of sensors and actuators. Consider the reasons for these developments and use examples. Why is the integration of electronics within sensors an issue? Produce a specification sheet for an anti-lock brake system wheel speed sensor, detailing what it must be able to do.

18.2.2.3 Multiple choice questions

The type of charge possessed by an electron is:

1. negative
2. positive
3. molecular
4. gravitational.

The base–emitter voltage of an NPN transistor when fully switched on is:

1. 0.3V
2. 0.6V
3. 1.2V
4. 2.4V.

Inductive sensors usually produce a:

1. square wave
2. saw tooth wave
3. sine wave
4. triangle.

The SI unit for power is the:

1. joule
2. watt
3. Newton
4. horsepower.

An electrical device, which restricts the flow of electrical current, is called:

1. an insulator
2. a conductor
3. an electrode
4. an earth connection.

The SI unit of electrical resistance is the:

1. volt
2. ohm
3. ampere
4. watt.

Ohm's law states: 'The current passing through a wire at constant temperature is proportional to the …':

1. power supplied
2. length of the circuit
3. resistance of the circuit
4. potential difference between its ends.

When comparing the current passed through a high resistance and the current passed through a low resistance, the current through a high resistance will be:

1. lower
2. higher

3. same
4. pulsing.

A component, which makes use of the magnetic effect of an electric current in a vehicle electrical system, is:

1. an ignition warning light
2. an alternator rotor
3. a fuel tank unit
4. an oil pressure gauge.

A Darlington pair of transistors is used to switch higher:

1. current
2. voltage
3. resistance
4. interest.

18.2.3 Tools and equipment

18.2.3.1 Questions

1. State five essential characteristics of an electrical test multimeter.
2. Describe why the internal resistance of a voltmeter should be as high as possible.
3. Make two clearly labelled sketches to show the waveforms on an oscilloscope when testing the output from an ignition Hall effect sensor at low speed and high speed.
4. Explain what is meant by a 'serial port'.
5. Describe briefly four advantages for the technician, of a standardized diagnostic plug.
6. State why a 'code reader' or 'scanner' is an important piece of test equipment.
7. Explain what is meant by an 'integrated diagnostic and measurement system'.
8. Using the six-stage diagnostic procedure discussed in this chapter, write out an example relating to testing a charging system.
9. Describe the meaning of accuracy in relation to test equipment.
10. List the main test connections required for an engine analyser and state the purpose of each.

18.2.3.2 Assignment

Examine how the service and repair environment has changed over the last 20 years and the different tools in use. Also, comment on what may be the situation in the next 20.

18.2.3.3 Multiple choice questions

An ohmmeter can be used to measure:

1. plug lead resistance
2. switch supply voltage
3. switch output current
4. all of the above.

Technician A says to check a switch measure the voltage at the input supply and the output. Technician B says to check a twin filament bulb use an ohmmeter to measure the resistance of the filaments. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

When looking at a waveform on an oscilloscope screen, the vertical scale represents:

1. voltage
2. time
3. current
4. resistance.

When measuring voltage, the term mV means:

1. megavolts
2. millivolts
3. microvolts
4. manyvolts.

A good multimeter, when set to read voltage, will have an internal resistance that is:

1. very low
2. low
3. high
4. very high.

A four-gas analyser will measure:

1. carbon monoxide, carbon dioxide, hydrocarbons, oxygen
2. methanol, nitrogen oxide, hydrocarbons, nitrogen
3. carbon monoxide, carbon dioxide, polycarbons, nitrogen
4. methanol, nitrogen oxide, polycarbons, oxygen.

Diagnostic procedures should always be:

1. logical
2. lunatic
3. laughable
4. laudable.

The output of most lambda sensors is a voltage that varies between:

1. 0.1 and 0.2 V
2. 0.2 and 0.8 V
3. 0.8 and 2.0 V
4. 2.0 and 12 V.

The CAN high and CAN low signals are connected to diagnostic socket pins:

1. 7 and 15
2. 4 and 16
3. 2 and 10
4. 6 and 14.

Digital scopes are usually preferred over analogue types because they can:

1. store readings
2. display quickly

3. increase reliability
4. reduce loading.

18.2.4 Electrical systems and circuits

18.2.4.1 Questions

1. Make a list of 10 desirable properties of a wiring terminal/connection.
2. Explain why EMC is such an important issue for automotive electronic system designers.
3. Describe why it is an advantage to consider vehicle systems as consisting of inputs, control and outputs.
4. Calculate the ideal copper cable size required for a fuel pump circuit. The pump draws 8 A from a 12 V battery. The maximum allowable volt drop is 0.5 V.
5. Explain what ‘contact resistance’ of a switch means.
6. State why a fuse has a continuous and a peak rating.
7. Describe the operation of a vehicle using the CAN system.
8. Explain the term ‘error checking’ in relation to a multiplexed wiring system.
9. State four types of wiring diagrams and list two advantages and two disadvantages for each.
10. Describe briefly the way in which a wiring colour code or a wiring numbering system can assist the technician when diagnosing electrical faults.

18.2.4.2 Project

Prepare two papers, the first outlining the benefits of using standard wiring looms and associated techniques and the second outlining the benefits of using a multiplexed system. After completion of the two papers make a judgement on which technique is preferable for future use.

Make sure you support your judgement with reasons!

18.2.4.3 Multiple choice questions

The output of a closed loop system has:

1. no effect on the input
2. a direct effect on the input
3. input characteristics
4. output tendencies.

A cable described as 14/0.3 will carry up to:

1. 3.75 A
2. 5.75 A
3. 8.75 A
4. 11.75 A.

A typical colour of a wire that is a main supply, according to the European code, is:

1. red
2. brown
3. black
4. white.

When discussing the amount of resistance offered by a conductor, Technician A says the greater the length of the conductor the smaller the resistance.

Technician B says the greater the cross-sectional area of the conductor the greater the resistance. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A relay can be thought of as a:

1. remote controlled switch
2. magnetic resistor
3. non-magnetic capacitor
4. heating device.

A latching device may be used on an electrical connector in order to:

1. increase resistance
2. reduce resistance
3. improve security
4. prevent security.

Technician A says the choice of cable size depends on the voltage it will have to carry. Technician B says as a rule of thumb, one strand of 0.3 mm diameter wire will carry 0.5 A safely. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A dirty electrical connection is likely to cause a:

1. high resistance
2. low resistance
3. short circuit
4. open circuit.

A multiplex wiring system will probably use:

1. three main wires
2. coaxial type wires
3. inductive type relays
4. changeover switches.

Controller area network protocols can be described as:

1. input or output types
2. high or low speed
3. reliable or limited
4. modern or old.

18.2.5 Batteries

18.2.5.1 Questions

1. Describe what a ‘lead-acid’ battery means.
2. State the three ways in which a battery is generally rated.
3. Make a clearly labelled sketch to show how a 12 V battery is constructed.
4. Explain why a battery is rated or described in different ways.

5. List six considerations when deciding where a vehicle battery should be positioned.
6. Describe how to measure the internal resistance of a battery.
7. Make a table showing three ways of testing the state of charge of a lead-acid battery together with the results.
8. Describe the two methods of recharging a battery.
9. State how the ideal charge rate for a lead-acid battery can be determined.
10. Explain why the 'energy density' of a battery is important.

18.2.5.2 Assignment

Carry out research into the history of the vehicle battery and make notes of significant events. Read further about 'new' types of battery and suggest some of their advantages and disadvantages. What are the main limiting factors to battery improvements? Why is the infrastructure for battery 'service and repair' important for the adoption of new technologies?

18.2.5.3 Multiple choice questions

A 12 volt lead-acid battery has:

1. cells connected in parallel, plates connected in series
2. cells connected in series, plates connected in parallel
3. cells connected in series, plates connected in series
4. cells connected in parallel, plates connected in parallel.

The gases given off by a lead-acid battery nearing the end of its charge are:

1. oxygen and nitrogen
2. oxygen and hydrogen
3. helium and hydrogen
4. nitrogen and hydrogen.

A lead-acid battery should be topped up with:

1. sulphuric acid
2. distilled water
3. sulphuric acid and distilled water
4. electrolyte at the correct relative density.

The electrolyte for a fully charged lead-acid battery has a relative density of approximately:

1. 1.000
2. 1.100
3. 1.280
4. 1.500.

The duration of a high rate discharge test should not exceed about:

1. 10 seconds
2. 30 seconds
3. 50 seconds
4. 70 seconds.

When a battery is disconnected, the earth lead should always be disconnected first because:

1. the circuit would still be a closed circuit
2. the mechanic could receive a shock

3. it reduces the chance of a short circuit
4. the battery will discharge quicker.

Connecting and disconnecting the battery leads with electrical systems switched on may cause:

1. a reduced risk of arcing
2. damage to electronic components
3. discharging the battery
4. low resistance connections.

When using a high rate discharge test on a 40 amp/hour capacity battery the current should be set to about:

1. 1 amp
2. 4 amps
3. 40 amps
4. 120 amps.

An ideal charge rate for a battery is:

1. 1/10th of the reserve capacity
2. 1/10th of the amp/hour capacity
3. 1/40th of the reserve capacity
4. 1/40th of the charger capacity.

When discussing the reasons why a change from 12 V to 42 V batteries is likely in the future, Technician A says this will produce an increase in power for an increased range of accessories. Technician B says this will provide an increase in power but also an increase in maintenance. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

18.2.6 Charging

18.2.6.1 Questions

1. State the ideal charging voltage for a 12 V (nominal) battery.
2. Describe the operation of an alternator with reference to a rotating ‘permanent magnet’.
3. Make a clearly labelled sketch to show a typical external alternator circuit.
4. Explain how and why the output voltage of an alternator is regulated.
5. Describe the differences between a star-wound and a delta-wound stator.
6. Explain why connecting two extra diodes to the centre of a star-wound stator can increase the output of an alternator.
7. Draw a typical characteristic curve for an alternator. Label each part with an explanation of its purpose.
8. Describe briefly how a rectifier works.
9. Explain the difference between a battery-sensed and a machine-sensed alternator.
10. List five charging system faults and the associated symptoms.

18.2.6.2 Assignment

Investigate and test the operation of a charging system on a vehicle. Produce a report in the standard format (as set out in Advanced Automotive Fault Diagnosis, Tom Denton (2000), Arnold).

Make recommendations on how the system could be improved.

18.2.6.3 Multiple choice questions

The purpose of a rectifier in an alternator is to:

1. change AC to DC voltage
2. control alternator output current
3. change DC to AC voltage
4. control alternator output voltage.

‘Star’ and ‘Delta’ are types of:

1. rotor winding
2. stator winding
3. field winding
4. regulator winding.

Technician A says an alternator rotor uses semiconductor components to rectify the direct current to alternating current. Technician B says a stator winding for a light vehicle alternator will usually be connected in a ‘star’ formation. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

The three auxiliary diodes in a nine-diode alternator provide direct current for the:

1. vehicle auxiliary circuits
2. initial excitation of the rotor
3. rotor field during charging
4. warning light simulator.

The purpose of the regulator in the charging system of a vehicle is to control:

1. engine speed
2. fuel consumption
3. generator input
4. generator output.

The function of the zener diode in the electronic control unit of an alternator is to act as a:

1. current amplifier
2. voltage amplifier
3. voltage switch
4. current switch.

The charging voltage of an engine running at approximately 3000 rpm should be:

1. 12.6 volts
2. 14.2 volts
3. 3 volts above battery voltage
4. the same as battery voltage.

Rotor windings are connected and supplied by:

1. soldered connections
2. crimped connections
3. adhesive bonding
4. brushes and slip rings.

An alternator has been dismantled and the rotor slip rings are blackened with carbon deposits. Technician A says clean them with a soft cloth and alcohol. Technician B says the rotor must be replaced. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

When fitting a new rectifier pack it is usual to:

1. remove the stator winding
2. replace the regulator
3. connect the battery lead
4. unsolder the connections.

18.2.7 Starting

18.2.7.1 Questions

1. State four advantages of a pre-engaged starter when compared with an inertia type.
2. Describe the operation of the pull-in and hold-on windings in a pre-engaged starter solenoid.
3. Make a clearly labelled sketch of the engagement mechanism of a pre-engaged starter.
4. Explain what is meant by 'voltage drop' in a starter circuit and why it should be kept to a minimum.
5. Describe the engagement and disengagement of an inertia starter.
6. State two advantages and two disadvantages of a permanent magnet starter.
7. Calculate the gear ratio of an epicyclic gear set as used in a starter. The annulus has 40 teeth and the sun gear has 16 teeth.
8. Describe the operation of a roller-type one-way clutch.
9. Make a sketch to show the speed torque characteristics of a series, shunt and compound motor.
10. Describe the difference between a lap- and a wave-wound armature.

18.2.7.2 Assignment

A starter motor has to convert a very large amount of energy in a very short time. Motors rated at several kW are in common use. However, the overall efficiency of the motor is low. A large saving in battery power would be possible if this efficiency were increased. Discuss how to improve the efficiency of the starting system. Would it be cost effective?

18.2.7.3 Multiple choice questions

The purpose of the pull-in winding in the operating solenoid of a pre-engaged starter motor is to:

1. hold the pinion in mesh
2. pull the pinion out of mesh

3. hold the pinion out of mesh
4. pull the pinion into mesh.

Technician A says a spring is used to hold a pre-engaged starter pinion in mesh when cranking the engine. Technician B says a holding coil holds the pinion in the engaged position during starting. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A one-way clutch in a pre-engaged starter motor:

1. prevents the engine driving the motor
2. prevents the motor driving the engine
3. stops the motor when the engine starts
4. starts the motor to turn the engine.

Technician A says permanent magnet starter motors are suitable for large diesel engines because of their low speed and high torque. Technician B says permanent magnet starter motors are suitable for small petrol engines because of their high speed and low torque. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

The effect of a planetary gear set fitted between the motor and drive pinion

1. modifies the speed characteristics only
2. modifies the torque characteristics only
3. modifies the speed and torque characteristics
4. has no effect on the speed or torque characteristics.

A voltmeter is connected between the main starter terminal and earth. On cranking the engine the reading should be:

1. no more than 0.5 V below battery voltage
2. approximately 0.5 V above battery voltage
3. the same as battery voltage
4. more than battery voltage.

A voltmeter connected between the starter motor body and the battery earth terminal should have a reading during engine cranking of:

1. more than 0.5 volts
2. not more than 0.5 volts
3. more than 12.6 volts
4. not more than 12.6 volts.

When fitting a phosphor bronze bush to a starter motor it is necessary to:

1. lubricate the bearing with oil before fitting
2. lubricate the bearing with grease before fitting
3. ream it to size before fitting
4. ream it to size after fitting.

The condition of starter solenoid contacts can be determined by operating the starter switch with:

1. a voltmeter connected across the solenoid contacts
2. an ammeter connected across the solenoid contacts
3. a voltmeter connected in series with the feed wire
4. an ammeter connected in series with the feed wire.

Solenoid windings may be checked for resistance with a:

1. resistance tester
2. ohmmeter
3. voltmeter
4. ammeter.

18.2.8 Ignition

18.2.8.1 Questions

1. Describe the purpose of an ignition system.
2. State five advantages of electronic ignition compared with the contact breaker system.
3. Draw the circuit of a programmed ignition system and clearly label each part.
4. Explain what is meant by ignition timing and why certain conditions require it to be advanced or retarded.
5. Make a sketch to show the difference between a hot and cold spark plug.
6. Describe what is meant by ‘mutual induction’ in the ignition coil.
7. Explain the term ‘constant energy’ in relation to an ignition system.
8. Using a programmed ignition system fitted with a knock sensor as the example, explain why knock control is described as closed loop.
9. Make a clearly labelled sketch to show the operation of an inductive pulse generator.
10. List all the main components of an ESA ignition system and state the purpose of each.

18.2.8.2 Assignment

Draw an 8 × 8 look-up table (grid) for a digital ignition system. The horizontal axis should represent engine speed from zero to 5000 rpm, and the vertical axis engine load from zero to 100%. Fill in all the boxes with realistic figures and explain why you have chosen these figures. You should explain clearly the trends and not each individual figure.

18.2.8.3 Multiple choice questions

The ignition component that steps up voltage is the:

1. capacitor
2. condenser
3. coil
4. king lead.

Setting spark plug gaps too wide will cause running problems because the firing voltage will:

1. increase and the spark duration will decrease
2. increase and the spark duration will increase
3. decrease and the spark duration will increase
4. decrease and the spark duration will decrease.

A spark is created as the coil primary winding is:

1. switched on
2. switched off
3. charged
4. stabilized.

Cruising conditions require the ignition timing to be:

1. retarded
2. reversed
3. allocated
4. advanced.

An inductive pulse generator in an ignition distributor will NOT produce an output voltage when the engine is:

1. running
2. cranking
3. stopped
4. over revving.

With the ignition switched on, a Hall effect pulse generator in an ignition distributor will produce an output voltage when the:

1. engine is running
2. engine is cranking
3. Hall chip is shielded
4. Hall chip is not shielded.

Technician A says a pulse shaper is used to shape the AC output from a pulse generator to a square wave pattern. Technician B says a Schmitt trigger is used to shape the AC output from a pulse generator to a square wave pattern. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A vehicle fitted with a system known as 'Limp Home' means that if a fault develops:

1. and you are in an ambulance, it is what you have to do if it breaks down …
2. the engine management system switches to just enough engine cylinders to keep you going
3. the driver will not even notice and the vehicle will keep going as normal
4. the engine management system switches in pre-set values to keep the vehicle driveable.

A 'hot running' engine must be fitted with a:

1. hot spark plug
2. cold spark plug
3. taper seat spark plug
4. washer seat spark plug.

Changes in pressure to a MAP sensor are converted in many cases to a:

1. variable voltage output
2. variable current output
3. steady state reading
4. steady waveform reading.

18.2.9 Fuel control

18.2.9.1 Questions

1. Explain what is meant by a lambda () value of 1.
2. State five advantages of fuel injection.
3. With reference to the combustion process, describe the effects of ignition timing.
4. With reference to the combustion process, describe the effects of mixture strength.
5. Draw a block diagram of a fuel injection system. Describe briefly the purpose of each component.
6. Explain the combustion process in a diesel engine.
7. Describe how electronic control of diesel fuel injection is achieved and state the advantages of EUI.
8. List all the main components of an electronic carburation control system and state the purpose of each component.
9. Make a clearly labelled sketch to show the operation of a fuel injector.
10. State six sources of emissions from a vehicle and describe briefly how manufacturers are tackling each of them.

18.2.9.2 Assignment

Draw an 8 × 8 look-up table (grid) for a digital fuel control system. The horizontal axis should represent engine speed from zero to 5000 rpm, and the vertical axis engine load from zero to 100%. Fill in all the boxes with realistic figures and explain why you have chosen these figures. You should explain the trends and not each individual figure.

18.2.9.3 Multiple choice questions

The ratio, by mass, of air to fuel that ensures complete and clean combustion is:

1. 14.7:1
2. 10:1
3. 1:10
4. 1:14.7

Exhaust gas products that are NOT harmful to the environment are:

1. carbon dioxide and water
2. water and carbon monoxide
3. carbon monoxide and hydrocarbons
4. hydrocarbons and oxides of nitrogen.

On an engine fitted with Electronic Fuel Injection, engine load may be determined by a:

1. MAP sensor
2. throttle position sensor
3. lambda sensor
4. vacuum capsule.

The type of petrol injection system which makes use of a single injector that sprays fuel towards a throttle is termed a:

1. single point system
2. rotary system

3. multi-point system
4. in-line system.

An injector pulse width, in milliseconds, is commonly:

1. 1.5–10
2. 1.0–30
3. 1.5–40
4. 2.0–30.

Technician A says the speed of flame spread in a diesel engine is affected by the air charge temperature. Technician B says the speed of flame spread in a diesel engine is affected by atomization of the fuel. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A valve fitted to the fuel rail in a petrol/gasoline injection system is used to:

1. bleed air
2. depressurize the system or test pressure
3. replace fuel after changing the filter
4. connect a compression tester.

Increased nitrogen oxides are formed when combustion:

1. temperatures are high
2. temperatures are low
3. speed is slow
4. speed is fast.

The function of a lambda sensor fitted in an exhaust system is to monitor:

1. carbon monoxide
2. oxides of nitrogen
3. carbon dioxide
4. oxygen.

Technician A says reduction in CO, NOx and HC has been achieved by reducing lead in fuel. Technician B says reduction in CO, NOx and HC has been achieved by using engine management systems. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

18.2.10 Engine management

18.2.10.1 Questions

1. Describe what is meant by ‘Engine management’.
2. State what the term ‘light off’ refers to in connection with catalytic converters.
3. Explain the stages of calculating ‘fuel quantity’ that take place in an ECU.
4. Make a clearly labelled sketch to show an exhaust gas recirculation system.
5. Draw a block diagram of an engine management system showing all the main inputs and outputs.
6. Describe the purpose of on-board diagnostics (OBD).
7. Make a simple sketch to show a variable length inlet manifold system.

8. State the information provided by a throttle potentiometer.
9. State four methods of reducing diesel engine emissions.
10. Explain the operation of a gasoline direct injection (GDI) system.

18.2.10.2 Assignment

1. Research the current state of development of 'lean-burn' technology. Produce an essay discussing current progress. Consider also the advantages and disadvantages of this method of engine operation. Make a reasoned prediction of the way in which this technology will develop.
2. Compare the early version of engine management with the latest Motronic or other systems and report on where, and why, changes have been made.

18.2.10.3 Multiple choice questions

Gasoline direct injection systems allow mixture in the cylinder to be:

1. homogenous
2. stratified
3. incremental
4. strong.

The main ECU 'input' parameters for calculating ignition timing and injector duration are:

1. speed and temperature
2. speed and load
3. pressure and temperature
4. pressure and load.

A throttle potentiometer provides information relating to:

1. throttle position and engine load
2. throttle position and driver intention
3. idle position and engine load
4. idle position and driver intention.

One design feature of an inlet manifold that ensures all cylinders are supplied with the same volume and air flow characteristics is the:

1. length and diameter
2. fitting of an air flow meter
3. fitting of a MAP sensor
4. material it is made from.

Atomization and distribution of fuel is generally improved if the air:

1. speed is reduced
2. pressure is reduced
3. is heated
4. is cooled.

A catalytic converter is fitted close to the exhaust manifold because:

1. it is the furthest point from the expansion box
2. it is protected from vibration
3. exhaust heat aids chemical reactions
4. exhaust gas speed is low at this point.

Measurement of exhaust emissions, just after starting the engine from cold, gives a higher than specification reading. The reason for this is:

1. the temperature of the catalyst is low
2. the catalyst is faulty
3. combustion temperature is always higher after start-up
4. compression pressures are higher after start-up.

A function that switches off the injectors during certain conditions is known as:

1. over-run fuel cut-off
2. deceleration reduction
3. under-run fuel cut-off
4. acceleration reduction.

An EGR system usually operates during:

1. cold starts
2. high vacuum conditions
3. fast accelerations
4. engine decelerations.

A correctly functioning lambda sensor will give readings between:

1. 0.002–0.008 volts
2. 0.02–0.08 volts
3. 0.2–0.8 volts
4. 2–8 volts.

18.2.11 Lighting

18.2.11.1 Questions

1. Describe briefly the reasons for fitting vehicle lights.
2. State four methods of converting electrical energy into light energy.
3. Explain the reason why headlights are fused independently.
4. Draw a simplified circuit of a lighting system showing the side- and headlight bulbs, light switch, dip switch and main beam warning light.
5. Make a clearly labelled sketch to show the 'aiming board' method of setting headlight alignment.
6. Describe the operation of a gas discharge lamp.
7. List the advantages and disadvantages of gas discharge lamps.
8. Explain the operation of infrared lighting and sketch a block diagram of the system components.
9. Define the term 'Expert or Intelligent lighting'.
10. Draw a typical dim-dip circuit and state the reason why it is used.

18.2.11.2 Assignment

Design a vehicle lighting system using technology described in this chapter. Decide which techniques you are going to use and justify your choices. For example, you may choose to use a single light source for all lights or you may decide to use neon lights for the rear and gas discharge for the front. Whatever the choice, it should be justified with sound reasons such as cost, safety, aerodynamics, styling, reliability and so on. Make sketches to show exterior views. Circuit diagrams are not necessary but you should note where components would be located. State whether the vehicle is standard or 'top of the range' etc.

18.2.11.3 Multiple choice questions

In a conventional incandescent bulb the filament is made from:

1. halogen
2. tungsten
3. quartz
4. non-resistive wire.

In a headlamp the bulb's filament position relative to the reflector ensures:

1. the correct beam direction
2. reduced electrical resistance
3. the correct beam colour
4. increased electrical resistance.

An asymmetric headlight gives a:

1. whiter light
2. dim-dip facility
3. diverging beam pattern
4. sharp cut-off line when on dip.

Technician A says dim-dip lighting is achieved with a simple series resistor. Technician B says dim-dip lighting is achieved by switching on and off fast. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

The main advantage of using light emitting diodes (LEDs) in vehicle lighting is:

1. the variety of colours available
2. that they produce whiter light
3. their long life
4. all of the above.

The wattage of a stoplight bulb is normally:

1. 5 W
2. 6 W
3. 12 W
4. 21 W.

One safety hazard associated with gas discharge lamps is related to the:

1. use of high voltages
2. use of kryptonite gas
3. length of time to cool down
4. length of time to discharge.

The headlights of a vehicle fail to illuminate when switched on. An initial visual check shows the wiring to be OK and the relay 'clicks'. Technician A says the fault is poor relay earth connection. Technician B says check the relay output. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

Correct headlamp beam alignment is necessary because:

1. it is a legal requirement
2. it ensures efficient operation
3. road safety is improved
4. all of the above.

Checking the stoplight switch can be done by removing the wires and:

1. bridging them with a jumper wire
2. bridging the switch terminals with a test lamp
3. bridging them with a voltmeter
4. bridging the switch terminals with an ammeter.

18.2.12 Auxiliaries

18.2.12.1 Questions

1. State four electrical systems considered to be 'auxiliaries'.
2. Describe briefly how a flasher/indicator unit is rated.
3. Make a clearly labelled sketch to show a typical wiper motor linkage.
4. Draw a circuit diagram of an indicator circuit, and label each part.
5. List five requirements of a wiper system.
6. Explain how off-screen parking is achieved by some wiper systems.
7. Describe what is meant by the term 'stall protection' in relation to wiper motors.
8. Draw a clearly labelled brake light circuit. Include three 21 W bulbs, a relay and fuse as well as the brake light switch.
9. Calculate the rating of the fuse required in Question 8.
10. Explain with the aid of a sketch what is meant by 'windscreen zones'.

18.2.12.2 Assignment

Investigate a modern vehicle and produce a report of the efficiency and operation of the washer and wiper systems (front and rear). Make a reasoned list of suggestions as to how improvements could be made. Consider for the purposes of lateral thinking that, in this case, money is not an issue!

18.2.12.3 Multiple choice questions

When checking the operation of a relay, an audible click is heard when the switch is operated. If there is no supply out from the relay this indicates:

1. that the relay is faulty
2. an open circuit supply
3. a faulty switch
4. all of these.

The operating frequency of an electronic flasher unit is:

1. 0.5 Hz
2. 1.5 Hz
3. 2.5 Hz
4. 3.5 Hz.

The wattage of an indicator bulb is normally:

1. 5 W
2. 6 W

3. 12 W
4. 21 W.

A wiper motor may use three brushes in order to:

1. increase torque
2. allow two speed operation
3. allow three speed operation
4. provide intermittent operation.

A thermal trip may be incorporated in a wiper motor in order to:

1. park the blades
2. protect the motor
3. provide intermittent operation
4. slow the blades in heavy rain.

When the two main brushes of a wiper motor are connected together via the limit switch, delay unit contacts and the wiper switch, this causes:

1. fast speed operation
2. slow speed operation
3. regenerative braking
4. none of the above.

Off-screen parking of wiper blades reduces:

1. current draw
2. voltage drop
3. aerodynamic drag
4. aerodynamic drop.

The delay time in a wiper control unit is set by a resistor and:

1. an inductor
2. a transistor
3. a diode
4. a capacitor.

A front screen wiper system can have:

1. only one motor
2. two motors
3. no motors
4. all of the above.

A vehicle horn produces sound because a tone disc is made to vibrate by:

1. electrostatics
2. electroplating
3. electrocuting
4. electromagnetism.

18.2.13 Instrumentation

18.2.13.1 Questions

1. State the main advantage of a thermal gauge.
2. Make a clearly labelled sketch of a thermal fuel gauge circuit.
3. Describe why moving iron and air-cored gauges do not need a voltage stabilizer.
4. Define the term, ‘driver information’.

5. Explain why digital displays are multiplexed.
6. Draw the circuit of a bulb failure system and describe its operation.
7. List five typical outputs of a trip computer and the inputs required to calculate each of them.
8. Describe with the aid of a sketch how a head-up display (HUD) operates.
9. Explain the operation of an air-cored fuel gauge system.
10. Describe what is meant by 'Telematics' and 'Telemetry'.

18.2.13.2 Assignment

Design an instrument display for a car. Choose whatever type of display techniques you want, but make a report justifying your choices. Some key issues to consider are readability, accuracy, cost and aesthetic appeal.

18.2.13.3 Multiple choice questions

When checking an NTC type temperature sensor, Technician A says remember resistance increases as temperature increases. Technician B says remember resistance decreases as temperature increases. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

One characteristic of a thermal type fuel gauge is its:

1. slow moving needle
2. almost instantaneous response
3. need for a reed switch type sensor
4. ability to be used for oil pressure measurement.

The component which prevents changes in the system voltage affecting a gauge reading is called a:

1. moving iron resistor
2. variable resistor
3. current regulator
4. voltage stabilizer.

An air-cored gauge uses the same principle as:

1. a compass needle lining up with a magnetic field
2. wind pushing a windmill blade
3. a bi-metal strip moving the needle when heated
4. none of these.

The instrument which uses pulses from the ignition primary circuit is a:

1. speedometer
2. tachometer
3. ammeter
4. odometer.

A vehicle condition monitoring system can monitor:

1. bulb operation by monitoring current drawn by the lights
2. door position by signals from switches
3. brake pad wear by contact wires in the friction material
4. all of the above.

One reason for using a dual resistance system is:

1. if one resistor breaks down the other will still operate
2. so that the circuit itself is checked
3. it reduces the operating temperature of the resistors
4. so the current flow in the circuit is increased.

The basic functions available on a trip computer include:

1. average fuel consumption, trip distance, elapsed time
2. trip distance, elapsed time, fuel remaining
3. elapsed time, fuel remaining, estimated time of arrival
4. fuel remaining, estimated time of arrival, date and time.

Technician A says advantages of LEDs are that they last a very long time and only draw a small current. Technician B says a disadvantage of LEDs is that they only produce red, yellow or green light. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

Backlighting of a liquid crystal display (LCD) is used in order to:

1. be able to read the display
2. prevent DC electroluminescence
3. display the light in a forward biased direction
4. increase vacuum fluorescence.

18.2.14 Heating ventilation and air conditioning

18.2.14.1 Questions

1. State the meaning of 'plenum chamber'.
2. Make a clearly labelled sketch to show the main components of an air conditioning system.
3. Explain the principle of refrigeration.
4. Draw a circuit showing how 'dropping' resistors are used to control motor speed.
5. Describe the operation of an air conditioning system.
6. State three potential benefits of an electrically driven air conditioning compressor.
7. Define: heat flow, radiation, convection and conduction.
8. Describe the reason for and the operation of a thermostatic expansion valve.
9. Draw a circuit of a screen heater that includes a timer relay.
10. List four functional requirements of a seat heater.

18.2.14.2 Assignment

In relation to heating and air conditioning systems, discuss why the temperature and climate settings in a vehicle may need to be changed under different external conditions to achieve the same 'perceived or ideal' feeling of comfort. Draw a block diagram of the system and add appropriate comments as to how this 'ideal' effect could be achieved.

18.2.14.3 Multiple choice questions

The function of a plenum chamber in a ventilation system is to:

1. circulate air inside the cabin
2. exhaust air from the cabin
3. hold air at a pressure higher than ambient pressure
4. hold air at a pressure lower than ambient pressure.

One reason for using recirculated air in a heating system is because it:

1. decreases warm up time
2. increases warm up time
3. reduces pollution
4. reduces traffic congestion.

Technician A says the heater motor speed is controlled by using dropping resistors. Technician B says the heater motor speed is controlled by using a variable resistor. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

The requirements of a seat heater system will include:

1. that the heater must supply more than the heat loss experienced by the person's body
2. that heat must be supplied only at the major contact points
3. that heating elements must be of a universal design to fit all seats
4. all of the above.

The refrigerant used in many air conditioning systems is known as:

1. CFC
2. Ozone
3. R134a
4. 26.3C.

The most likely position of a condenser on a car is:

1. outside the car in the air stream
2. inside the car, behind the dashboard
3. bolted to the engine
4. none of the above.

The change of state that takes place in the evaporator is:

1. solid to liquid
2. liquid to gas
3. gas to liquid
4. liquid to solid.

Which of the following is a true statement:

1. an air conditioning compressor is controlled by an electromagnetic clutch
2. an air conditioning compressor is permanently driven
3. the compressor is always mechanically controlled
4. the compressor is always manually controlled.

The component that controls the flow of refrigerant as demanded by the system is called the:

1. compressor
2. condenser

3. evaporator
4. expansion valve.

The current drawn by a blower motor running at full speed is likely to be about:

1. 0.5 amps
2. 5 amps
3. 25 amps
4. 150 amps.

18.2.15 Chassis electrical

18.2.15.1 Questions

1. Describe the three main control phases of an ABS system.
2. Describe what is meant by ‘black box fault-finding’.
3. Explain with the aid of a labelled sketch the operation of a wheel speed sensor.
4. State four advantages of electric power steering.
5. Draw a graph to show the effectiveness of traction control when only the throttle is controlled.
6. Make a simple sketch of a block diagram for an electronically controlled automatic transmission (ECAT) system and state the purpose of each part.
7. List eight chassis systems that can be controlled by electronics.
8. Define: ‘Total vehicle dynamics’.
9. Describe the operation of an active suspension system.
10. State three possible disadvantages of an ABS system.

18.2.15.2 Assignment

Investigate the possibilities of producing a vehicle with a central control unit (CCU) that is able to control ALL operations of the vehicle from engine management to instrumentation and stability control. Produce a report for the board of a major vehicle manufacturer showing the possible advantages and disadvantages of this approach. Make a clear recommendation to the board as to whether they should make this idea into a reality – or not. Justify your decision.

18.2.15.3 Multiple choice questions

Technician A says an anti-lock braking system must recognize poor road conditions, such as when a vehicle is aquaplaning and react accordingly. Technician B says an anti-lock braking system can increase the stopping distance when on poor road conditions, such as loose gravel or snow. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

The task of an ABS control module is to compare signals from wheel speed sensors by determining wheel:

1. speed
2. deceleration
3. linear speed
4. percentage slip.

When a wheel locks during the braking of a vehicle fitted with ABS, the modulator action will be:

1. release, hold, build-up
2. hold, build-up, release
3. build-up, release, hold
4. none of the above.

An oscilloscope connected to a wheel speed sensor should show a:

1. sine wave pattern
2. cosine wave pattern
3. high resistance
4. low resistance.

On most vehicles, disconnecting the ABS fuse for 10 seconds will:

1. disable the ABS
2. de-activate the ABS
3. reset the ABS fault memory
4. test the ABS sensors.

Electronically controlled automatic transmissions can prevent surging by controlling:

1. hydraulic vacuum
2. hydraulic pressure
3. feedback vacuum
4. feedback pressure.

Electric power steering that does not have a mechanical connection between the steering wheel and the front wheels is known as:

1. a crazy idea
2. a good idea
3. steer-by-wire
4. scare-by-wire.

A system that improves the grip of driven wheels when accelerating is known as:

1. ABS
2. ECU
3. TCR
4. ECAT.

Brake assist systems help to apply the brakes under:

1. all conditions
2. inclement conditions
3. anti-lock conditions
4. emergency conditions.

An electronically controlled clutch:

1. reduces wear and improves performance
2. reduces wear but reduces performance
3. increases wear but improves performance
4. increases wear and reduces performance.

18.2.16 Comfort and safety

18.2.16.1 Questions

1. State what is meant by active and passive safety.
2. Draw a simple motor reverse circuit and explain its operation.

3. Describe briefly six features of a high-end ICE system.
4. State five sources of radio interference.
5. Explain why fault-finding sometimes involves ‘playing the odds’.
6. Describe the operating sequence of a driver’s airbag.
7. Define ‘Latching relay’.
8. Describe, with the aid of a block diagram, the operation of a cruise control system.
9. State four advantages of an intelligent airbag.
10. Explain the key features of a top-end alarm system.

18.2.16.2 Assignment

Produce a report on the main issues connected with active cruise control. A good technique for starting on this type of assignment is to ask the question: ‘What is good or bad and who gains and who loses?’

18.2.16.3 Multiple choice questions

An electric window has a Hall type sensor fitted. Technician A says this is used to determine the window position. Technician B says this is part of the ‘bounce back’ safety feature. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

A window lift motor drives through a worm gear because this:

1. increases speed and torque
2. reduces speed and torque
3. increases speed and reduces torque
4. reduces speed and increases torque.

The frequency reproduction from a ‘tweeter’ type speaker would be described as:

1. high
2. middle range
3. low
4. very low.

In order for a radio to interrupt listening and broadcast traffic announcements it will receive signals described as:

1. AM
2. RDS
3. CD
4. PC.

When discussing ways in which to disable a vehicle to prevent theft, Technician A says two ways to do this are ignition circuit cut-off and fuel system cut-off. Technician B says starter circuit cut-off and engine ECU code lock. Who is right?

1. A only.
2. B only.
3. Both A and B.
4. Neither A nor B.

Which of the following would provide an input signal to an alarm system:

1. volumetric sensor
2. volumetric transmitter
3. ignition immobilizer
4. unbroken loop circuit.

Which of the following would be regarded as a passive safety feature:

1. airbag
2. seat-belt
3. belt tensioner
4. all of the above.

Which of the following would be regarded as an active safety feature:

1. good road holding
2. side airbags
3. seat-belt tensioner
4. all of the above.

Following a frontal impact, the time taken to fully inflate an airbag will be approximately:

1. 10 ms
2. 20 ms
3. 30 ms
4. 40 ms.

To prevent the risk of accidental deployment of an airbag:

1. remove the SRS fuse and wait 10 minutes
2. remove the SRS fuse and discharge the capacitors manually
3. wait 10 seconds and remove the SRS fuse
4. wait 10 seconds and discharge the capacitors manually.

18.2.17 Alternative fuel, hybrid and electric vehicles

18.2.17.1 Questions

1. State what is meant by ZEV.
2. Describe briefly the term ‘Hybrid’.
3. Explain what is meant by, and the advantages of, inductive charging.
4. Describe with the aid of sketches the different ways in which a hybrid vehicle can be laid out.
5. Explain the term ‘Power density’.
6. List five types of EV batteries.
7. Describe with the aid of a sketch the operation of a synchronous induction motor.
8. State four types of EV drive motor.

18.2.17.2 Assignment

A question often posed about so-called ZEVs: as the electricity has to be generated at some point, often from burning fossil fuels, then how can they be said to produce no emissions? Research and comment on this issue.

18.3 Simulation program

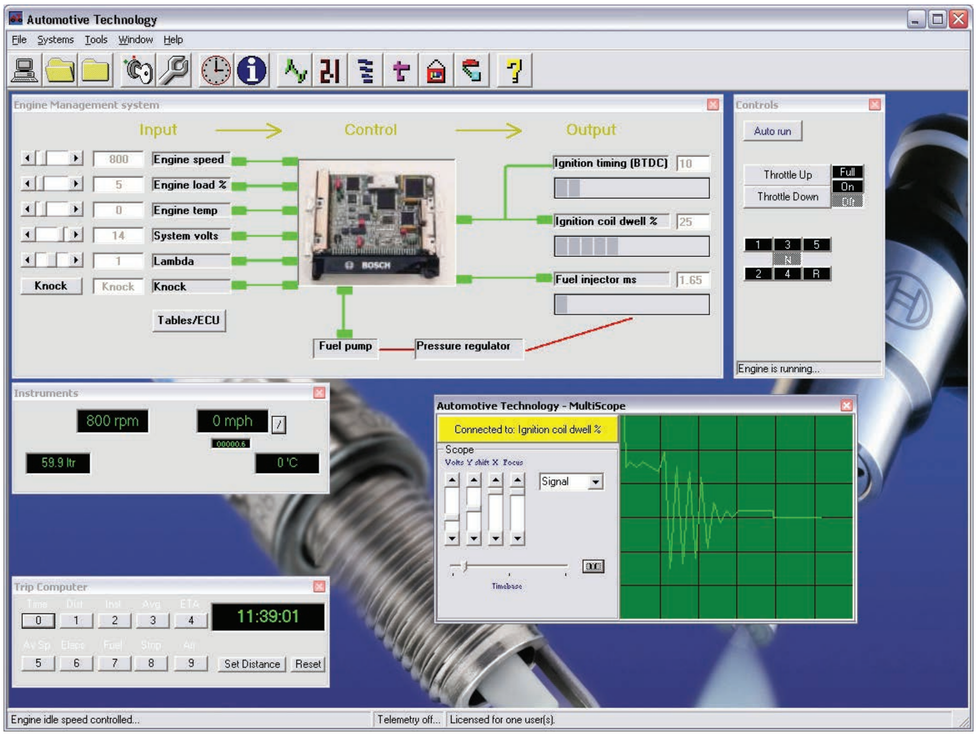
The Automotive Technology (AT) program is all about learning how complex automotive systems work – and how to fix them when they do not! AT Electronics helps you learn how systems (engine management in particular) operate, how the inputs to a system affect its outputs, and what the effects are when a fault occurs. Diagnostic routines, which are built into the program, will allow you to put into practice some of the skills you develop but ensure that you work in a logical way.

The MultiScope feature allows you to examine signals from sensors and those supplied to actuators. It also contains a scanner and multimeter to show typical readings. A telemetry screen, text and pictures window can also be used. Learning tasks, which are part of the help file, will help you work your way through the program.

The program allows you to control the inputs to systems and note the effect this has on the outputs. In this way you will start to understand the operation of automobile electronic systems. Figure 18.2 shows the charging system simulation. In this case an example of the inputs would be engine/alternator speed and an example of the outputs would be the system voltage.

Diagnostics are possible by creating a fault and carrying out tests to locate it! A database is built into the program to assist with this and MultiScope has lots of functions to help. The methods used are appropriate for use on real systems. This is an ideal training system for trainees and students. The main simulation windows relate the engine management, starting and charging – but others are ‘under construction’.

The shareware program can be downloaded from www.automotive-technology.co.uk



800 Engine speed

5 Engine load %

0 Engine temp

14 System volts

1 Lambda

Ignition timing (BTDC) 10

Ignition coil dwell % 25

Fuel injector ms 1.65

Engine is running...

800 rpm

0 mph

00000.6

59.9 ltr

0 'C

11:39:01

Connected to: Ignition coil dwell %

Engine idle speed controlled... Telemetry off... Licensed for one user(s).

Figure 18.2 ‘It is all about INPUTS and OUTPUTS’

18.4 Last word

Key fact

Modern motor vehicles are highly sophisticated machines that incorporate the latest developments in electrical, electronics, software and mechanical engineering.

Modern motor vehicles are highly sophisticated machines that incorporate the latest developments in electrical, electronics, software and mechanical engineering. They are a marvel of modern engineering practice and truly show how all these technologies can be integrated and harmonized for maximum benefit to the end user. It is clear that this level of technology produces the safest, quietest and most efficient road vehicles we have ever known. Diagnostic systems are also advancing. Further reading: Advanced automotive fault diagnosis (Denton 2011).

Well, that's it, if you have arrived here, after having read all the book, done all the assignments, completed all the practical tasks, used the website: www.automotive-technology.co.uk resources and can remember everything, then well done.

Or did you just start reading the book from the back?

References

Ayala, D., J. Lin, et al. (2010). *Communication Reduction for Floating Car Data-based Traffic Information Systems*, IEEE.

Denton, T. (2010). *Advanced Automotive Fault Diagnosis*. Oxford, Elsevier, Oxford.

Dohle, U. (2003). *New Generations of Injection Systems: Piezoelectrics and More Makes Diesel Even Cleaner and More Fuel Efficient*, Robert Bosch GmbH.

Flipsen, S. (2006). 'Power sources compared: the ultimate truth?' *Journal of power sources* **162**(2): 927-934.

Frantzeskakis, P., T. Krepec, et al. (1994). 'Specific analysis on electric vehicle performance characteristics with the aid of optimization techniques.' *SAE transactions* **103**: 339-339.

HaloIPT. (2011). Retrieved 20/04/2011, from www.haloipt.com.

Heitzer, H.-D. (2003). *Development of a Fault Tolerant Steer-By-Wire Steering System*, AutoTechnology.

Huber, W., M. Lädke, et al. (1999). *Extended Floating-car Data for the Acquisition of Traffic Information*.

Kelling, N. and P. Leteinturier (2003). *X-by-wire: Opportunities, Challenges and Trends*, SAE.

Korolov, M. (2011). 'Driving Monitor.' *Treasury & Risk*: 14-14.

McLaren. (2011). 'ATLAS.' Retrieved 10/05/2011, from http://www.mclarenelectronics.com/Products/All/sw_atlas.asp

Mindl, P. (2003). *Super-capacitor in Hybrid Drive*, CVUT FEL Research Centre.

Plapp, G. and J. Dufour (2003). *New Functions for Brake Control Systems*, Robert Bosch GmbH.

Steckmann, K. (2009). *Extending EV Range with Direct Methanol Fuel Cells*, World Electric Vehicle Journal.

Tesla Motors. (2011). Retrieved 20/03/2011, from www.teslamotors.com.

Transonic Combustion. (2011). Retrieved 09/03/2011, from www.tscombustion.com.

This page intentionally left blank

Index