Automotive Microelectronics

BOSCH

Automotive Technology

- Basic principles
- Components
- Development and production

Imprint

Published by:
© Robert Bosch GmbH, 2002
Postfach 30 02 20,
D-70442 Stuttgart.
Automotive Aftermarket Business Sector,
Department AA/PDT5.
Product Marketing,
Diagnostics, Test Equipment.

Editor-in-chief:
Dipl.-Ing. (FH) Horst Bauer.

Editorial team:
Dipl.-Ing. (BA) Jürgen Crepin,
Dipl.-Ing. Karl-Heinz Dietsche,
Dipl.-Holzw. Folkhart Dinkler.

Authors:
Dipl.-Ing. Adolf Fritz
(Control Unit Development),
Dr. rer. nat. Ulrich Goebel
(Hybrid Circuit Board Production),
Armin Hess
(Example Circuit),
Karlheinz Isert
(Printed Circuit Board Production),
Dr. rer. nat. Ulrich Schaefer
(Basic Principles, Components, Microcontroller)

and the editorial team in co-operation with the responsible technical departments of Robert Bosch GmbH.

Unless otherwise stated, the above are all employees of Robert Bosch GmbH, Stuttgart.

Printed in Germany.
Imprimé en Allemagne.

1st Edition, August 2001.
English translation of the German edition dated: February 2001.
(1.0)

Automotive microelectronics

Robert Bosch GmbH

Contents

Motor vehicles are now inconceivable without electronic control units and their associated sensors and actuators. Such components have revolutionised the automotive world. All essential vehicle functions are electronically controlled by systems and components that occupy only a tiny space. There are systems that control the function of engine and gearbox, the safety and security systems and a large number of comfort and convenience systems, and the number of "intelligent" vehicle systems is growing at a rapid pace. Nor is there any foreseeable end to this trend.

Electronic control systems open the door to a vast array of possibilities. They improve vehicle safety and ride comfort. At the same time, they make vehicles more economical and kinder to the environment.

The triumphant advance of electronics has created a vocabulary of terms with which we are bombarded every day. It is becoming more and more difficult to understand what precisely those terms refer to.

This publication in the Bosch Yellow Jacket "Technical Instruction" series explains the specialist terminology of microelectronics in detail.

It starts with a short introduction followed by an explanation of the principles of semiconductor technology and how microelectronic components interact. A practical example of a circuit helps to consolidate the theoretical concepts described. This is then followed by a description of the development and production of control units.

Finally, the glossary provides a quick guide to the most important microelectronics terms.

The functions of the individual electronic systems (e.g. MED, EDC, ESP) are described in detail in separate publications in the Yellow Jacket "Technical Instruction" series.

Automotive microelectronics

Microelectronics have revolutionised motor vehicle technology. Initially, mechanical components were replaced by electronic devices in order to make systems more reliable. This was the case with the contact breaker points in the conventional coil ignition system, for example. Gradually, however, more and more new vehicle systems were introduced which simply would not have been possible without the use of electronics. The impetus for these new developments was provided by increasingly demanding requirements placed on the exhaust-emission characteristics of the internal-combustion engine (e.g. emission-control systems), on comfort and convenience (e.g. climate control and navigation systems) and on safety (e.g. antilock braking system [ABS] and airbags).

Overview

Cars and commercial vehicles equipped with the latest available technical improvements are absolutely brimming with electronic systems. Those systems can be subdivided into the following areas of application:

- Engine and drivetrain
- Safety
- Comfort and convenience
- Communication and multimedia.

Fig. 1 provides an overview of the electronic systems that can be found on modern motor vehicles. Many of the systems referred to are now standard equipment on all new vehicles. By contrast, there are others that remain (as yet) the preserve of the most expensive luxury models.

Electronic systems can be subdivided into the following functional areas:

- Sensors and setpoint generators
- Control units (ECUs)
- Actuators
- ECU communication links (networks) and
- Electronic diagnosis

Sensors and setpoint generators

Sensors detect operating conditions (e.g. engine speed, wheel speed, temperature). They convert physical variables into electrical signals. Setpoint generators (e.g. controls operated by the driver) specify desired settings.

Control units (ECUs)

Control units process the information received from the sensors and setpoint generators using specific mathematical calculation sequences (control algorithms). They control the actuators by means of electrical output signals. Control units also form the interface with other systems and the vehicle diagnostics.

Actuators

Actuators convert the electrical output signals from the control unit into physical variables. Examples of actuators are:

- gasoline-engine fuel injectors
- diesel-engine fuel injectors
- electric motors (e.g. for driving the power-window regulator mechanism or as throttle-valve positioner on vehicles with ETC (Electronic Throttle Control))
- fans

Networks

As the number of electronic systems in vehicles grows, so does the amount of interconnecting wiring needed. The total length of the wiring in the wiring harness of a mid-range car is now roughly 1.6 km on average and incorporates up to 300 connectors with a total of around 2000 connector pins.

Networking of the various systems reduces the overall length of the wiring required. A shared data bus consisting of only two wires (e.g. CAN bus) carries data that is read by all bus users. Depending on the specific requirements, a vehicle may also have separate busses for engine and drivetrain, comfort and convenience systems, and communication systems.

Motor vehicles are now inconceivable without electronic control units and their associated sensors and actuators. Such components have revolutionised the automotive world. All essential vehicle functions are electronically controlled by systems and components that occupy only a tiny space. There are systems that control the function of engine and gearbox, the safety and security systems and a large number of comfort and convenience systems, and the number of "intelligent" vehicle systems is growing at a rapid pace. Nor is there any foreseeable end to this trend.

Electronic control systems open the door to a vast array of possibilities. They improve vehicle safety and ride comfort. At the same time, they make vehicles more economical and kinder to the environment.

The triumphant advance of electronics has created a vocabulary of terms with which we are bombarded every day. It is becoming more and more difficult to understand what precisely those terms refer to.

This publication in the Bosch Yellow Jacket "Technical Instruction" series explains the specialist terminology of microelectronics in detail.

It starts with a short introduction followed by an explanation of the principles of semiconductor technology and how microelectronic components interact. A practical example of a circuit helps to consolidate the theoretical concepts described. This is then followed by a description of the development and production of control units.

Finally, the glossary provides a quick guide to the most important microelectronics terms.

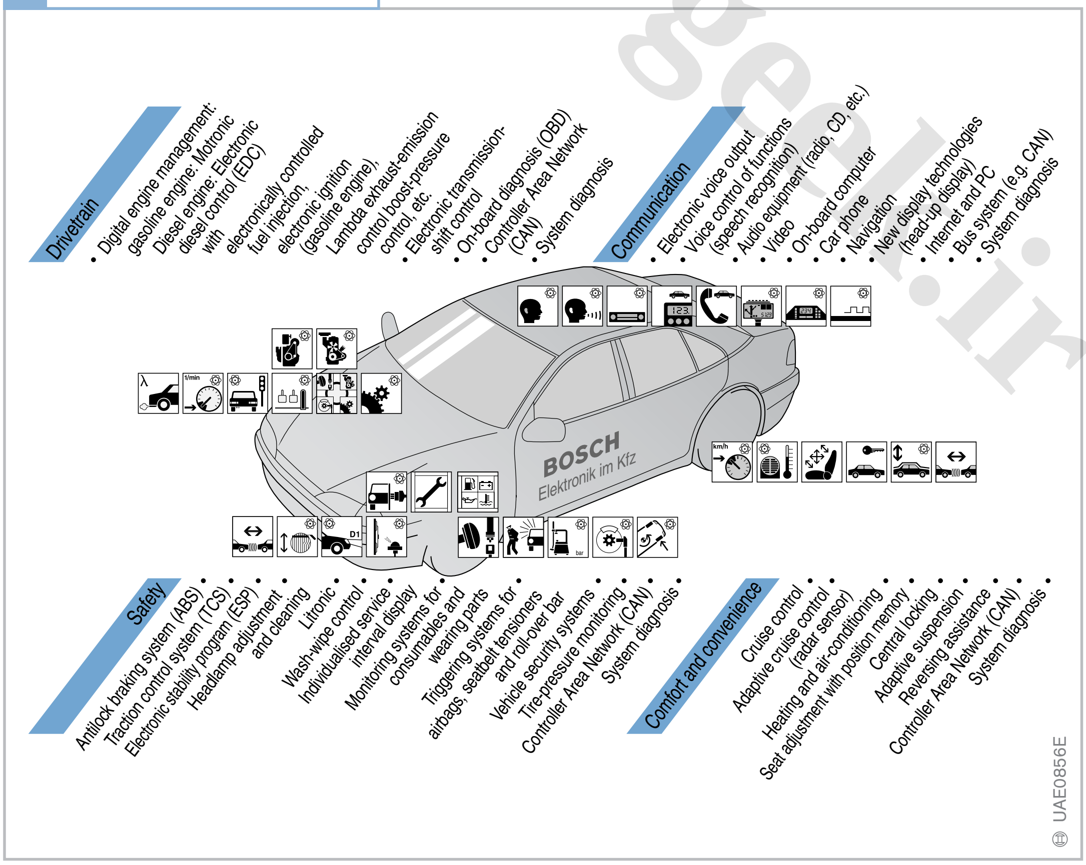
The functions of the individual electronic systems (e.g. MED, EDC, ESP) are described in detail in separate publications in the Yellow Jacket "Technical Instruction" series.

Another advantage of the bus system is that sensor signals need only be analysed by a single control unit. For example, the instrument cluster can use the individual wheel-speed signals delivered by the ABS control unit to calculate the vehicle's road speed. This complex calculation even takes account of cornering differences and spinning wheels. The road-speed signal is transferred on the CAN bus to all other users (e.g. the ABS control unit which uses it for controlling brake application, the engine-management ECU which uses it as the basis for cruise control, or the car radio for the speed-related volume adjustment).

Electronic diagnosis

Electronic diagnosis functions on the control unit continuously monitor the operation of the system and its components. Any faults that occur (e.g. short circuits in the wiring, sensor failure) are stored in the control unit's fault memory. Those stored faults can then be read out in the course of service using a dedicated system tester that is connected to the control unit's diagnosis interface. The system tester can also be used to scan sensor signals and send commands to actuators to perform specific operations. Electronic diagnosis thus enables faults to be located more quickly and easily in the course of vehicle service.

1 Electronics in Motor Vehicles (Cars)



Demands on electronic systems

Electronic systems in motor vehicles are exposed to extreme stresses (e.g. due to extreme temperature variations, unusual climatic conditions, poor road surfaces and the effect of corrosive substances). These are some of the requirements they must meet in order to be able to function reliably and without faults over long periods:

- Resistance to temperatures ranging from –40 °C ... 125 °C
- EMC (electromagnetic compatibility): immunity to external interference (e.g. mobile phone signals) and no emission of electromagnetic radiation likely to cause interference on other equipment
- Resistance to shocks and vibration
- Resistance to water and damp
- Resistance to corrosive fluids (e.g. oils and salt-water spray)
- Light weight
- Economical production costs and
- Secure and trouble-free mounting

History of development

The amount of electronic equipment in motor vehicles is continually increasing. Fig. 2 provides an overview of the growth of electronic equipment expressed as a proportion of vehicle cost.

Because of their cost, electronic systems were initially reserved for vehicles at the luxury end of the market. This explains why in 1980 electronic equipment accounted for only half a percent of vehicle cost. From that time onwards, and particularly in the 1990s, that proportion grew rapidly as the price of electronic equipment continually dropped with the result that more and more systems could be fitted to mid-range and even small cars.

Gradually, more and more electronic systems were fitted to motor vehicles (Table 1). And the trend continues.

The new science of mechatronics deals with the interaction between mechanical, electronic and data processing devices.

2 Proportion of cost per vehicle (overall average)

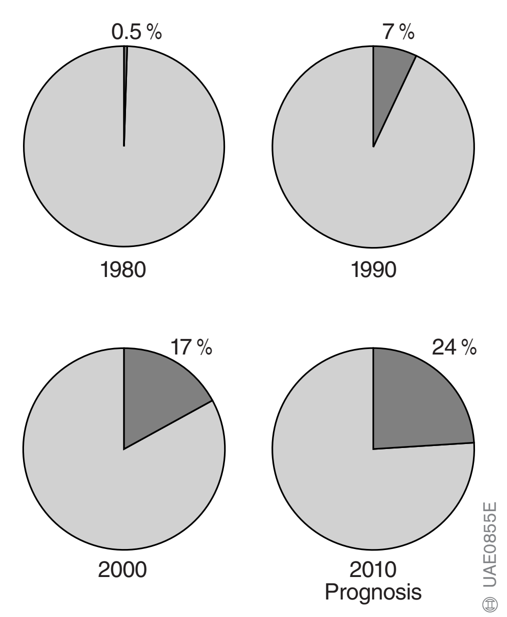


Table 1

1 Historical development of electronic systems in motor vehicles (examples)

1958	DC generator with variode
1962	3-phase alternator with variode
1965	Transistorised ignition
1967	D-Jetronic gasoline-injection system (pressure-controlled)
1978	Antilock braking system (ABS)
1979	Motronic (combined ignition and fuel-injection system)
1982	Electronic ignition system
1982	Knock control
1986	Electronic diesel control (EDC)
1986	Electronic throttle control (ETC)
1987	Traction control system (TCS)
1989	Electronic transmission-shift control (“stand-alone” system Tiptronic)
1989	CAN (Controller Area Network)
1989	Vehicle navigation system (Travelpilot)
1991	Litronic
1994	ME-Motronic (integrated ETC)
1997	Electronic stability program (ESP)
2000	MED-Motronic (gasoline direct injection)
2000	Adaptive Cruise Control

Miniaturisation

Transistorised ignition was first used on gasoline engines in 1965. It did away with the negative effect on ignition timing accuracy of erosion caused by electrical arcing between the contact-breaker points. The transistor had gained a foothold in the motor vehicle, heralding the start of the electronic age. But is wasn't until electronic componentry was miniaturised that the decisive step was taken in making electronic systems in automobiles capable of the levels of performance that are taken for granted today. Enormous advances in miniaturisation were made in the area of semiconductor components in particular, making it possible to integrate more and more functions within a component that occupied only a tiny amount of space. Every ECU contains microcontrollers that combine millions of transistor functions on semiconductor chips that take up only a few square millimetres.

It has also been possible to substantially reduce the dimensions of power components such as output stages for controlling actuators. For example, multiple ignition output stages are now combined in a single component. This means that an external ignition output stage is no longer required. It is now integrated in the engine-management ECU. Consequently, the external ignition output stage previously used can be dispensed with. The associated reduction in the number of components also improves the reliability of the system.

The size of discrete components (resistors, capacitors) has similarly been significantly reduced. SMDs (**S**urface **M**ounted **D**evices) are soldered or bonded to the circuit board without wire connections.

In spite of the continual growth in the number and complexity of the functions performed, the miniaturisation of electronic equipment has meant that the size of the ECUs continues to shrink (Fig. 3).

Memory capacity

Whereas a memory capacity of 4 kilobytes was adequate for the modest requirements of, for instance, a management system for a gasoline engine in the late 1970s, 10 years later the figure had reached 30 kilobytes. The incorporation of more and more functions in the engine-management ECU led to an explosion in the demand for memory capacity. By the year 2000, the required capacity had reached 500 KB. Other automotive electronic systems have followed a similar pattern of development. And there is no foreseeable end to this trend.

3 Gasoline-engine ECUs

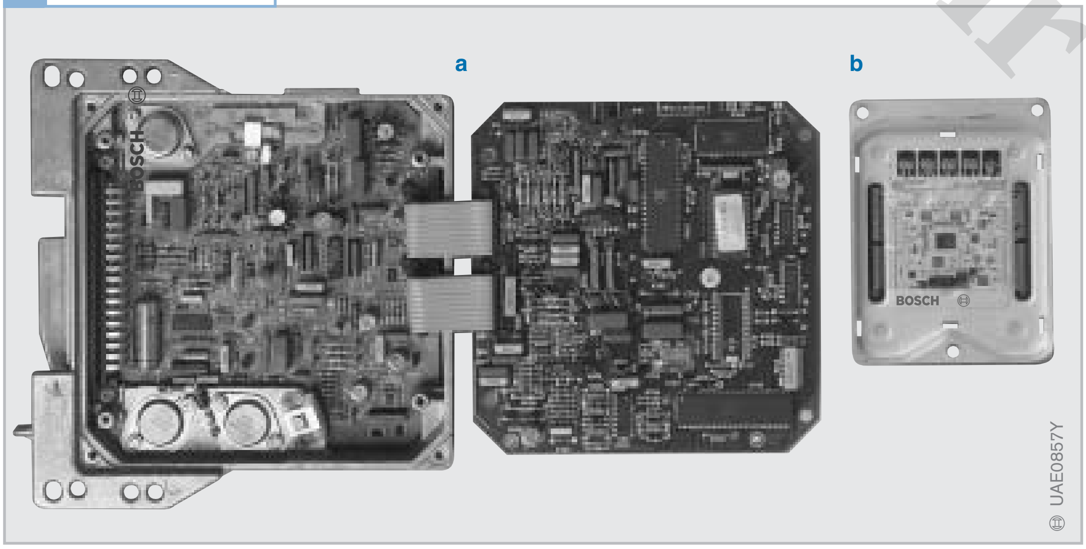


Fig. 3
a 1979 Jetronic ECU with 290 components and a weight of 1.14 kg
b 1996 Motronic hybrid ECU with 82 components and a weight of 0.25 kg

Basic principles of semiconductor technology

Semiconductors occupy a unique position between conductors and non-conductors. The electrical conductivity of semiconductors is dependent on pressure, temperature, intensity of incident light or the proportion of foreign atoms in the semiconductor material. Those properties are the basis for semiconductor technology.

Terminology

Electronics

According to the definition of the IEC (International Electrotechnical Commission), electronics is the branch of science and technology that deals with the study and utilisation of the physical phenomena in gases, solids and vacuums that are related to the flow of electricity.

Microelectronics

According to DIN 41 857, microelectronics is a branch of technology that deals with the conception, design, technology, manufacture and use of highly miniaturised electronic circuits.

It is evident from that definition that microelectronics is concerned with miniature circuits made up of multiple individual components rather than miniaturised discrete components (i.e. components with clearly definable individual functions). Only integrated film and semiconductor circuits and composite microcircuits (hybrid circuits) fall into that category.

Electrical conductivity

The specific suitability of different materials for conduction of electricity is determined by the number and mobility of free charge carriers that they contain. The electrical conductivity of solids at room temperature can vary by 24 powers of ten between different materials.

Solids are subdivided into three classes of material according to their electrical conductivity (Table 1).

Table 1

1 Classes of material based on conductivity (with examples)

Conductors (Metals)	Non-conductors (Insulators)	Semiconductors
Silver	Teflon	Silicon
Copper	Quartz glass	Germanium
Aluminum	Aluminum oxide	Gallium arsenide

All solids contain around 10^{22} atoms per cubic centimetre which are held together by electrical forces.

Conductors (Metals)

In metals, the number of free charge carriers is very large (one or two free electrons per atom). Their level of mobility is moderate. The electrical conductivity of metals is high. In good conductors it can be as much as 10^6 siemens/cm.

Non-conductors (Insulators)

In insulators, the number of free charge carriers is practically zero and consequently the electrical conductivity virtually nonexistent. The conductivity of good insulators is of the order of 10^{-18} siemens/cm.

Semiconductors

The electrical conductivity of semiconductors is somewhere between that of conductors and insulators. Under normal conditions they also have a very small number of free charge carriers, but that number can be substantially increased by the application of energy from an external source. Their conductivity, therefore – in contrast to that of metals and insulators – is heavily dependent on:

- pressure (affects the mobility of charge carriers),
- temperature (affects the number and mobility of charge carriers),
- exposure to light (affects the number of charge carriers), and
- added impurities (affects the number and type of charge carriers).

The sensitivity of semiconductors to pressure, temperature and light makes them suitable for use as sensors.

The ability to accurately modify and localise the conductivity of semiconductors by the controlled introduction of impurities that affect electrical properties (doping) is the basis of semiconductor technology. The electrical conductivity that can reliably be brought about in silicon by doping ranges from 10^4 to 10^{-2} siemens/cm.

As silicon is by far the most important semiconductor material, the explanations that follow will restrict themselves exclusively to that material. When solid, silicon consists of a crystal lattice in which each silicon atom is linked to four equally spaced adjacent atoms. Every silicon atom has four outer electrons (Fig. 1). Each pair of adjacent atoms is linked together by two shared electrons. In such a perfect crystal lattice, therefore, there are no free charge carriers, which means that the silicon is a non-conductor. This condition is changed fundamentally by the introduction of suitable impurities (doping), or energy from an external source.

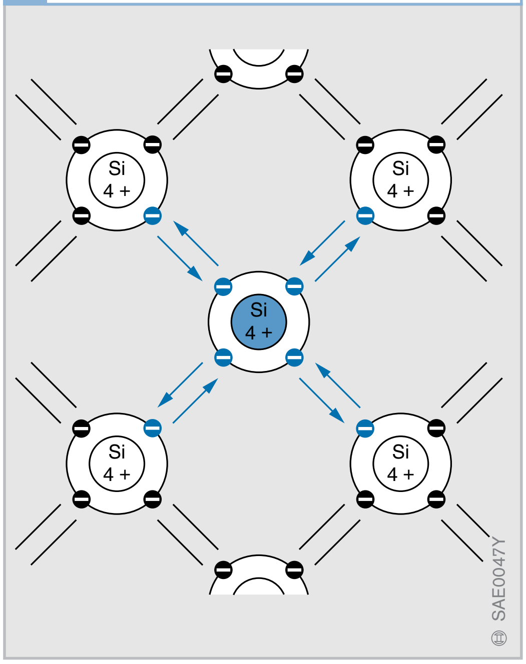
n-type doping

The addition of foreign atoms with five outer electrons (e.g. phosphorus) introduces free electrons because only four are required to bind each atom within the silicon crystal lattice. Each phosphorus atom introduced therefore supplies one free, negatively charged electron. The silicon becomes negatively conductive (Fig. 2). It is then referred to as n-type silicon.

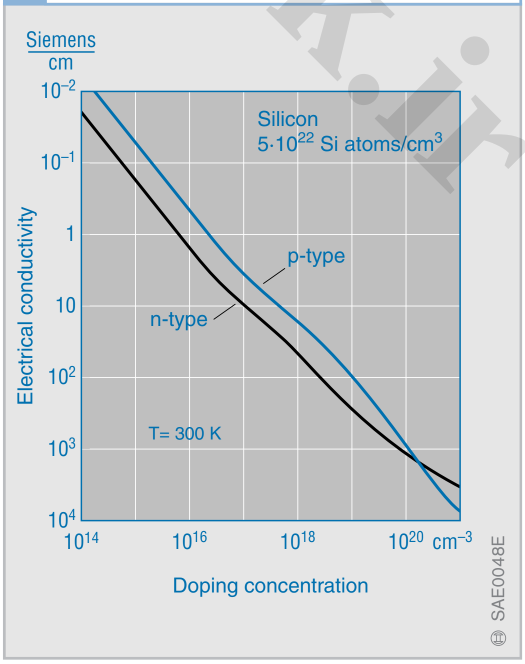
p-type doping

The addition of foreign atoms with three outer electrons (e.g. boron) creates electron gaps. The boron atom is one electron short of the number required to properly bind it within the silicon crystal lattice. The result-

1 Structure of a perfect, pure-silicon crystal lattice with four outer electrons (valance electrons ●)



2 Electrical conductivity of silicon relative to concentration of added doping substance



ing gaps are referred to as "holes". In silicon, those holes are mobile, and within an electrical field they move in the opposite direction to electrons. Holes therefore act like free positive charge carriers. Each boron atom introduced therefore supplies one free, positively charged hole. The silicon becomes positively conductive (Fig. 2) and is therefore referred to as p-type silicon.

The conductivity of n-type silicon is around 10 times as high as that of p-type silicon assuming other factors are equal (e.g. number of foreign atoms, temperature, pressure) because electrons can move about 10 times as fast as holes.

Intrinsic conductivity

The application of heat or light can generate free charge carriers even in undoped silicon. They consist of electron-hole pairs and make the semiconductor intrinsically conductive. Such conductivity is generally low compared with that produced by doping. The number of electron-hole pairs increases exponentially with rising temperature and ultimately erases the electrical differences between areas of p-type and n-type silicon created by doping. Consequently there are maximum limits for the operating temperatures of semiconductor components, as shown in the following table:

Material	Max. operating temperature
Germanium	90 ... 100 °C
Silicon	150 ... 200 °C
Gallium arsenide	300 ... 350 °C

An n-type semiconductor always has some holes and a p-type semiconductor always has some free electrons. Such minority charge carriers are fundamental to the way in which almost all semiconductor components work (refer to section entitled "Electronic Components").

p-n junction

The boundary between the p-type and n-type regions of the same semiconductor crystal is called the p-n junction. Its properties are fundamental to almost all semiconductor components.

p-n junction without external voltage

In the p-type region there are a large number of holes (○) and few free electrons. In the n-type region, by contrast, there are extremely few holes and a large number of free electrons (●) (Fig. 3). Due to the concentration differentials, the mobile charge carriers in each region diffuse into the other region in each case (diffusion currents). As a result, the p-type region is negatively charged and the

3 p-n junction without external voltage

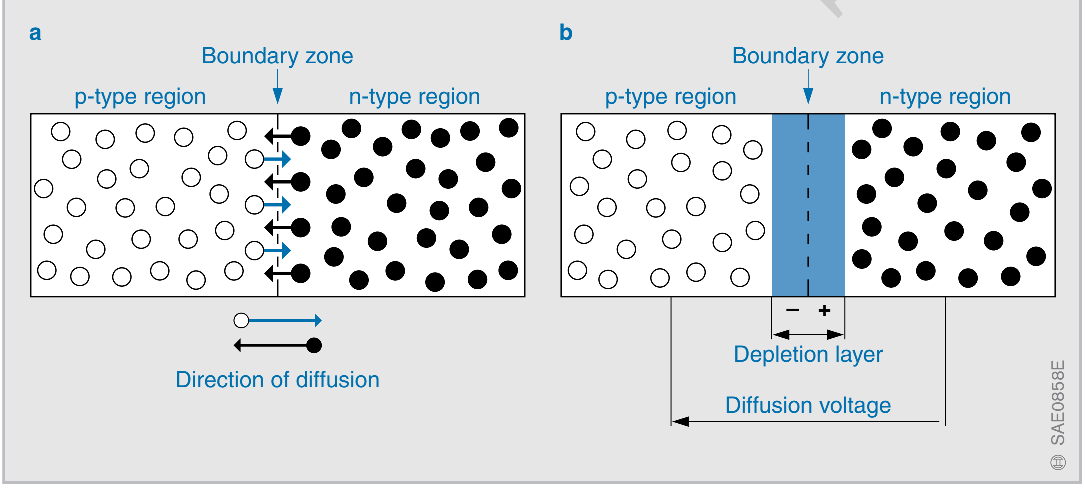


Fig. 3
a Diffusion of mobile charge carriers
b Formation of depletion layer (space-charge region)
○ Holes
● Electrons

n-type region positively charged. A potential difference (diffusion voltage) is thus created between the p-type region and the n-type region which counteracts migration of the charge carriers. This brings the equalisation of holes and electrons to a halt. As a result, a region that is deficient in mobile charge carriers, and therefore has poor electrical conductivity, is created at the p-n junction. It is referred to as the depletion layer or space-charge region. Due to the diffusion voltage, the depletion layer has as strong electric field.

p-n junction with external voltage

If an external voltage is applied to a p-n junction, it produces the effects explained below (Fig. 4).

Reverse bias

If the negative terminal is connected to the p-type region and the positive terminal to the n-type region, the width of the space-charge region increases. Consequently, current flow is largely inhibited apart from a very small residual current (reverse current) produced by minority charge carriers.

Forward bias

If the positive terminal is connected to the p-type region and the negative terminal to the n-type region, the depletion layer is broken down. When the diffusion voltage is exceeded, the charge carriers "flood" the p-n junction and a large current flows in forward direction.

Breakdown voltage

The breakdown voltage is the reverse-direction voltage above which a small increase in voltage brings about a steep rise in the reverse current.

The cause of this effect is the release of bound electrons from the crystal lattice in the space-charge region due to the high field strength (Zener breakdown) or due to surges of accelerated electrons. The accelerated electrons strike other electrons, breaking them free of their bonds and starting an avalanche-like increase in the number of charge carriers ("avalanche breakdown" or "first breakdown"). Both effects are reversible. The nature of the cause and the level of the breakdown voltage are dependent on the doping concentration profile.

A second breakdown occurs if there is localised heating of a semiconductor component caused by current constriction so that the area concerned becomes more conductive. This results in a self-accelerating increase in current and leads to the destruction of the semiconductor component.

4 p-n junction with external voltage

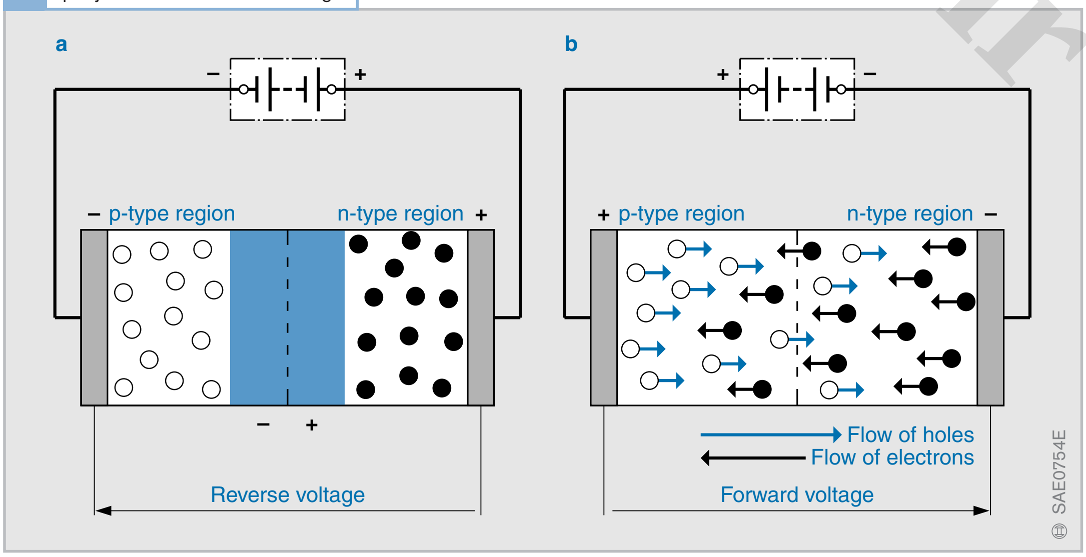


Fig. 4
a Reverse bias
b Forward bias

Electronic components

Electronic devices are made up a large number of components that can be subdivided into two main categories: passive components and semiconductor components (active components); the latter can be further subdivided into subcategories.

Passive components

Ohmic resistors, capacitors and inductances are classed as passive components.

Ohmic resistors

Ohmic resistors are generally made of materials with conductive properties similar to metals e. g. carbon (carbon-film resistors) or special metal alloys (metal-film resistors). They are constructed so as to reduce to the absolute minimum the effect of voltage, current and temperature on their electrical resistance. The conductor has a very small cross-section relative to its length, this being achieved either by the application of thin films to insulators or by winding wire into a coil.

In electronic circuits, resistors limit the current flow *I* or produce a voltage *U* proportional to the current. The resulting energy loss is converted into heat. An extreme example of this is a light bulb, in which an ultra-fine tungsten filament is heated to such a degree that it starts to glow.

The resistance in a circuit is referred to by the symbol *R* and its unit of measurement is the ohm (Ω).

Capacitors

The most simple type of capacitor consists of two parallel metal plates that are not in contact with one another. The area and separation of the plates as well as the medium separating them (dielectric) determine the quantity of charge carriers that can be stored by the capacitor (its capacitance). Using a vacuum (or air) as the dielectric offers the lowest capacitance. The capacitance can be substantially increased by the use of other insulating materials. The "amplification factor" is referred to as the relative permittivity or dielectric constant ε_r. Since, in many cases, very large plates would be required for the frequently used capacitance levels, capacitors are normally made by winding long strips into a coil (wound capacitors) or packing a large numbers of small plates together (multi-layer capacitors). The electrolytic capacitor uses a thin layer of oxide as the dielectric. This method makes it possible to manufacture small capacitors with large capacitances.

When direct current is applied to a capacitor, it stores up charge until the limit of its capacitance is reached; at that point current can no longer flow. Thus – for a certain period at least – the capacitor stores electrical energy that is then available if the external power supply fails. This effect is utilised in the circuit for triggering the airbag, for instance. Even if the wires connecting the airbag triggering unit to the vehicle's battery are severed in a serious accident, the capacitor still holds sufficient electrical energy to ensure that the airbag is deployed.

In alternating-current circuits, a capacitor has a similar effect to a resistor. Its resistance is dependent on the capacitance and the frequency of the alternating current. As the frequency decreases, the resistance increases. If the frequency is reduced to its lowest limit, i.e. zero (direct current), the resistance is equal to infinity and, consequently, no current flow is possible. This relationship is utilised, for example, by frequency filters in speaker systems in order to filter out the high-frequency sounds from the lower-frequency ones (high-pass).

The capacitance in a circuit is referred to by the symbol *C* and its unit of measurement is the farad (F).

Inductances

When an electric current flows through a coil, a magnetic field is created. The strength of the magnetic field depends on the strength of the current, the number of windings in the coil and the properties of the coil core (generally a ferrite or iron core). Inductance is the characteristic of a coil which indicates the amount of magnetic energy the coil can store for a given current.

If the current or the magnetic flux changes, a voltage is induced in the coil which counteracts the generation of the magnetic field. Once the magnetic field of a coil carrying direct current has reached its maximum strength, current flow is no longer restricted. Thus inductance does not represent a lasting hindrance to the flow of direct current.

In an alternating-current circuit, because of the constant generation and collapse of the magnetic field (and the energy contained within it), an inductance acts as a frequency-dependent resistor, the characteristics of which, however, are exactly the opposite of those of a capacitor. In this case, the higher the frequency, the greater is the resistance. This means that in frequency filters such as those referred to above in the description of capacitors, inductors can be used to filter out the low-frequency sounds from the high-frequency ones (low-pass).

The inductance in a circuit is referred to by the symbol *L* and its unit of measurement is the henry (H).

Semiconductor components

Semiconductor components are active components that are generally subdivided into four categories (Table 1). These categories are further subdivided according to the way in which the components are manufactured and their function. The first two "classical" categories are discrete semiconductor components and monolithic integrated circuits (see Table 1 for examples).

Discrete semiconductor components are self-contained, individually distinguishable components (discrete being derived from "discretus", the past participle of the Latin verb "discernere" meaning "to distinguish"). According to the currently accepted definition, semiconductor components with fewer than 100 transistor functions are referred to as discrete.

Monolithic integrated circuits
ICs (Integrated Circuits) are active components which incorporate more than 100 individual functions on a single chip (mono-

1 Semiconductor components (Examples)
Discrete active components
– Diodes
– Semiconductor resistors
– Transistors
– Thyristors
Integrated circuits (ICs)
– Analog circuits
– Digital circuits
– Mixed-signal circuits
Opto-electronic components
– Photoresistor
– Photodiode
– Photovoltaic cell
– Laser diode
– Phototransistor
– Charge-coupled device
Micromechanical sensors (Examples)
– Pressure sensors
– Acceleration sensors
– Yaw rate/angle sensors
– Flow sensors
– Temperature sensors
– Position/angle sensors (Hall-effect sensors)
– Gas sensors

Table 1

lithic literally meaning "made from a single stone" from the Greek "monolithos" meaning "single stone").

Optoelectronic components form the third category. They are so distinctly different from the classical active components in terms of their method of production and usage, that classification as a separate category makes sense.

Micromechanical sensors using MST (Micro-System Technology) or MEMS (Micro ElectroMechanical Systems) have more recently come into being as a fourth category. In this case too, the methods of production and the type of use differ substantially from conventional active components.

Diodes

Diodes are semiconductor components with a p-n junction and two connections – one to the p-type region and one to the n-type region ("diode" means "two ways"). A diode utilises the characteristics of the p-n junction. The pattern of doping impurity concentration within the crystal determines the specific characteristics of diodes.

Diodes designed for a forward current of more than 1 A are referred to as power diodes.

1 Diode characteristics

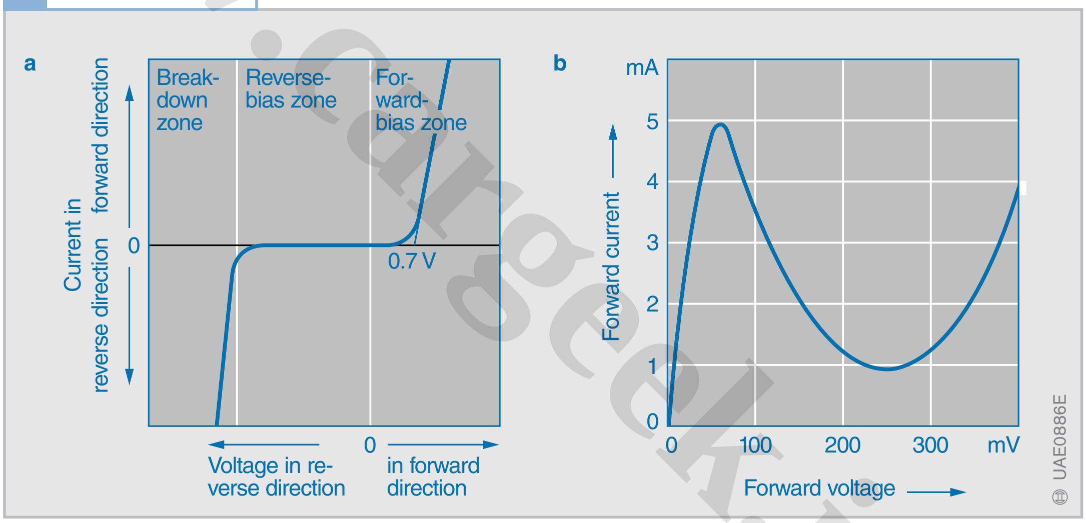


Fig. 1
a Characteristic of diodes, e.g. rectifier diode, Zener diode and Schottky diode
b Section of tunnel-diode characteristic

2 Rectifier circuits using diodes

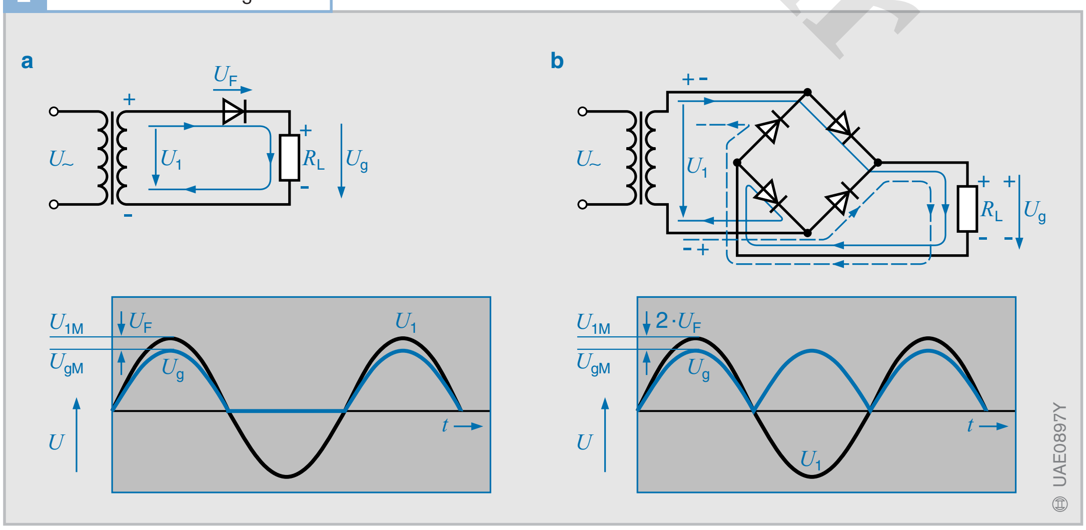


Fig. 2
a Half-wave rectifier
b Bridge rectifier
U_F Diffusion voltage (approx. 0.7 V)
$U_\sim$ AC voltage
U_1 Transformed AC voltage
U_g Rectified voltage
R_L Load resistance
U_{1M} Amplitude of U_1
U_{gM} Amplitude of U_g
t Time

Rectifier diode

A rectifier diode allows current to pass in one direction (positive terminal connected to p-type region) but not in the other (positive terminal connected to n-type region). It acts like a flow control valve and is therefore the ideal component for rectifying alternating current (Figs. 1a and 2).

The current flowing in reverse direction (reverse current) is much smaller than the forward current (approx. 1/100 the strength). It increases rapidly with rising temperature.

Amongst other applications, rectifier diodes are used in automotive alternators to convert the alternating current into direct current. Because of the high ambient temperatures to which the alternator is subjected, the reverse current is a critical factor and has to be taken into account in the design of the diodes used.

Rectifiers for high reverse voltages

In order to obtain a high reverse voltage in rectifiers, at least one region of the rectifier must have low conductivity. However, that means a high resistance in forward direction and consequently high power loss and excessive heating.

Creating a very lightly doped i-type region between the heavily doped p and n-type regions produces a p-i-n rectifier which, despite having a high reverse voltage, has a low forward resistance (conductivity modulation). The i-type region acts like intrinsically conductive silicon.

Uses: all rectification applications involving high voltages.

Zener diode

The Zener diode is a semiconductor diode in which the reverse current rises abruptly upwards of a certain voltage as a result of Zener breakdown and/or avalanche effects. Even though large numbers of these diodes depend for their function on avalanche effects rather than Zener breakdown, they are still referred to as Zener diodes.

Zener diodes are rated for different breakdown voltages depending on their particular application. They are designed for continuous operation at the breakdown voltage. They are used chiefly for voltage limitation and for generating a voltage reference, e.g. in direct-current power supply units.

Variable-capacitance diode

The space-charge region at the p-n junction acts like a capacitor; the semiconductor material depleted of charge carriers acts as the dielectric. An increase in the applied voltage increases the width of the depletion layer and reduces the capacitance; reducing the voltage increases the capacitance. Capacitance diodes are used mainly for resonant circuit tuning and frequency multiplication (e.g. in tuners).

Schottky diodes

The Schottky diode is a semiconductor diode with a metal-semiconductor junction. Because electrons transfer more easily from the n-type silicon to the metal film than in the opposite direction, an electron-depleted boundary layer known as the Schottky barrier is created in the semiconductor. Transmission of charge is performed exclusively by electrons; as a result, extremely high switching speeds are achieved because no minority storage effects occur.

Schottky diodes are suitable for use as fast switches and microwave rectifiers.

Tunnel diode

The tunnel diode (or Esaki diode) is a semiconductor diode with a very heavily doped p-n junction in which the tunnel effect, which is explainable only by quantum mechanics, occurs to such a degree that when operated in forward direction a negative differential conductivity occurs over a certain range of the current/voltage characteristic (Fig. 1b).

Tunnel diodes are used as oscillators in the gigahertz (GHz) range and for low-noise amplifiers.

Semiconductor resistors

In contrast with ohmic resistors, these components are voltage, current and temperature-dependent. They generally consist of polycrystalline semiconductor materials. The effects that occur in semiconductor resistors are based in part on depletion-layer properties that appear at the crystallite boundaries.

Varistors

The resistance of a varistor decreases as voltage increases. The polarity of the voltage makes no difference. A varistor consists of polycrystalline ZnO or SiC powder mixed with a binder, compressed and sintered.

Uses: e. g. voltage stabilisers, surge protectors for semiconductor circuits.

NTC resistors (thermistors)

NTC (**N**egative **T**emperature **C**oefficient) resistors (thermistors or thermal resistors) are, as the name suggests, resistors that have a marked negative temperature coefficient. Their electrical resistance decreases as the temperature increases and consequently, NTC resistors conduct electricity better at high temperatures than at low temperatures.

NTC resistors are made from polycrystalline metal oxides such as Fe_2O_3, $ZnTiO_4$ or $MgCr_2O_4$ by a process of compression and sintering. Their temperature coefficients can be as much as –6 %/K.

Uses: e. g. temperature sensors.

PTC resistors (thermistors)

PTC (**P**ositive **T**emperature **C**oefficient) resistors are resistors that have a positive temperature coefficient (of the order of a few ‰/K). Their electrical resistance increases as the temperature increases and consequently, PTC resistors conduct electricity better at low temperatures than at high temperatures. Most metals are thermal resistors with very low temperature coefficients.

The term PTC resistors refers to resistors made of semiconductor materials. PTC resistors made of ferro-electrical ceramic material (e. g. polycrystalline barium titanate) have a relatively narrow temperature range and a very high positive temperature coefficient (+6 to +60 ‰/K). PTC resistors made of silicon have wide temperature range and a virtually constant positive temperature coefficient (approx. +0.8 ‰/K).

Uses: e. g. liquid level sensors, heating regulators.

Magnetoresistors

A magnetoresistor is a magnetically controllable semiconductor resistor. Its resistance increases as the magnetic flux density *B* increases (Fig. 3).

The indium antimonide film (approx. 25 µm thick) contains minute needles of nickel antimonide with very high electrical conductivity. The flux density prevents the charge carriers from following a direct path. They move at an angle from one needle to the next. Within the metallic needles, differing charge-carrier densities are immediately equalised. As the flux density *B* increases, the current paths become more and more angled and the distance travelled by the charge carriers therefore greater. As a result, the resistance of the magnetoresistor increases.

Uses: e. g. magnetic-field sensors, controllable resistors.

3 Magnetoresistor

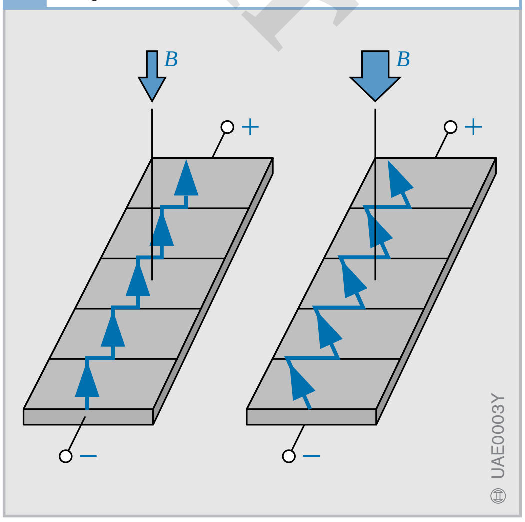


Fig. 3
Current path under the influence of a low (left) and a high (right) magnetic flux density *B*

Transistors

Transistors can be used control a large current with a small control current. Consequently, these semiconductor components can be used as power amplifiers or switches.

Transistors have three or more connections. Figure 4 illustrates the classification of the transistor family. "Transistor" is a contraction of the term "**transfer** resis**tor**".

Bipolar transistors

Bipolar transistors consist of three semiconductor regions of differing conductivity arranged to form a p-n-p or an n-p-n configuration. These regions (and their connections) are called the emitter (E), the base (B) and the collector (C).

Depending on their applications, transistors are categorised as low-signal transistors (up to 1 watt power loss), power transistors, switching transistors, low-frequency transistors, high-frequency transistors, microwave transistors, phototransistors, etc.

Such transistors are called bipolar because they make use of charge carriers of both polarities (holes and electrons). In an n-p-n transistor, positive charge carriers (holes) in the base current control roughly 100 times as many negative charge carriers (electrons) flowing between the emitter and the collector. This corresponds to a current amplification factor of around 100. Bipolar transistors are thus controlled by means of the base current.

Method of operation of a bipolar transistor
(with reference to an n-p-n transistor):
The emitter-base junction (EB) is polarised for forward bias (Fig. 5 overleaf). This means that electrons migrate into the base region. The base-collector junction (BC) is polarised for reverse bias. This creates a space-charge region with a powerful electrical field.

Discernible coupling (transistor effect) occurs if the two p-n junctions are very close to one another (less than 10 µm apart in silicon). In that case the electrons crossing the EB junction pass through the base to the collector. As soon as they come within the

4 Transistor family with symbols

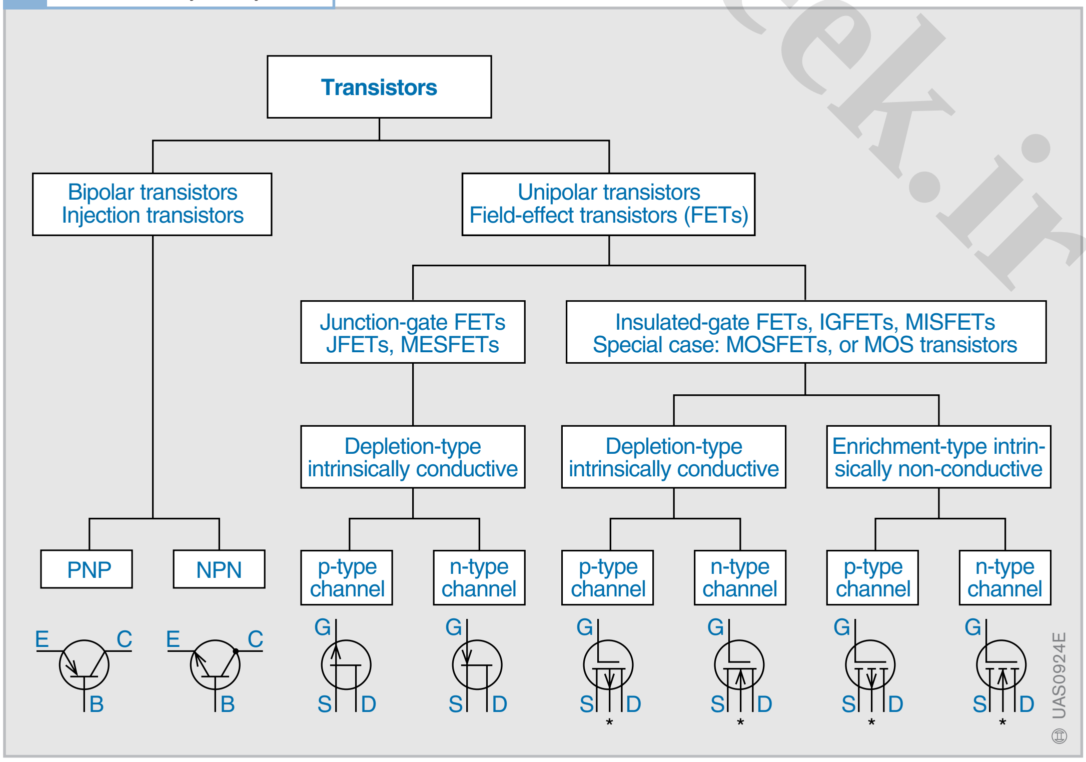


Fig. 4
B Base
E Emitter
C Collector
S Source
D Drain
G Gate
* Bulk connection (connection designation not usual)

range of the electrical field at the BC junction, they are accelerated into the collector region and travel onwards as the collector current. The concentration differential within the base thus remains and, therefore, so does the impetus for continued electron migration from the emitter to the collector (Fig. 5).

In conventional transistors, 99 % or more of electrons originating from the emitter pass into the BC space-charge region and form the collector current. The few that are lost have found their way into the electron gaps in the p-type base while passing through it. If this effect were left unchecked, the base would become negatively charged and the resulting repulsion forces would completely prevent the continued flow of electrons within an extremely short space of time (50 ns). Such a build-up of negative charge can be partially or entirely counteracted in a transistor by a low base current of positive charge carriers (holes). Small changes in the base current bring about large variations in the emitter-collector current.

With moderate base currents, the n-p-n transistor acts as a bipolar, current-controlled, amplifying semiconductor component. If the base current alternates abruptly between very low and very high levels, the transistor acts as a switch.

5 Bipolar n-p-n transistor

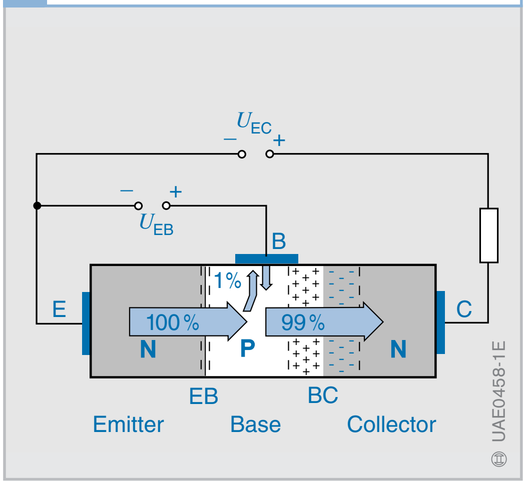


Fig. 5
N n-type silicon
P p-type silicon
E Emitter
B Base
C Collector

Field-effect transistors (FETs)

In a field-effect transistor (Fig. 6a), the current flowing from the source (S) to the drain (D) through a conductive channel is controlled by an electric field. That field is generated by a voltage applied via a control electrode known as the gate (G). The use of an electric field to control the current flow explains the origin of the term "field-effect transistor" (FET). In contrast with bipolar transistors, field-effect transistors use only one type of charge carrier (either electrons or holes) and are therefore also referred to as unipolar transistors. Field-effect transistors are subdivided into

- junction-gate field-effect transistors (junction **FET**s, JFETs)
- insulated-gate field-effect transistors, in particular **m**etal-**o**xide **s**emiconductor (MOS) field-effect transistors (MOSFETs).

MOSFETs are particularly suited to use in highly integrated circuits. Power FETs have superseded bipolar transistors for most applications.

Method of operation of a junction-gate FET
(with reference to a JFET with an n-type channel, Fig. 6a):
DC voltage is applied to the ends of an n-type crystal. Electrons flow from the source, S, to the drain, D. Two p-type regions at the sides forming the gate (G), and the negative voltage applied to them, determine the width of the channel. If the negative gate voltage is increased, the space-charge regions extend further into the channel and constrict the current path. The voltage at the control electrode, G, thus controls the current between the source, S, and the drain, D. The FET only requires charge carriers of a single polarity in order to function. The current is controlled virtually without power consumption. The junction FET is thus a unipolar, voltage-controlled component.

Method of operation of a MOS FET
(with reference to a MOS FET with p-type enrichment): while no voltage is present at

the gate electrode, no current flows between the source, S, and the drain, D; the p-n junctions are reverse biased (Fig. 6 b). A negative voltage applied to the gate, G, forces the electrons in the n-type region below the electrode into the centre of the crystal and draws holes – which are always present as minority charge carriers even in n-type silicon – to the surface. A narrow p-type layer is created below the surface, in effect a p-type channel. Current, which is comprised only of holes, can now flow between the two p-type regions (source and drain).

As the gate voltage acts through an insulating oxide layer, no current flows in the control circuit and control therefore requires no power. The MOS FET is thus a unipolar, voltage-controlled component.

PMOS, NMOS and CMOS transistors

In addition to the **p**-type channel **MOS** transistor, or PMOS transistor, there is the NMOS transistor in which the p and n-type regions are reversed in comparison with the PMOS transistor. Due to the higher mobility of electrons, NMOS transistors are faster than PMOS transistors which, for physical reasons, are easier to manufacture and were therefore available first.

If PMOS and NMOS transistors are created in pairs on the same silicon chip, they are referred to as CMOS transistors (complementary **MOS** transistors, Fig. 6 c). Their particular advantages are very low power loss, high interference immunity, TTL-compatibility (transistor-transistor logic in electrical circuits), low power-supply voltage and suitability for analog signal processing.

The low power consumption, particularly for digital circuits, is due to the fact that current only has to flow when a digital cell changes (e.g. switches from “0” to “1”). Retention of the information requires only that a voltage is applied. Since, with this configuration, the PMOS and NMOS transistors are connected to one another, the leakage current is minimal. More than 80 % of ICs are now manufactured using CMOS technology.

6 Transistor designs (Schematic)

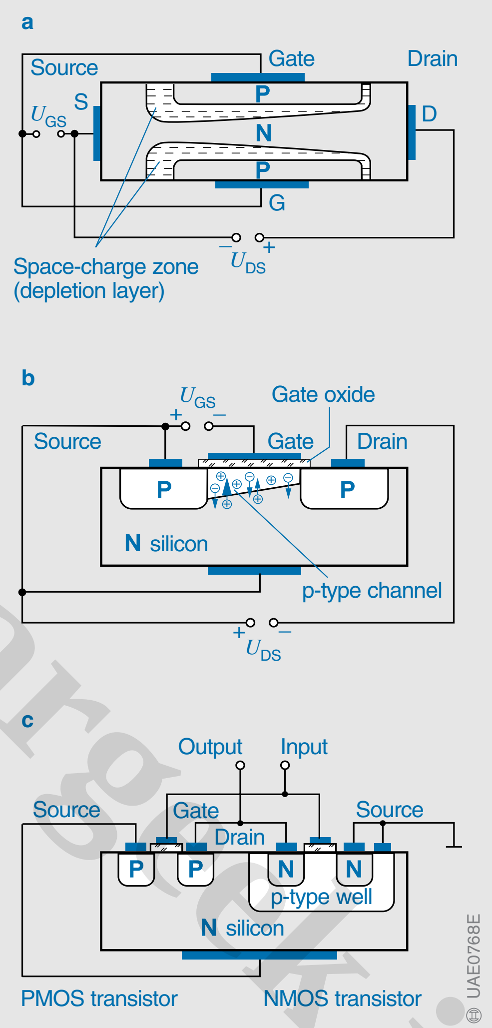


Fig. 6
- a Depletion layer, field-effect transistor
- b PMOS transistor
- c CMOS transistor pair
- N n-type silicon
- P p-type silicon
- S Source
- G Gate
- D Drain
- U_{GS} Voltage between G and S
- U_{DS} Voltage between D and S

Thyristors

Thyristors are semiconductor components with at least three p-n junctions (one of which may also be replaced by a suitable metal-semiconductor contact) that can be switched from a reverse-bias condition to a forward-bias condition (or vice versa). The term thyristor is used as the generic term for all types of component which conform to that definition. It is a contraction of the two words **thyr**atron (gas-filled tube triode) and res**istor**.

Uses in power electronics: speed and frequency control, rectification and conversion, switching.

7 Development of monolithic integration on a single chip in the period from 1970 to 2000 (logarithmic representation in a and b).

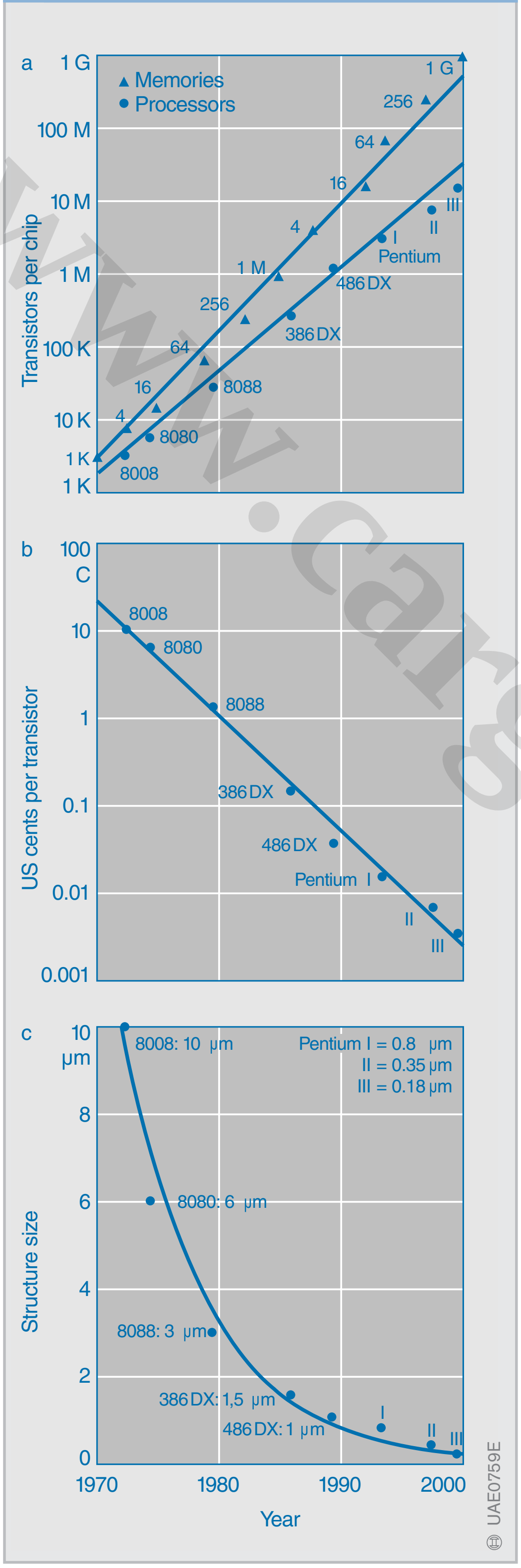


Fig. 7

a Increase in number of transistors per chip in memories and processors (Gordon Moore's Law)

b Fall in cost of transistors

c Decreasing size of structural units of new processors

Monolithic integrated circuits (ICs)

Monolithic integration

The planar process is a method of manufacturing semiconductors which makes it possible to produce all components of a circuit (resistors, capacitors, diodes, transistors) and the conductive connections between them on a single silicon chip in a single production process. It involves the creation of multiple p and n-type layers in a multi-stage doping process using a pre-doped single-crystal silicon disc or wafer.

An integrated circuit IC (Integrated Circuit, see section entitled "Manufacture of Semiconductor Components and Circuits") does not contain any "separate" (discrete) components but rather switching elements or functional elements.

Degree of integration

The degree of integration is defined as the number of functional elements, transistors or gates on a single chip. The following categories are defined based on the degree of integration (and chip surface area):

- SSI (small-scale integration): up to around 1000 elements per chip, average chip area 3 mm^2 (varies considerably depending on level of power loss)
- MSI (medium-scale integration): up to around 10,000 elements per chip, average chip area 8 mm^2
- LSI (large-scale integration): up to 100,000 elements per chip, average chip area 20 mm^2
- VLSI (very-large-scale integration): more than 100,000 elements per chip, average chip area 30 mm^2

The number of functional elements on VLSI chips is constantly increasing. Microprocessors can now have up to 10 million transistor functions per chip. Memory modules (DRAM) can even contain several hundred million transistors. The number of such functional elements is subject to an exponential increase over time. This fact was first recognised by Gordon Moore

(co-founder of the company Intel) and is expressed by "Gordon Moore's Law" (Fig. 7a):
Elements per chip = $2^{(\text{year} - 1956) \cdot 2/3}$

This law states that the number of functions per chip will double every 18 months. Since the structural units in which the chips are manufactured are becoming smaller at the same time as the number of functional elements is growing, the size of the ships alters very little and the costs per transistor are similarly dropping exponentially (Fig. 7b). Thus the structural units of new processors have shrunk over a period of 28 years from 10 µm in 1972 to 0.18 µm in the year 2000 (Fig. 7c).

The degrees of integration LSI and VLSI demand methods such as CAD (**c**omputer-**a**ided **d**esign). VLSI circuits in particular can only be created with the help of highly sophisticated programs which can convert entire function blocks into corresponding circuit subdivisions. Such program systems are called HDL (**H**ardware **D**escription **L**anguage) or VHDL (**V**isual **HDL**).

IC classification

There are various systems for classifying ICs, although there are invariably mixed categories within each system.

- *Classification by method of production:* Bipolar/Unipolar (MOS); mixed category: e. g. BiCMOS, BCD
- *Classification by function:* Analog/Digital; mixed category: e. g. mixed-signal IC
- *Classification by application:* Standard IC/ASIC mixed category: e. g. ASSP

ASIC (**a**pplication-**s**pecific **IC**): developed and produced specifically for a particular application and exclusively for a particular customer
ASSP (**a**pplication-**s**pecific **s**tandard **p**roduct): developed and produced specifically for a particular application for a particular customer but is also sold to others for the same type of application.

Apart from a few exceptions, ICs are now always based on MOS or combined technologies, and for that reason bipolar ICs are not described in any further detail at this point.

Analog circuits

Analog circuits are required wherever electronic equipment has to communicate with the outside world – which is always analog. This applies in particular to the areas of "preparation" and "processing" of input and output signals. Examples of analog-input signal generators are microphones (telephone) and sensors. Examples of analog-output signal receivers are speakers or actuators.
Figure 8 illustrates the stages of development of an analog IC from conception to installation in an ECU.

8 Development of an analog IC from conception to installation

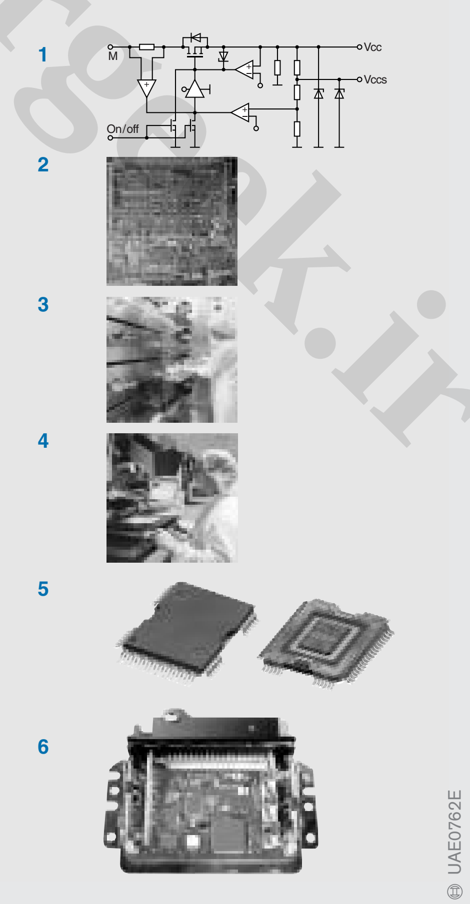


Fig. 8
1 Development and simulation
2 Layout
3 Production
4 Testing, analysis and release
5 Packing, final testing
6 Installation in an ECU

Basic structures: stabilised-voltage sources, stabilised-current sources, differential amplifier stages, coupling circuits, potential shifting circuits, output stages
Application-based ICs: operational amplifiers, voltage regulators, comparators, timers, transducers, interface circuits
Specialised ICs: voltage references, broadband amplifiers, analog multipliers, function generators, phase-locking circuits, analog filters and switches

Digital circuits

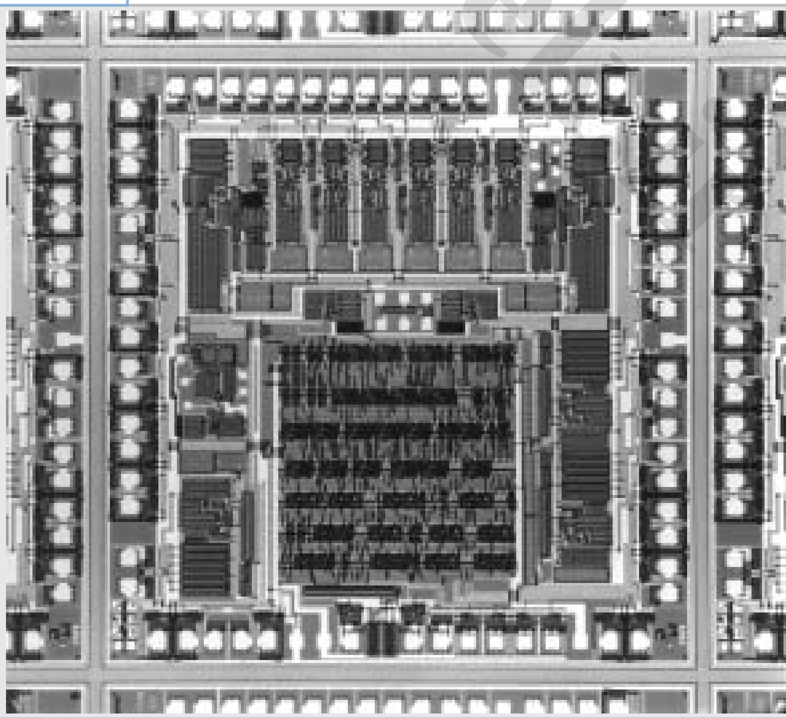
Digital circuits (Fig. 9) are used wherever large volumes of data have to be processed within a short space of time. High-performance microcontrollers in particular can perform several hundred million computational operations per second. They are thus able to utilise the input data to provide the required, highly accurate response at the output of the particular device. Similarly, large volumes of data can be transmitted via a cable link in coded form, e.g. several thousand telephone calls can be transmitted simultaneously via a fibre-optic cable without cross-talk.

Since inputs and outputs for communication with the real world always have to be analog, analog-digital converters (ADCs) are used at the inputs of a digital circuit, and digital-analog converters (DACs) at the outputs.

Digital systems make use of a whole series of recurring basic circuits and variations of them. The range extends from simple gates to memories, microprocessors and microcontrollers.

Digital modules can only be connected up to form a complete system if power supply voltages, logic signal levels, switching speeds and signal transmission times are compatible. This requirement is met within a family of circuits. Since, apart from a few specialised applications mainly in military systems, CMOS logic circuits are the only type now used, such compatibility is no longer a problem.

9 Digital circuit on a chip



UAE0013Y

Mixed-signal ICs

ICs in which both analog and digital circuit elements are combined are called mixed-signal ICs. They are an essential requirement for the combination of entire electronic systems on a single piece of silicon (SoC: System-on-a-Chip).

With ever-increasing degrees of integration, this type of circuit is becoming more and more attractive and the number of applications is continually increasing. It is now conceivable, for example, to integrate the entire electronic circuitry for a mobile phone or a complete Internet-access system on a single chip. Where only low levels of electrical power are required, such as in the examples quoted, both the analog and the digital sections are manufactured using CMOS technology.

Mixed-signal ICs for applications involving higher electrical power levels, such as are frequently encountered in automotive systems, are typically hybrids of various technologies, such as BCD systems (where B stands for bipolar, C for CMOS and D for DMOS).

Figure 10 shows how the functions, "sensing", "analysing" and "acting" are integrated in a mixed-signal IC. In such circuits, the bipolar section is used for converting the analog inputs. The CMOS section of the IC performs the logical processing operations using digital technology. The DMOS section enables high analog output performance.

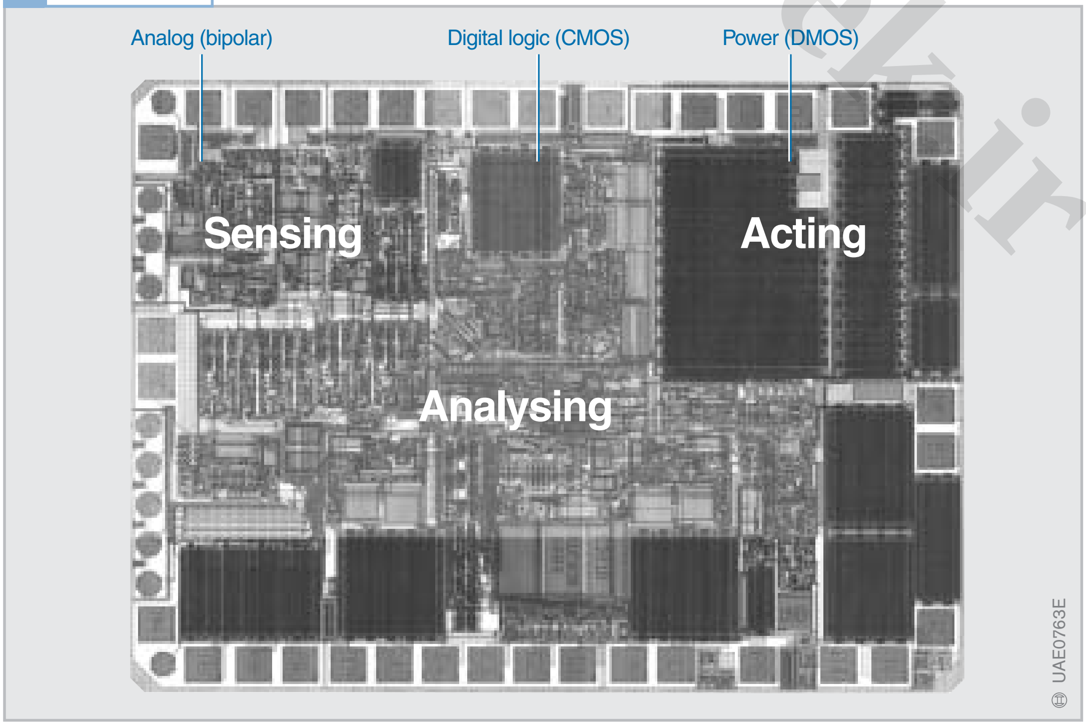
Elements of digital circuits

Digital circuits assign voltage signals one of the two signal levels "0" (low voltage level) or "1" (high voltage level) and process those signals digitally. As digital circuits make up the greater part of all ICs and our day-to-day lives are no longer conceivable without them, a few typical circuit elements are described below.

Gates

Gates are logical-operation circuits with two or more inputs. The input signals are logically linked to one another so that their various combinations determine the output signals according to a defined logic (e.g. AND gate).

10 Mixed-signal IC



Inverters
Inverters are digital circuits that convert the digital input signal “0” into the output signal “1” and vice versa.

Bus
The bus links up the individual elements of a digital circuit; this relates both to connections within an integrated circuit and the interconnection of separate digital ICs.

A bus can connect up a large number of separate modules with different functions but with electrically identical interfaces. A serial bus (e.g. for connecting up serial EEPROMs) transmits the data along a single lead. A parallel bus (e.g. address bus, data bus, control bus) is a bundle of parallel lines. The number of separate items of information that can be simultaneously transmitted (equal to the number of bus lines) is, together with the data transmission speed, a measure of the performance capacity of the data bus. The most common types of bus are 8-bit, 16-bit and 32-bit.

The dimensioning of the data bus is determined by the capacity of the CPU (**c**entral **p**rocessing **u**nit). In order for a system to achieve its maximum performance capability (i.e. maximum processing speed) the CPU and data bus should have the same capacity, i.e. an 8-bit CPU should use an 8-bit bus, a 16-bit CPU a 16-bit bus, etc.

In addition, the bus should be able to transmit the data as quickly as the CPU can process it. This is often not the case with external busses (e.g. CPU clock speed f_{CC} = 400 MHz and bus clock speed f_{CB} = 133 MHz).

Only one bus node at a time can write data to the bus. According to the bus type, this addresses the data in such a way that it can only be read by the addressee, or else it switches all other bus nodes off before sending the information and then on again afterwards.

Clock-pulse generator (CPG)
A clock-pulse generator ensures that all operations on the microcomputer are synchronised with a defined timing pattern. The clock-pulse generator must be matched to the required speed of computing operations.

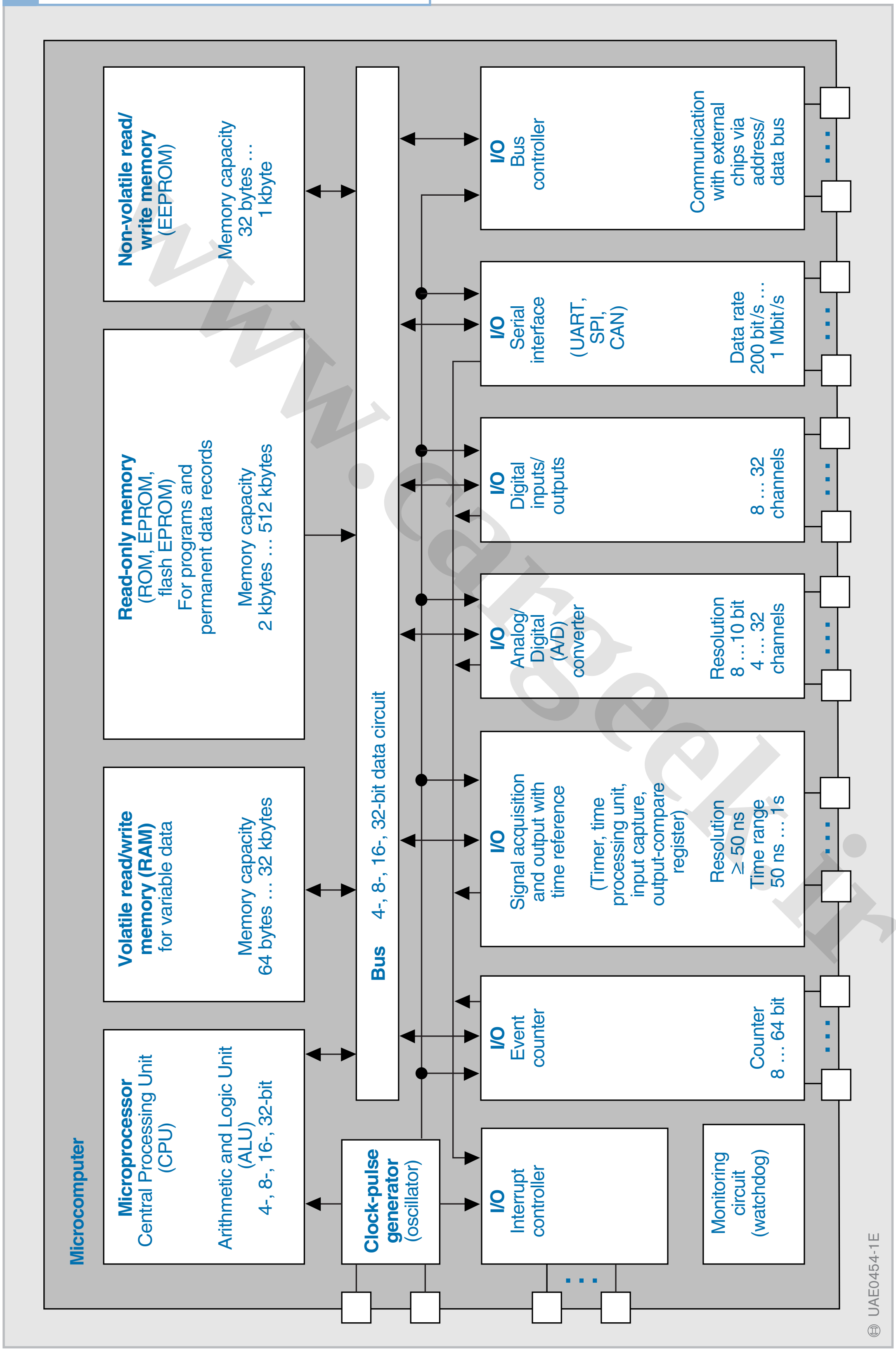
Input/output (I/O) unit
The I/O unit handles data exchange with the outside world. The input signals are scanned at the required frequency. The output signals are transmitted at the speed required by, and in the optimum order for, the application concerned or are held in temporary storage until called for.

Microcomputers

The microcomputer consists of the following interacting components (Fig. 11):

- A *microprocessor* as the CPU (**c**entral **p**rocessing **u**nit). For its part, the microprocessor consists of the control unit and the arithmetic and logic unit. The arithmetic and logic unit performs arithmetical and logical operations, as the name suggests, while the control unit carries out the instructions from the program memory.
- *Input and output units* (I/O units) which handle the exchange of data with peripheral devices. Peripheral devices include input and output devices and external data storage media.
- A *program memory* in which the operating program (user program) is permanently stored (ROM, PROM or EPROM).
- *Data memories* for holding the data being processed at any particular time. This data changes continually (RAM).
- The *bus system* which links up the individual components of the microcomputer.
- A *clock-pulse generator (oscillator)* which ensures that all operations on the microcomputer are synchronised with a defined timing pattern.
- *Logic circuits* which are modules with specialised tasks such as program interrupts. They are integrated in individual I/O units.

11 Microcomputers



The chief components of a microcomputer are generally separate modules connected to one another on a printed-circuit board. For simpler tasks such as are more and more frequently required for Internet access on wireless communication devices, single-chip computers are becoming increasingly common. They integrate the functions detailed above on a single silicon chip (System-on-a-Chip). The relatively small amount of RAM that can be accommodated on a chip at reasonable cost limits the performance capabilities of such highly integrated systems.

Microprocessor

A microprocessor is a central processing unit (CPU) in the form of an integrated circuit on a chip. The concept of the microprocessor avoids individualisation despite the high degree of integration, and enables adaptation to the multiplicity of practical demands by reliance on programming. A microprocessor is not capable of functioning on its own; it is always part of a microcomputer. At present, there are 16-bit, 32-bit and 64-bit microprocessors. Processors are subdivided into two main categories:

- PCs use CISC (**C**omplex **I**nstruction-**S**et **C**omputer) processors. These are extremely versatile and permit unrestricted programming.
- Workstations normally use RISC (**R**educed **I**nstruction-**S**et **C**omputer) processors. These perform specific tasks such as are frequently required on workstations much faster, but are distinctly slower for all other tasks.

Microcontrollers

The microcontroller is a component that incorporates the following elements on a chip:

- CPU,
- random-access memory (RAM),
- peripheral modules (input/output, interrupt, timer, serial interfaces) and
- an optional read-only memory (ROM).

With these integrated components, the microcontroller is capable of stand-alone operation. For that reason, it is also referred to as a single-chip microcomputer.

The microcontroller is used to control self-regulating systems such as an engine-management system.

At present, there is a choice of 4-bit, 8-bit, 16-bit or 32-bit microcontrollers for the various types of application. Depending on application, they may also have expansion modules connected to them (e.g. additional memory for data and program code).

The program that is run by the CPU is permanently fixed in the read-only memory and is not changed for different applications. This distinguishes the single-chip microcontroller from the PC.

Transputer

A transputer is a special type of microprocessor that is particularly suited to the construction of parallel computer networks. In addition to the usual components of a microprocessor, it has communication and process-handling hardware on a chip.

Programming

The only command form capable of direct interpretation by a microprocessor is a bit pattern, i.e. the binary representation of a number. Since, however, this form of instruction is not easy to work with for a programmer, and is therefore susceptible to errors, easily memorable abbreviations (mnemonics) are used. These are automatically translated by an assembler program into bit patterns (machine code) that can be understood by the microprocessor. Microcomputers for single-purpose applications are programmed in the assembler language specific to the processor.

For more complex systems and programs, “high-level programming languages” such as C are needed, as otherwise it would be impossible to keep extensive programs manageable and free of errors. Such languages require sophisticated translation programs (compilers) which convert the text of the high-level language into a form that can be processed by the processor.

Semiconductor memories

Applications

Memories are used to store large volumes of

- Digital signals representing data (I/O data, statuses, intermediate results involving frequent and rapid reading and writing),
- Program code (usually permanently stored), and
- Constants (permanently stored)

Storage involves

- recording (writing),
- permanent retention (actual storage), and
- location and retrieval (reading) of information.

Memories make use of physical effects that make two different statuses clearly distinguishable and easy to generate and identify. The advantage of semiconductor memories lies in their technological compatibility with the modules used in other parts of a computer, and thereby in new opportunities for functional integration.

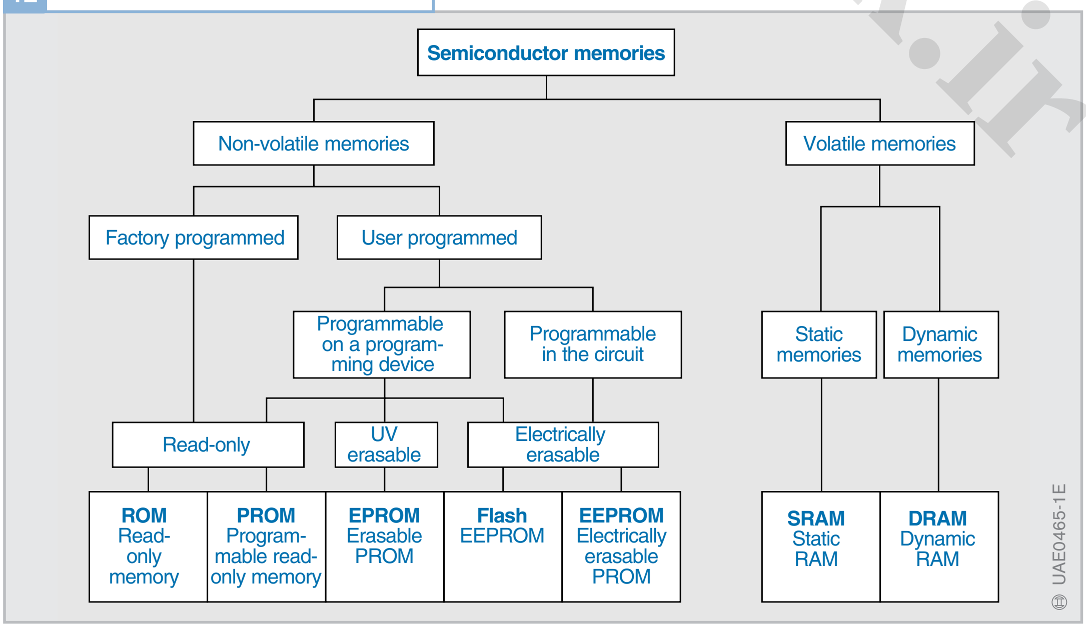
Classification of semiconductor memories

Digital signals are stored by utilising the status alternatives conducting/not conducting or charged/uncharged. The most important terms are explained below according to their standardised definitions, where applicable, or their most common usage (see Fig. 12 for overview).

Random-access memory (RAM)

Random-access memory or RAM is a short-term memory that allows direct access to any storage location. Information can be read/written from/to the memory any number of times. However, it must exist in binary form, i.e. encoded as a series of "yes or no" statuses (logical "1" or logical "0"). Such a "yes or no" unit of information is called a bit (**b**inary dig**it**). Random-access memories, like read-only memories, are organised on a bit or word basis depending on the particular application. A "word" is a group of bits that can be processed as a single unit. The word length is equal to the number of bits processed as a single unit. On microcomputers, word lengths of 4, 8, 16, 32 and 64 bits are common. 8 bits make up one byte, i.e. 1 byte = 8 bits.

12 Classification of semiconductor memories



Memories can be organised on the basis of a variety of word lengths. The way the memory is organised is generally indicated by a multiplication sign, e.g. an 8 M × 8 RAM or 8-megabyte RAM
≈ 8 million bytes
= 64 million bits
= 64 Mbit

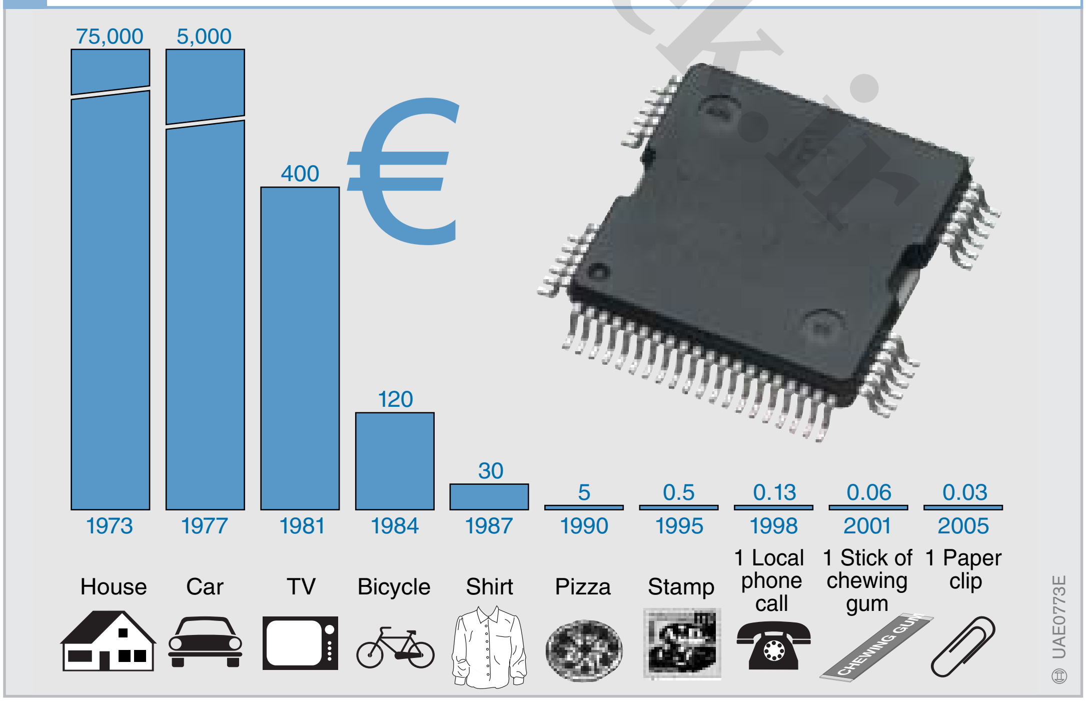
That means that the memory has eight data inputs and eight data outputs at which the eight bits of one of the eight million stored data words are simultaneously present. Its word length is thus 8 bits. Since all memory specifications are based on the binary system, the exact number of bits is as follows:
64 Mbit
= 8 Mbyte × 8 bit
= 2^{23} × 8 bit
= 2^{26} bit
= 67,108,864 bits

Static RAM (SRAM)
Static RAMs use bistable switching elements as the data storage cells. Their function is similar to that of a flip-flop, a simple circuit with two transistors, of which either the one (logical "1") or the other (logical "0") conducts at any one time. In SRAM, the information remains stored until the storage cell concerned is addressed and overwritten, or the power supply is switched off.

Static RAMs are currently available in sizes up to 16 Mbit. Although considerably more expensive than dynamic RAMs, because they can be written into and read significantly faster they are used as high-speed access computing memories (cache memory) for the CPUs in computers. Typical read-write times are currently less than 30 ns.

Dynamic RAM (DRAM)
The information in a DRAM is stored as an electrical charge in the gate capacitor of a CMOS transistor which either passes (logical "1") or does not pass (logical "0") a current as a result. As such capacitors are susceptible to leakage, the charge is gradually lost. In order that the information (charge) is retained, the memory has to be refreshed at regular intervals (every few ms).

13 Growth in productivity of dynamic RAM (DRAM) in hindsight.
Cost of 1 Mbit of memory capacity in Euro (€)



DRAMs are currently available commercially in sizes up to 256 Mbit (on one chip), and in laboratory conditions up to 1 Gbit (gigabit = 10^9 bit). Since every storage cell consists of a transistor and a capacitor, this requires more than a billion transistors on a single chip.

DRAMs are volatile memories. This means that they lose their information when power is switched off.

DRAMs are very widely used nowadays as system memories for all types of computer. The fall in price for 1 Mbit of memory over the last few decades has been so dramatic that ample memory can now be fitted in any PC without substantially affecting its price (Fig. 13).

Read-only memory

Read-only memory (ROM) is permanent-storage memory that allows any memory location to be accessed directly but – as the name indicates – allows the information only to be read and not altered.

A ROM is a non-volatile memory, i.e. the information it contains is retained even when the power supply is switched off. It is usually used to store program codes (control programs) and fixed data (function tables, encoding rules, character generators, engine characteristic-data maps) that need to be retrievable at any time. The information may be indelibly entered in the memory either by the manufacturer (in one of the last stages of production) or by the user by appropriate programming of specially prepared memories (PROMs or programmable **ROMs**).

Erasable ROM

There are also ROMs whose contents can be erased and reprogrammed as outlined below.

EPROM (Erasable **PROM**)
This type of erasable read-only memory can have its contents completely wiped by irradiation with UV light and can then be reprogrammed. Such reprogramming is only possible with special equipment at relatively high cost.

EEPROM (Electrical **EPROM**)
EEPROM is also referred to as E^2PROM (E squared PROM). This type of erasable read-only memory can be electrically wiped and reprogrammed. The wiping and reprogramming operation can be performed either in a separate unit or in situ. Every storage cell of an EEPROM be individually overwritten. For that reason, this type of memory module can also be used as non-volatile data memory (e.g. for status information in engine-management systems).

Flash EEPROM
Yet another variation of EPROM and EEPROM is the flash EEPROM. In this case, electrical flash pulses are used to erase specific storage areas or the entire contents of the memory. The erased areas can subsequently be reprogrammed.

The flash memory can be reprogrammed on a programming station. However, the advantage of flash EEPROM is that is can also be reprogrammed while still inside the sealed control unit. When this is done, the memory area which contains the programming routines must not be erased, or alternatively, the programming routines must be transferred to the RAM before the memory is wiped. The microcontroller then works with the RAM as the program memory. Flash EEPROM is used wherever relatively large volumes of data need to be stored but also have to be changed from time to time (e.g. in mobile phones, digital cameras and as program memories in electronic control units in motor vehicles).

Opto-electronic components

Photoresistor

A photoresistor is a resistor whose resistance decreases when exposed to light (Fig. 14). Light (photons) generates free charge carriers in the semiconductor as soon as the energy of a photon is equal to the energy holding a charge carrier within the lattice (internal photoelectric effect).

Photoresistors generally consist of polycrystalline semiconductor materials in which the effect is particularly marked such as CdS, CdSe, PbS, PbSe, CdTe, ZnO, Se, InSb, InAs, Ge or Si.

A photoresistor is not equally sensitive to all wavelengths. Its maximum sensitivity lies within a narrow wavelength band that is specific to the material used.

Photoresistors are mainly suitable for use in cameras as light meters.

Photodiode

A photodiode is a semiconductor diode which utilises the depletion-layer photoelectric effect. Reverse voltage is applied to the p-n junction. Incident light releases electrons from the crystal lattice. As a result, additional free electrons and holes are produced. They increase the reverse current (photoelectric current) in proportion to the intensity of the light (Fig. 15).

Photodiodes can be used in light meters, photoelectric beams and in positioning and remote control applications that use infrared light.

Photovoltaic cell

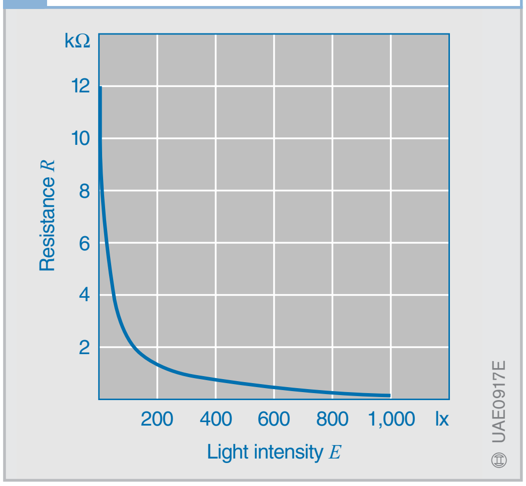
Like the photodiode, the photovoltaic cell releases charge carriers when exposed to light; no external voltage is applied to the p-n junction, however. If the electrons and holes reach the p-n junction, they are separated by the internal field of the space-charge region. A photoelectric voltage is generated which produces a photoelectric current in the external circuit. Light energy is thus converted into electrical energy.

Photovoltaic cells are used to measure light intensity and to generate electricity with solar panels.

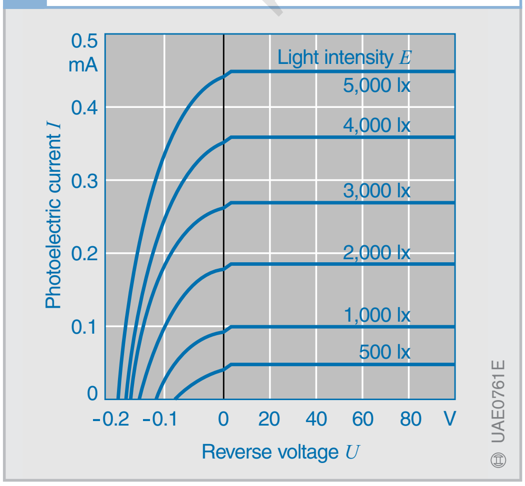
Light-emitting diode

The light-emitting diode or LED emits light when a forward current is passed through it. In other words, it operates in the opposite way to the photodiode effect – free electrons and holes are recombined. The energy released is emitted in the form of light.
The semiconductor materials most commonly used are gallium compounds such as gallium arsenide (GaAs) and gallium phosphide (GaPh). The wavelength of the light emitted by an LED depends on the semiconductor material used. There is now a choice

14 Photoresistor (characteristic)



15 Photodiode (characteristics)



of colours available (infrared, red, yellow, green, blue).

LEDs are used for numerical and alphabetical displays as well as for warning lamps. Since it has been possible more recently to substantially increase the amount of light generated, LEDs are now also used in motor vehicles to illuminate the instrument panel or as a third (high-level) brake light. They have the advantage of a relatively high light yield and, as a result of their very long service life, do not have to be replaced.

Laser diode

With the appropriate mechanical/optical design, LEDs can also be made to emit laser light, i.e. a parallel beam of monochromatic (of a single wavelength) and coherent (the waves are in phase) light. Such properties are required for the transmission of data by fiber-optic cables. Laser diodes were therefore a necessary requirement for the high data flow rates such as occur in computer networks with access to the Internet or the transmission of television signals by cable.

Phototransistor

A similar effect to that used by the photodiode also occurs in the phototransistor. Incident light striking the base alters the conductivity between the emitter and the collector in such a way that a current proportional to the intensity of the light is able to flow through the transistor. This type of component can thus be used as a switch for photoelectric beams, for instance.

There are also phototransistors that can be operated the opposite way around so that they emit light. The advantage compared to the LED is that the phototransistor can be switched on and off very quickly. It is therefore particularly suited for use in the manufacture of flat screens. There is a transistor for every screen dot (determined by the screen raster in a cathode ray tube). As the areas required are generally substantially larger than the silicon slices available, and such slices would, in any case, be much too expensive, flat screens are manufactured using thin-film techniques (vapour deposition through a screen). The advantages of flat screens over cathode-ray tubes are their slimmer dimensions, lower power consumption and the absence of flicker.

Charge-coupled device (CCD)

CCDs are integrated circuits arranged as a charge-coupled array. They are used to record images in video and digital cameras as well as in scanners. Thousands or even millions of individual elements laid out in a matrix pattern (array) are created on a chip (Fig. 16). Digital cameras can record up to 3.5 million pixels with a colour resolution of 16 million colours. CCDs are not only capable of converting the visual information into electrical signals, they can also store the data until it can be transferred to an external storage medium. To do this, the output register (5) is scanned at a high rate.

16 CCD, image-sensor structure

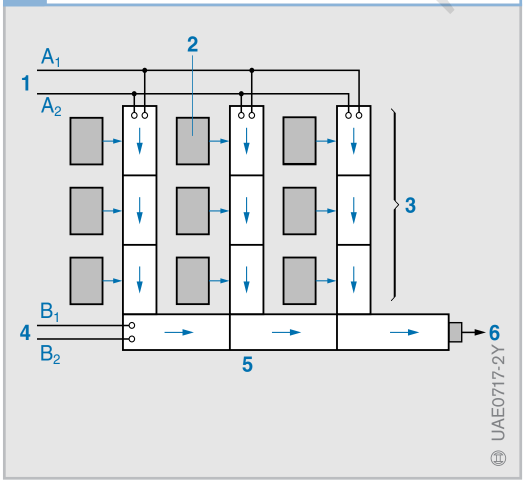


Fig. 16
1 Column clock pulse A_1/A_2
2 Photosensors
3 CCD array
4 Row clock pulse B_1/B_2
5 Output register
6 Video output

Micromechanical sensors

It has been possible for a number of years now to manufacture some sensor components using silicon technology. Since a large number of sensors can be produced simultaneously on a single silicon wafer in similar fashion to electronic components, the unit costs are substantially lower than in the production of conventional sensors. In addition, there is also the possibility of having the sensor signals electronically processed directly on the same chip.

Examples of micromechanical sensors:

- Magnetic-field sensors measure an external magnetic field by purely electrical means.
- Acceleration and pressure sensors measure the relevant variables by the physical deformation of specific areas of the silicon.
- Temperature sensors utilise the change in conductivity of the semiconductor material according to temperature.
- Chemical sensors make use of the effect whereby certain chemicals (gases, liquids) alter the electrical conductivity of some semiconductor materials in a very specific way.

Some of the more important examples of sensors and their design are described below.

Hall-effect sensor for detecting magnetic fields

An electrical potential U_H is generated at the edges of a thin plate through which a current is flowing if a magnetic field is acting perpendicular to the current flow I_V (Fig. 17). It is described by the formula

$$U_H = R_H \cdot I_S \cdot B/d$$

where

U_H Hall-effect voltage
R_H Hall-effect constant
I_V Supply current
B Flux density of the magnetic field
d Thickness of the plate

In metals, the Hall-effect constant is very small: $R_H \approx 10^{-9}$ m³/As. In semiconductors, R_H is five orders of magnitude greater, e.g.

- Indium antimonide (InSb) $R_H \approx 2.4 \cdot 10^{-4}$ m³/As,
- Indium arsenide (InAs) $R_H \approx 1.2 \cdot 10^{-4}$ m³/As.

Hall-effect sensors can be used in motor vehicles as position sensors (e.g. for sensing camshaft position). The advantage of this measurement principle is that position can be detected even when the component concerned is not moving. By comparison, inductive speed sensors only produce an electrical signal above a certain minimum rotational speed.

17 Hall effect

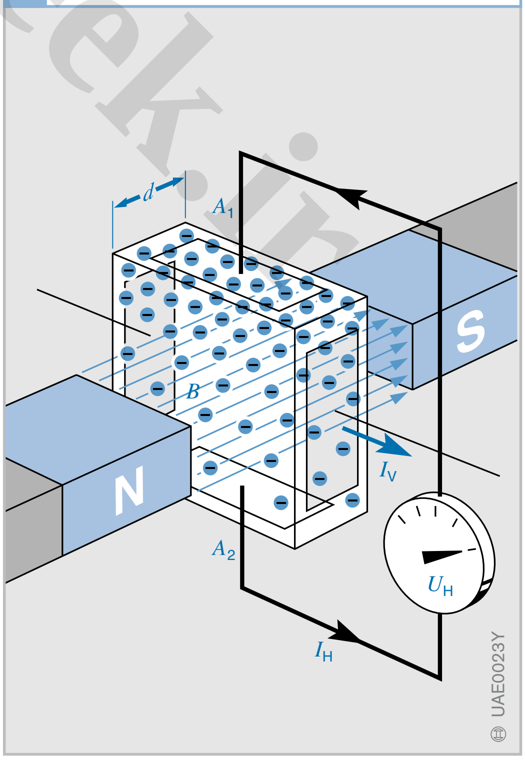


Fig. 17
A_1, A_2 Measurement points for the Hall-effect voltage
B Flux density of the magnetic field
I_V Supply current
I_H Hall-effect current
U_H Hall-effect voltage
d Thickness of the plate

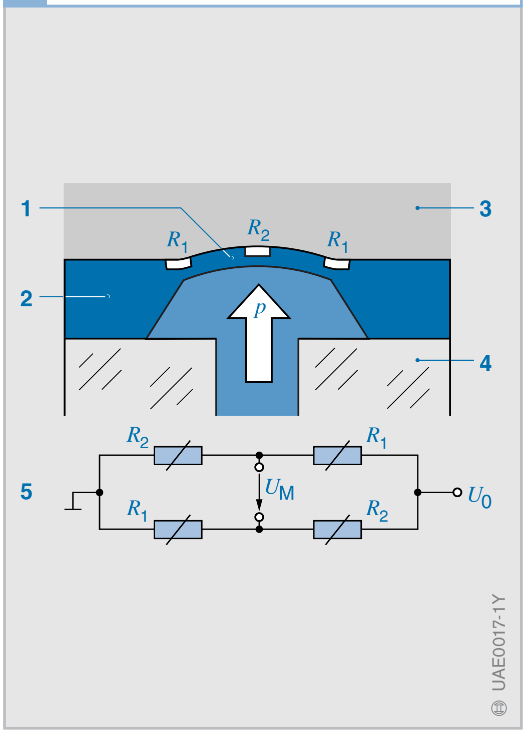
Pressure sensor

Figure 18 shows the structure of a micromechanical pressure sensor. The silicon chip (2) is attached to a glass base (4) through which there is a hole which acts as the pressure channel for the pressure *p* that is to be measured. At the point where the measured pressure acts on the silicon chip, it has been hollowed out on the underside. Consequently, the chip is in effect a thin diaphragm (1) at that point, and deforms under the action of pressure.

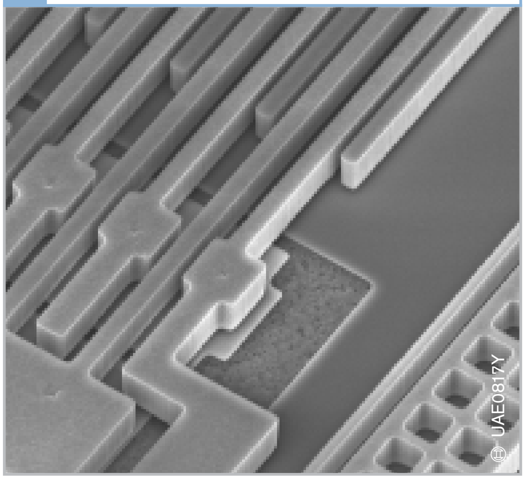
Attached to the top of the diaphragm there are resistors which change their resistance in response to physical deformation (piezoresistive effect). By measuring the resistance, the pressure can be calculated with the aid of the integrated bridge circuit.

Pressure sensors are used in motor vehicles for such tasks as measuring the intake manifold pressure and the atmospheric pressure.

18 Pressure-sensor cell with reference vacuum on the structural side



19 Detail of finger structure of a micromechanical silicon acceleration sensor



Acceleration sensor

Finger-like structures are etched into the surface of a silicon ship (Fig. 19). Every other “finger” is rigidly attached to the chip, while the alternate fingers are only attached at one end so that they can oscillate freely at the other. These fingers are connected up as a multi-layer capacitor. If the sensor is accelerated in a particular direction, the distance between the fixed and the movable fingers changes, thus altering the capacitance. That change can be measured as an electrical signal from which the acceleration can then be calculated.

Acceleration sensors are used in motor vehicles wherever changes in the vehicle’s dynamic behaviour need to be detected.

They are employed, for example, to trigger the airbags in the event of an accident (abrupt deceleration).

The Electronic Stability Program (ESP) uses a yaw rate sensor to detect the rate of rotation of the vehicle around its own axis and counteracts such yawing motion at an early stage by controlled operation of the brakes.

Fig. 18
1 Diaphragm
2 Silicon chip
3 Reference vacuum
4 Glass base (Pyrex)
5 Bridge circuit
U_0 Supply voltage
U_M Measurement voltage
Strain resistors R_1 (compressed) and R_2 (stretched)

Microcontrollers

Microcontrollers are single-chip computers that are used for event-driven control systems or for controlling technical processes. Modern-day life is no longer conceivable without them and they are to be found wherever control systems for electrically operated devices are required.

Microcontroller developments

Five billion microcontrollers were produced in 1999, that is roughly equivalent to one for every person on the Earth. Its triumphant march into every aspect of our day-to-day lives is clearly demonstrated by the number of microcontrollers produced in the last ten years (Fig. 1).

A microcontroller consists of the integration of a CPU (central processing unit) with memories and peripheral circuits on a single chip. These are referred to as "embedded systems". Frequently such systems incorporate a microprocessor that was originally developed for computer applications and subsequently adapted and modified so as to be marketed as part of an embedded system.

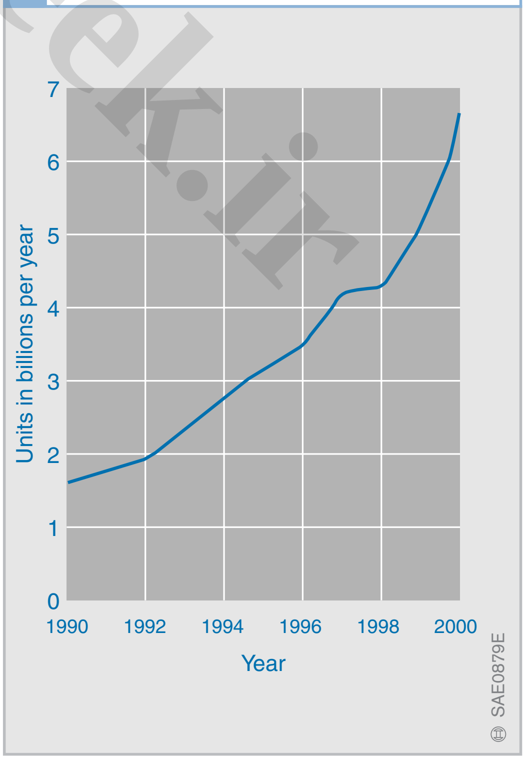
Nevertheless, there are also many microcontroller architectures that were developed from the outset as controllers. They include the M-Core™, the TriCore™ and the SH7000, 8051 and C166 families. Such microcontrollers are characterised by rapid data exchange with their environment, low power consumption, high data density and low production costs.

What then is the actual origin of microcontrollers? Even before the transistor was invented, most of the devices in which microcontrollers are now used were already in existence – cars, telephones, radios, televisions and household appliances. The difference was that for most applications the control systems were mechanical and only in very rare and very costly exceptions were functions controlled by electron tubes and relays. In addition to their high cost, the problem of such control systems was that their reliability was heavily influenced by manufacturing tolerances, environmental conditions and wear.

In the 1960s, Intel, followed shortly afterwards by Motorola, developed the first microprocessors – initially for use in cash registers. Following the development of integrated circuits and the first microprocessors, the inexorable progress of digital technology began in earnest in the 1970s. For the first time, it was now possible to eliminate a large number of interference factors.

The end of the 1970s saw the dawn of the age of the single-chip microcomputer. Prominent representatives of that period were the Intel 8048 and Motorola 6800 microcontrollers, the structure of which is still the basis for many of today's microcontrollers. In the succeeding period, advances in high-integration technology allowed more and more peripheral circuits to be incorporated in the chip. These were the be-

1 Unit production figures for microcontrollers



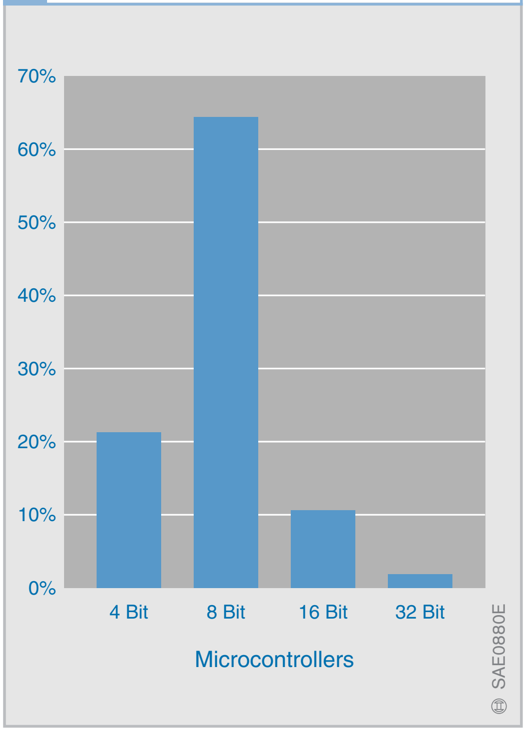
ginnings of the microcontroller as we know it today.

In parallel with these developments, scientists realised that single-chip computers could be used to perform control tasks in motor vehicles. The first applications involving high data volumes were the management systems for internal-combustion engines, followed later on by antilock braking systems (ABS) and entertainment systems.

At the end of the 1980s, the term "embedded control system" was introduced to distinguish them from infinitely programmable systems such as the PC. Embedded systems are designed exclusively for a specific application so that the entire program together with the vehicle-specific data in a ROM or EPROM are a permanent part of the product that cannot be altered by the user.

Most microcontrollers sold today have an 8-bit CPU. Although the numbers of more powerful 16 and 32-bit controllers are continually growing, simple 4-bit designs are still used in large quantities (Fig. 2).

2 Proportion of microcontroller production volume by processor type (1999 figures)



Microcontrollers are produced in a variety of designs for a vast range of applications. The various types differ by virtue of the choice of integrated memory and the peripheral modules offered. The range of integrated memories available encompasses RAM, ROM, EEPROM and, more recently, flash EPROM.

To make it worthwhile to develop a stand-alone microcontroller design, the production figures should run to at least 1 million units a year. The unit price will then be between 0.5 and 15 Euro.

Applications that hold the promise of such production volumes include control systems for automotive applications, communications devices, computer peripherals, domestic appliances and entertainment systems. In addition, microcontrollers are also popular in industrial electronics, automation systems and testing and measuring equipment, although recourse is generally made to standard products.

Microcontroller components

The microcontroller is a programmable electronic module that contains all the necessary components for a microcomputer system. It consists of the

- *CPU,*
- the *memories* for instructions (program memory) and data (data memory), and
- the *peripheral modules.*

Those components exchange data and monitoring information via bus links (Fig. 1).

CPU

The CPU (central processing unit; also referred to as the microcontroller core) is the programmable unit for addressing and manipulating data and for controlling the timing and logical sequence of a program.

Memory

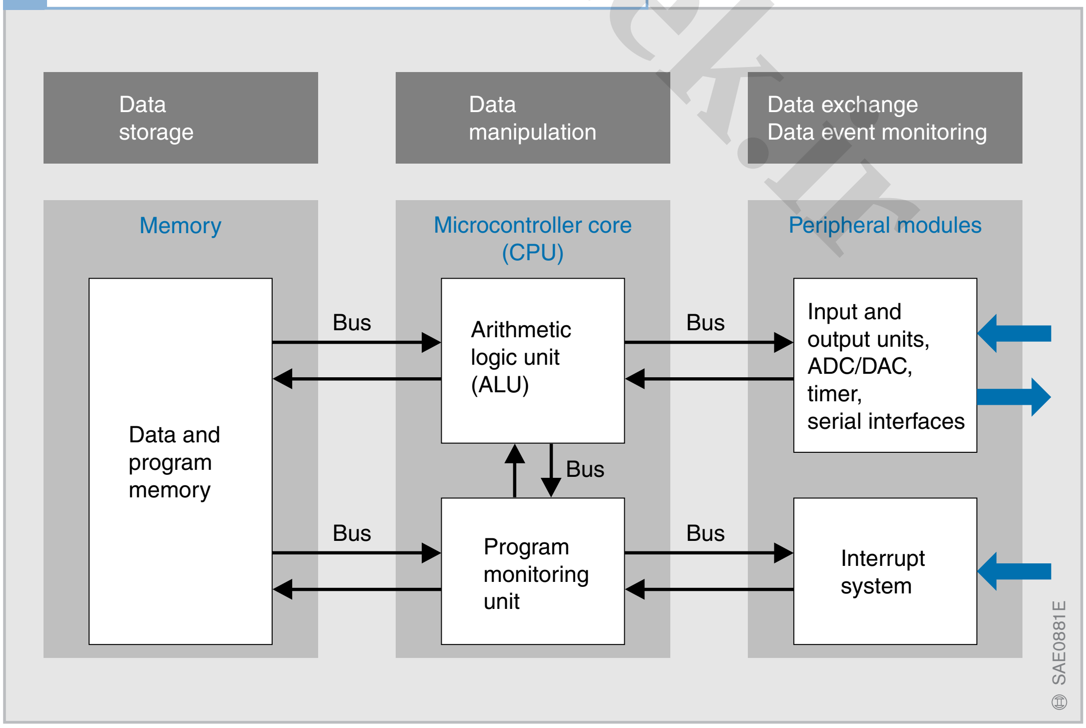
The memory is used to store data and program instructions. The memory for variable data is a random-access memory (RAM). The memory used for program instructions and unvarying data is a read-only memory (e.g. ROM or PROM). In addition, microcontrollers have a small register memory integrated in the CPU for rapid access (cache memory).

Peripherals

The peripherals are used for the input and output of data originating from or destined for external systems. The peripherals are programmable to a limited degree in order that their functions are adaptable to the requirements of the application.

Typical peripheral modules digitise analog external signals or convert internal digital signals into analog signals for output (analog-digital and digital-analog converters). Counters and timers count external pulses and time intervals between events. Communication interfaces are used for the exchange

1 Structure of a microcontroller with the "Von-Neumann architecture"



of data with other modules via standardised bus links (e.g. CAN bus).

There are, of course, many other functions that can be integrated in the microcontroller depending on the requirements of the user concerned.

Main operations

The three blocks illustrated in Figure 1 enable the four main operations of the microcontroller, i.e.

- data manipulation (data processing),
- data storage,
- data exchange with external systems (data movement), and
- data event monitoring (control mechanism).

These functions enable the microcontroller to be used to transfer, store and manipulate data (both in the memory and externally). The sections which follow describe the various modules of the microcontroller that make those operations possible.

Design and operating concept

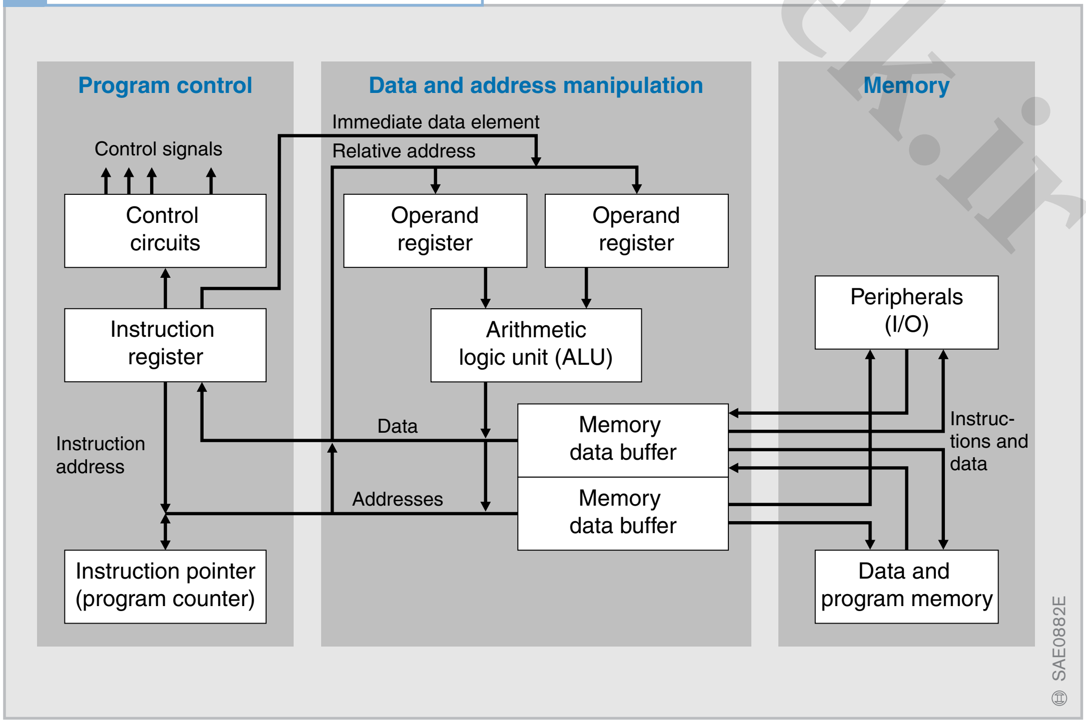
CPU (Central Processing Unit)

The CPU processes the data received from external sources via the peripherals and monitors the data flow. In the CPU there is a small memory (register) in which operands, results and addresses are stored. Figure 2 shows the basic structure of a CPU, which may also be extended by additional components in order to increase the processing speed.

Programming model

The "programming model" of a microcontroller refers to the sum total of all registers that are available to, i.e. are "visible" to, the programmer. In principle, there does not have to be any "visible" register in the CPU. But in that case, every alteration to the program would necessitate modification of the hardware, which would be very involved, expensive and time-consuming. Rarely altered configurations are therefore set by means of bits in special control registers. The control

2 Basic structure of a CPU (microcontroller core)



registers are thus quasi-static extensions to the instructions. The interrupt control register, for example, specifies which interrupts are allowed and which are barred. Other control registers define the function of the ALU (**a**rithmetic **l**ogic **u**nit) or the peripheral modules.

Various operations can alter the operating status of the microcontroller. If, for example, a signal is received from an external source (interrupt), this generally causes the program sequence to divert to a defined memory address. While the interrupt service routine at that location is being processed, only signals with a higher priority can interrupt that procedure. All other interrupt requests are stored and only processed once the interrupt service routine in progress has been completed.

The status information that accrues in the process could be temporarily stored in the instruction memory. However, this would result in very long instructions. For that reason, special registers that store the status of the CPU are integrated in it in addition to the control registers. These status registers include, among others, the program status register, the interrupt status register, the multiplier status word, etc.

In order to reduce the number of loading and storage operations by the microcontroller, the register file architecture incorporates several accumulators (special computation registers in the CPU). These enable interim results or important variables to be held in the CPU. This increases the maximum clock speed and reduces power consumption.

Operand memories

Depending on the instruction-set architecture, there are various possible ways of providing the operands (data to be linked) involved in mathematical calculations or logical operations before and after the computing operation.

Memory-memory architecture

The memory-memory architecture uses the general random-access memory (RAM) to provide the operands. When doing so, it encodes the memory addresses of the operands and the result of a mathematical operation (function that links the operands) explicitly in the instruction. In that way, for example, two numbers that are both stored in the memory can be added together with a single instruction. The result can then be written back to the memory immediately. The term "memory-memory architecture" is derived from the storage location of the operands.

Accumulator architecture

With the accumulator architecture, the CPU has an integral memory cell that is permanently defined as both the source and the destination of every mathematical operation. This memory cell is called the accumulator. Only the address of the second operand is encoded in the instruction. Before every mathematical operation, the first operand must be copied from the memory to the accumulator by a load command. Following the operation, the result is copied back from the accumulator to the memory.

Memory-register architecture

In a memory-register architecture, there is a whole series of special memory cells called registers integrated in the CPU. Both operands are explicitly encoded in the instruction. However, only one of the two operands can be addressed directly in the memory using the memory address. The second operand and the result are addressed in a register. As in the case of the accumulator architecture, one of the operands must be copied from the memory to a register before the mathematical operation is performed. Following the operation, the result must then be written back to the memory. If, however, the number of registers is large enough, interim results can be held in registers rather than being continually copied backwards and forwards. The term "memory-register architecture" is derived from the location of the operands.

Register-register architecture

This last architecture category – also called "load-store architecture" – addresses both operands of an operation explicitly in the registers. Before any mathematical operation, therefore, both operands first have to be loaded into a register. The result is then copied back to the memory.

Operand addresses

Another important distinguishing feature among instruction-set architectures is the possible number of implicitly and explicitly encoded addresses. The operation C = A + B requires three addresses:

- the address of operand A,
- the address of operand B, and
- the address of the result operand C.

Instruction-set architectures which allow free choice of these three addresses (possibility of encoding three addresses) are referred to as "non-destructive instruction-set architectures". But as three addresses normally occupy too many bits in the instruction code, many architectures use an implicit method of addressing.

With implicit addressing, one of the addresses of the two source operands is also used as the destination address. Thus, the address of one of the source operands is used to store the result of the operation, and that means that the operand in question is overwritten, i.e. destroyed. That "destruction" has led to the adoption of the term "destructive instruction-set architecture".

Requirements placed on instructions

Instruction-set architectures differ from one another not only by virtue of their operand memories and operand addresses, but also on the basis of the length of their instructions. The following two very different requirements must be met in this connection:

Number of bits available for encoding

Different operations require very widely varying numbers of bits for encoding instructions of different levels of complexity. For example, the null operation NOP (No Operation) requires no operators (no addresses) nor any additional bits for defining the precise function. The operation MAC (Multiply & Accumulate: A = A + (B × C)), on the other hand, requires three operators and additional bits for encoding the function, such as the behaviour in the event of a mathematical overflow (saturation addition), identifying the data format (Q format), etc.

Decoding complexity

Instructions of differing lengths are difficult to decode. First of all, the length of the current instruction has to be established. Then there is a check as to whether the instruction has been completely read. Both operations generally require several clock cycles and limit the possible processing speed to a substantial degree.

Length of instructions

There are currently three different concepts for the length of instructions.

- *Fixed-length instruction set:* every instruction is the same length, e.g. always 16 bits.
- *Multiple instruction sets:* it is possible to alternate between two different fixed-length instruction sets, e.g. 16 bits and 32 bits. The first bit of each instruction then indicates the format.
- *Multiple instruction length:* the length of the instructions varies according to complexity, e.g. the NOP command can have a length of 16 bits and the intersegment jump a length of 32 bits.

Instruction set

The complete instruction set of a microcontroller enables it to execute any expression of a higher-level programming language (e.g. C) by a sequence of instructions. The instructions are classified as follows:

- *Data processing operations:* mathematical, logical and conversion instructions.
- *Control instructions:* jumps, comparisons, etc.
- *Input/output instructions:* instructions for inputting and outputting data.
- *Memory instructions:* instructions for reading and writing data from/to the memory.

The bits available in an instruction for addressing the operands/result are limited. Embedded applications, however, require ever larger programs and data volumes, which leads to larger and larger numbers of addresses. As a result, the addresses become longer and longer and encoding them becomes more and more complex. There are a number of different methods for solving this problem.

Instruction execution

Execution of instructions on modern microcontrollers

The various phases of instruction execution are illustrated in Figure 3.

- *Fetch-instruction phase(1):* before an instruction can be executed, it has to be fetched from the program memory to the CPU. This involves first calculating its address, which is required in order to load the instruction into the CPU.
- *Decode-instruction phase (2):* the instruction transferred to the CPU is analysed. The relevant function is then initiated.
- *Fetch-operand phase (3):* the operands are fetched from the registers or memories to the executing unit (e.g. ALU).
- *Execute phase (data operations) (4):* once the operands have been transferred to the executing unit, the data operation specified by the instruction is executed.
- *Store-operand phase (5):* the result of the data operation is written back to a register or memory in this phase (which is therefore also called the "write-back phase").

Execution of instructions on older microcontrollers

On older microcontrollers, instructions are processed purely sequentially (CISC architecture: Complex Instruction-Set Computer). This requires a complete clock cycle for each phase:

- Phase 1: calculating the next instruction address and reading the instruction
- Phase 2: decoding the instruction

3 The phases of instruction execution on modern microcontrollers

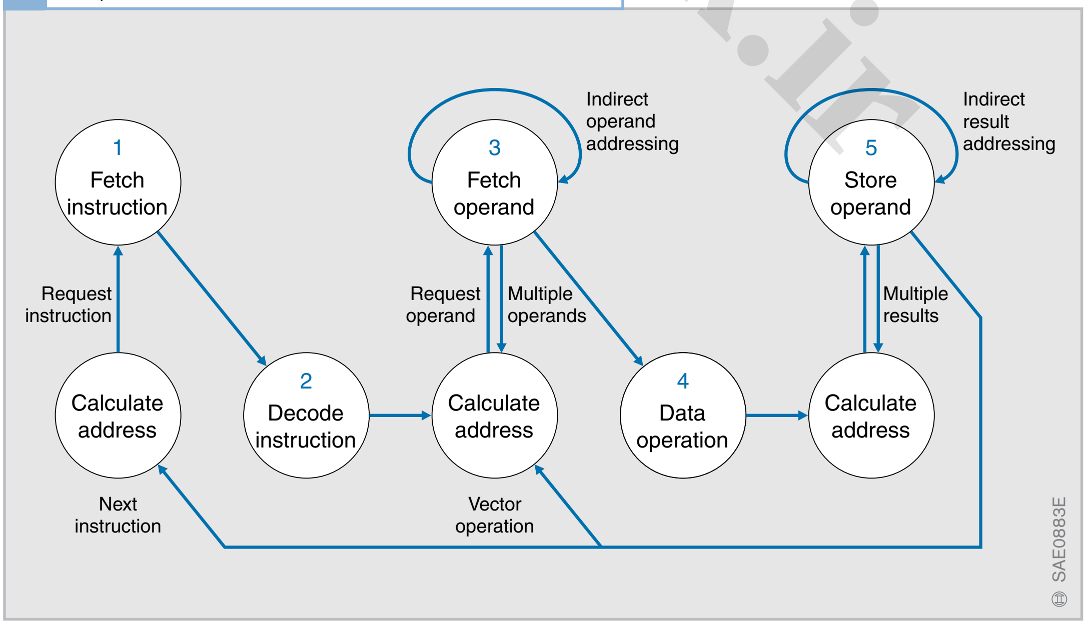


Fig. 3
The numbers indicate the five phases of execution of an instruction

- Phase 3: calculating the operand address and reading the operand
- Phase 4: executing the data operation
- Phase 5: calculating the address and storing the result
- Phase 6: additional cycles for calculating indirect addresses.

Sequential execution is necessary with this process because certain units of the CPU are used a number of times in different phases. The arithmetic logic unit (ALU) in particular is used to calculate all addresses (instruction, operand, result, indirect address) and to execute the data operation. That means that the ALU is in use in every phase.

Parallel execution of instructions
By adding extra address calculation units (address ALUs), the various operations can be performed simultaneously. Thus, while one instruction is being decoded, the next is being addressed and loaded (pipelining). This is the way in which RISC (**R**educed **I**nstruction **S**et **C**omputer) architectures generally operate. A RISC architecture is distinguished by the following characteristics:
- reduced number of instructions,
- large number of general-purpose registers, and
- optimised pipeline.

In a microcontroller with a pipeline architecture, the execution of multiple consecutive instructions is handled simultaneously. Rather like on a conveyor belt, the various instructions are at differing stages of execution and it is not possible to reduce the execution time of individual instructions.

In order to ensure that one instruction is completed in each cycle, all stages of a pipeline must operate simultaneously and independently of one another. Once the CPU pipeline is full, one instruction is completed in each clock cycle.

Microcontroller memories

It is the job of memories to store both data (constants and variables) and the instructions of a program and to make them available when required.

There are a number of different ways in which memories are implemented. This range of alternatives is necessary because there is no single implementation which suits all requirements. In order to satisfy the different requirements, microcontrollers usually have several different memory blocks.

The features by which memories are distinguished include the implementation location, the method of addressing data, the access method and volatility/non-volatility.

Implementation location

- External to the microcontroller (external memory),
- Integrated as separate module on the microcontroller (internal memory) or
- Integrated as a submodule in a peripheral module on the microcontroller (e.g. timer) or in the CPU (register)

Data addressing

- Free access (reading or writing) to all data by allocation of a fixed address (RAM: **r**andom-**a**ccess **m**emory).
- Direct addressing: specification of the memory address by an external source (e.g. the programmer)
- Indirect addressing: the microcontroller calculates the address itself according to external events
- Content-related access to data by addressing based on the data content. For example, address the word whose first bit is "one" (CAM: **c**ontent **a**ddressable **m**emory, or associative memory). This type of memory is used in CAN communication modules, for example.

- Sequential access to data with the aid of a hard-wired buffer structure (e.g. FIFO: first in, first out; LIFO: last in, first out). These memories are primarily used for synchronising separate processor elements (e.g. microcontroller for antilock braking system [ABS] with two CPUs).

Access
- Read-only memory (ROM),
- Random-access memory (RAM), or
- Erasable read-only memory (EPROM).

Volatility/Non-volatility
- Volatile: data is lost when the power supply is switched off, or
- Non-volatile: data is retained when the power supply is switched off

A RAM that has a back-up power supply from a battery can act as a non-volatile memory.

Microcontroller peripheral modules

In addition to the CPU and the memory, the peripheral modules (input/output modules or I/O modules) are essential components of microcontrollers. They make it possible to acquire external signals in a variety of ways and to manipulate the controlled variables by means of outgoing signals. Peripheral modules represent an "intelligent" link between the microcontroller and its environment. Each module has an internal connection to a system bus and external connections to sensors and actuators.

Figure 4 shows a schematic diagram of the structure of a peripheral module. Its tasks can be subdivided into the following:
- Communication with the internal system bus,
- Communication with the environment,
- Data storage,
- Monitoring and timing, and
- Fault detection.

Peripheral modules: Addressing

Isolated peripherals (isolated I/O)
There are two separate address areas for memory and peripherals. Programming of the peripherals is strictly limited in this case as only special instructions can be used for peripheral modules.

Memory-mapped peripherals (memory-mapped I/O)
Peripherals and memory modules share a common address area. The advantage of this is that the large number of instructions for addressing memories can also be used for peripheral modules. However, this uses up valuable address space which can be disadvantageous with 4-bit and 8-bit microcontrollers.

Microcontrollers with 16-bit or 32-bit data width now almost exclusively use memory-mapped peripheral architectures.

Peripheral modules: Operating mode

Another characteristic for distinguishing between I/O modules is the mode of operation, which can be one of the following:

Programmed I/O

The I/O module is controlled directly by the CPU. The microcontroller controls all functions by means of a program. In this case, the CPU has to wait while the I/O unit performs an operation. This mode of operation is therefore only used with microcontrollers that exclusively perform input and output operations (I/O processors).

Polled I/O

The I/O module is capable of performing independent operations and the input/output data is temporarily stored in special buffer memories. The CPU periodically checks the module status and transfers new data as required. This mode of operation is primarily suited to RISC microcontrollers as they only have a software interrupt system.

Although the periodic scanning of the I/O module demands a large proportion of the CPU resources, polled I/O is ideally suited to time-based operating systems (time-triggered architecture or TTA), e.g. OSEK+ (Offene Systeme und deren Schnittstellen für die Elektronik im Kraftfahrzeug [Open Systems and their Interfaces for Electronics in Motor Vehicles]).

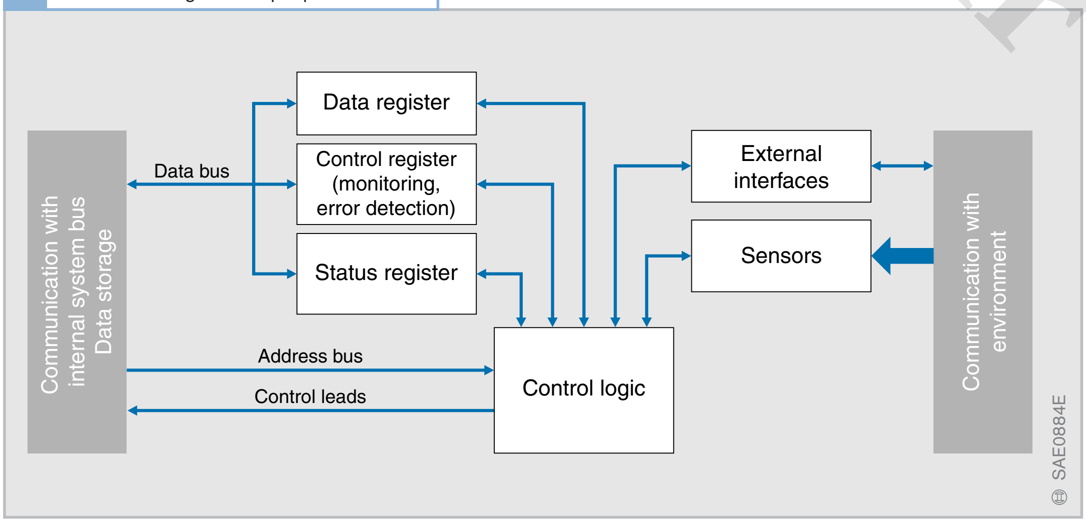
Interrupt-driven I/O

The peripheral module processes all input and output operations independently and signals to the CPU via a dedicated line (interrupt line) when new information is received or when an action by the CPU is required. The basic advantage of this is that the CPU and the peripheral modules can work in parallel. The CPU program only has to be interrupted when the peripherals need assistance. Checking of the programs for errors, however, is extremely time-consuming because all interrupt constellations have to be checked. This architecture supports event-driven operating systems such as Windows CE, OSEK or OS9.

DMA (Direct Memory Access) I/O

With this mode of operation, the I/O modules can exchange data directly with the memory, without the involvement of the CPU. This mode is primarily used for "high-end" microcontroller systems. As with interrupt-driven I/O, this mode of operation requires hardware that prioritises all pending requests.

4 Schematic diagram of a peripheral module



Microcontroller busses

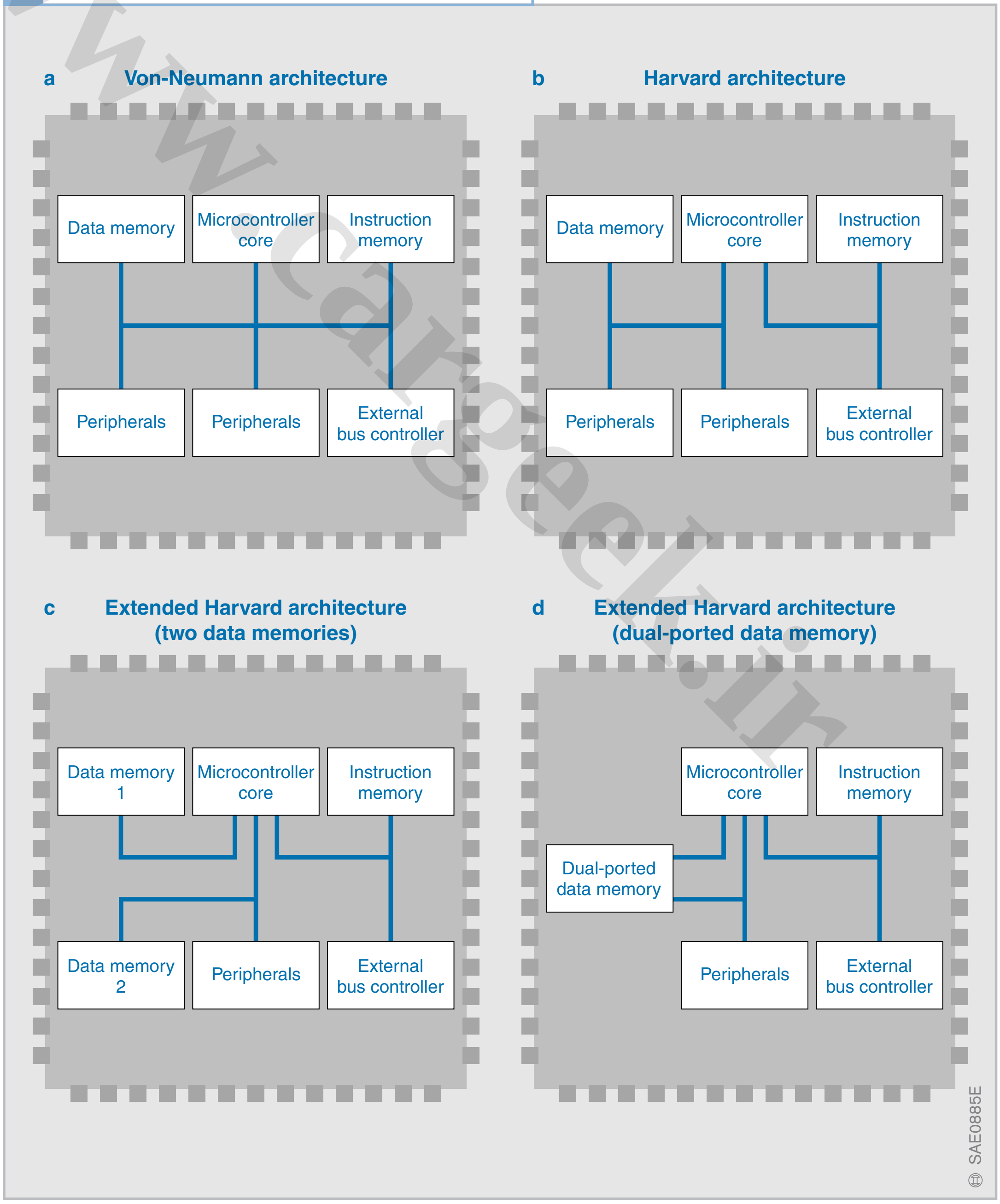
The electrical connections between the modules of a microcontroller (e.g. CPU, peripherals, memory, I/O modules, external bus modules) are referred to as busses.

Design

Busses generally consist of two groups of connections:

- a data bus for transferring data,
- an address bus for transferring the addresses of data, and
- a control line for controlling communication between the sender and receiver of the data.

5 Common bus architectures for embedded microcontrollers



Operating concept

The transfer of data involves the following sequence of operations:

- The data is specified by a signal placed on the address bus.
- Signals on control lines indicate the validity of the address.
- The address is read by the receiver/sender of the data element (data source).
- The sender places the data element on the data bus.
- The sender of the data element signals the validity of the data element.
- The receiver of the data element reads the data.

In order to achieve faster cycle times, these operations can be performed simultaneously on some busses (interleaved or pipelined busses). With interleaved busses, only the unidirectional runtime (drain → source or source → drain) has to be considered, rather than the bi-directional runtime (drain → source → drain).

Common bus architectures

Von-Neumann architecture

If a microcontroller has only *one* bus via which the instructions for the CPU are read and the data is exchanged, this arrangement is referred to as the "Von-Neumann architecture" (Fig. 5a). This architecture is characterised by the small amount of space it requires on the semiconductor chip.

The data memory, the program memory and the I/O modules are arranged across a linear address range. Because the bus is shared, only one signal, one item of data or one instruction can be read from the memory at a time. Instructions that read data from the memory therefore require at least two cycles.

Harvard architecture

Where a microcontroller has two busses (for data communication and transfer of instructions), it is said to have a "Harvard architecture" (Fig. 5b).

The Harvard architecture uses separate address areas for data memory and peripherals on the one hand and program memory on the other, with each area being accessed by a separate bus. This means that an instruction can be read at the same time as an item of data is read or written. For embedded applications, however, it has become evident that separation reduces performance. Most Harvard architectures therefore incorporate bridges which allow transfer of data between the busses.

Digital signal processing generally involves operations performed on two operands (e.g. filter coefficient, item of data). In order not to have to read these two operands one after the other via the same bus, multiple data bus arrangements referred to as *extended Harvard architectures* were introduced. With such arrangements, two items of data can be read simultaneously.

Extended Harvard architectures may also have two different data memories connected to the two data busses (Fig. 5c) or one data memory with two simultaneous access paths (dual-ported data memory) as illustrated in Figure 5d.

Circuit example

The interaction between a circuit and microcontrollers (µC) is best explained by means of an example. The circuit used in the example chosen is referred to as a development circuit because it is used to test user programs developed on a PC. Such circuits are sometimes also called evaluation boards.

Overview

A development circuit is part of a development system used in conjunction with a PC. Figure 1 shows the overall layout of a development workstation with the circuit described in this example.
The advantage of a development circuit is that all components are easily accessible. In addition, the signals of the various inputs and outputs can be tapped at the terminals. In that way individual signals can easily be checked.

Development environment

Program development

The user program is created on a PC using a text editor (source code, Fig. 2a). A compiler then translates the source code from the programming language in which it is written (e.g. C++) into an assembly language based on mnemonics. The assembly language is the preliminary stage to the machine language and is still easy to read. An assembler is then used to convert the assembly language into a digital code that can be understood by the development circuit's microcontroller. The precise memory addresses are specified by a linker. The machine code or object code can then be transferred to the circuit by the PC.

Assembler programs can also generate listing files in which each line of the original source code is shown alongside the corresponding object code. This is useful for locating errors.

1 Components of a development workstation with the development circuit

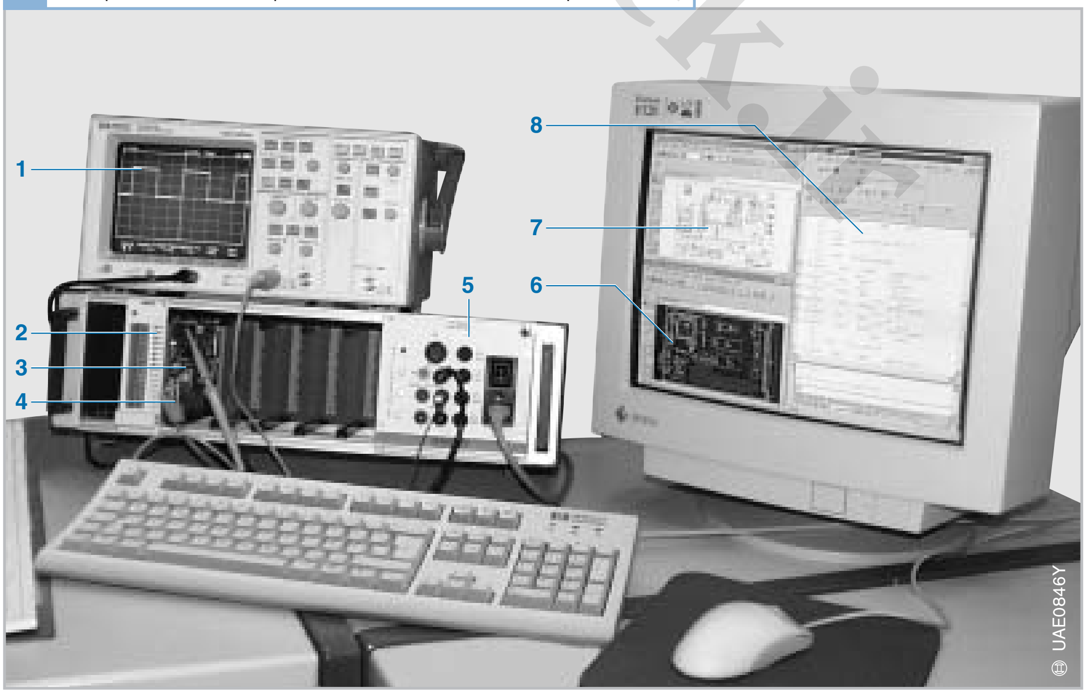


Fig. 1
1 Circuit signal on the oscilloscope
2 Output indicators (LEDs)
3 Development circuit
4 Connection to PC (connector X2)
5 19-inch module rack system with power supply unit
6 Circuit layout
7 Circuit diagram
8 Example program

Transferring the object code

The object code is transferred (downloaded) by starting a program on the PC that transmits the file containing the object code via the PC's RS232 interface. This interface is connected by a lead and the connector X2 to the circuit (Fig. 2b).

Monitoring program (basic program)

A fixed monitoring program running on the microcontroller reads the data sent by the PC and stores the object code in the development circuit's RAM (external RAM).

The monitoring program is stored in the ROM (external EPROM) and is started when the power supply is switched on. It can also send data back to the PC if requested to do so by the PC.

Control units used for automotive applications do not require a monitoring program. They run an invariable program that is stored in the non-volatile memory (e.g. EPROM).

Hardware configuration

Following download, the user program is stored in the external RAM module. Normally this type of memory module is used only for data storage. In development circuits, however, the components are connected in such a way that the RAM can operate as a program memory as well. This makes it possible to download the user program from the PC and store it on the development circuit without having to alter the actual program memory (the EPROM).

Starting the user program

To start the user program, the monitoring program that the microcontroller runs has to branch to the RAM. The start command and start address are entered on the PC and transferred to the development circuit ("g address"; e.g. g 8000, i.e. go to address 8000). The monitoring program reads this information and then executes the instruction to go to the user program (Fig. 2c).

Stopping the user program

The user program can be stopped by pressing the reset button which forces the microcontroller to revert to the monitoring program (Fig. 2d).

2 Working with the development circuit

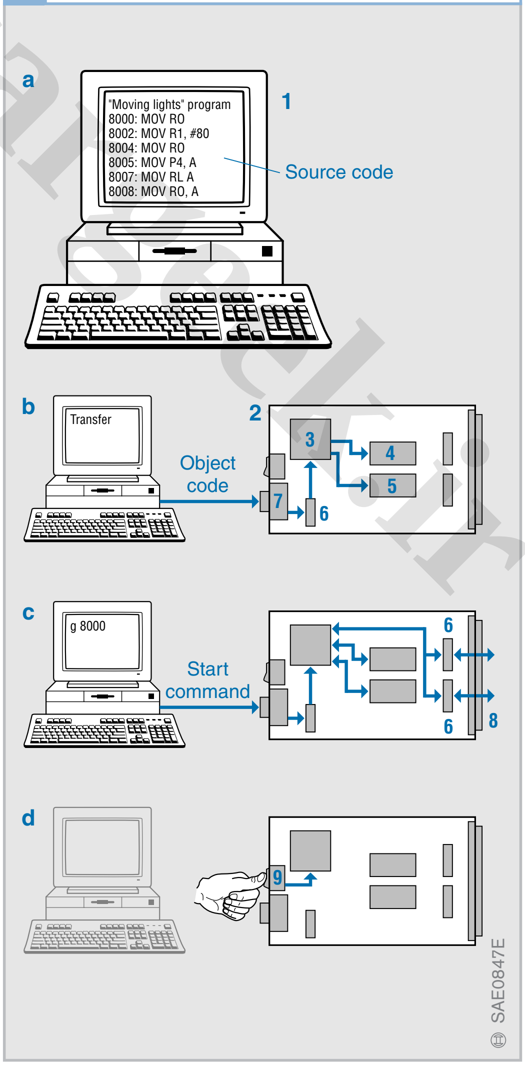


Fig. 2
- a Program creation
- b Transfer of object code
- c Starting and running the program
- d Stopping the program

1. PC
2. Development circuit
3. Microcontroller
4. EPROM with monitoring program
5. RAM
6. Interface module
7. Connector X2
8. Connector X1
9. Reset button

Design and construction

The development circuit is a stand-alone system. That means it is self-contained and can operate independently. A connection to the PC is required only to download and start the program.

Figure 1 at the end of this section shows the circuit diagram for the development circuit. The layout of the circuit and the arrangement of components can be seen in Figure 2. The full list of circuit components is shown in Table 3.

Control units in motor vehicles are, for the most part, fitted with SMDs (**s**urface **m**ounted **d**evices). They contain more signal input and output components than the circuit example shown here.

Microcontroller (IC1)

The central component of the development circuit is the 80C535 8-bit microcontroller (CMOS module). It belongs to the 8051 family of processors. Modules of this type are used in a wide variety of applications, the most important of which are

- industrial electronics (machine controllers, measurement and control equipment),
- electronic entertainment systems,
- data processing and
- automotive electronics (control units for gasoline and diesel engines, ABS, etc.).

8-bit, 16-bit and 32-bit microcontrollers are used in automotive applications. The fundamental design and method of operation of those microcontrollers is the same. However, the 8-bit microcontroller is not as complex and is therefore easier to use as an example of the interrelationships.

The microcontroller controls the running of the program. In addition to the CPU (**c**entral **p**rocessing **u**nit), it also contains input and output channels (I/O ports), timer units, RAM, serial interfaces and other peripheral modules integrated on a common microchip. There are also microcontrollers which have the program memory on the same chip.

An integrated oscillator circuit that is connected to an external quartz crystal generates the system clock pulse.

The 80C535 microcontroller (μC) has a total of seven I/O ports. They are designated port P0 to port P6. Each of those ports consists of 8 bits. Each bit is designated by a digit from 0 to 7 after a decimal point following the port designation (e.g. P0.0 = first bit of port P0). Many of the port connection pins have more than one function. Such alternative functions cannot be used simultaneously. For example, the 16 leads P0.0 to P0.7 and P2.0 to P2.7 are used to transfer addresses and data if external program or data memories are employed (address/data bus function). If only the microcontroller's integral memories (RAM, ROM) are accessed, those leads are spare and can be used as I/O ports for the input/output of signals.

Temporary storage memory (IC2)

In order to reduce wiring, the transfer of addresses and data can staggered so that the same cable can be used for both operations. Such "multiplex operation" requires a memory in which the addresses can be stored temporarily so that addresses and data are simultaneously available.

The 74HCT573 8-bit latch serves as a temporary store for addresses. Copying of the addresses sent by IC1 to the latch is controlled by the ALE (**A**ddress **L**atch **E**nable) signal.

EPROM (IC3)

The EPROM (**E**rasable **P**rogrammable **R**ead-**O**nly **M**emory) is a 27C256 32-kbyte EPROM (32 × 8 kbyte = 256 kbit). The EPROM on the development circuit board holds the monitoring program.
The microcontroller accesses the EPROM via the address/data bus. It executes the monitoring program instructions sequentially. That information is invariable data which cannot be altered during operation.

The EPROM can have its contents erased by irradiation with UV light and can then be reprogrammed using a programming device.

RAM (IC4)

The 62256 32-kbyte RAM (**R**andom-**A**ccess **M**emory) acts as supplementary working memory in addition to the microcontroller's internal RAM for storing variable data (variables such as calculated values and signals) and as a program memory. The RAM is connected to the microcontroller via the address/data bus.

When the power supply is switched off, the RAM loses all the information placed in it (volatile memory). For that reason, this circuit incorporates a back-up battery (see section headed "Back-up battery" overleaf).

Level converter (IC5)

For serial data transmission, the PC has an interface which enables it to communicate with other devices. That interface conforms to the RS232 standard. The interface cable can carry TxD (**T**ransmit **D**ata) and RxD (**R**eceive **D**ata) signals. However, the signal levels of these transmissions are unsuitable for use with the microcontroller. The MAX232 level converter or interface module adjusts the signal levels used by the microcontroller's serial interface (0 ... 5 V) to those used by the PC RS232 interface. The capacitors C11 ... C14 function as voltage stores.

Logic modules (IC6, IC7)

The NAND (negative AND) gates of the 74HCT132 (IC6) process the control signals from the microcontroller. Those gates form a logic circuit that permits access to the following physical addresses on the two memory modules IC3 and IC4:

- EPROM addresses $0000_H ... 7FFF_H$ and
- RAM addresses $8000_H ... FFFF_H$.

The two gates on the 74HCT125 (IC7) process the control signals RD\ and WR\ from the microcontroller and output them via the connector X1.

Output modules (IC8, IC9)

Addresses and data are outputted via the 74HCT245 bi-directional power drivers. They are thus available at connector X1 in a 19-inch module rack system. Other peripheral devices can be connected via these outputs.

On IC8, the lead for the direction of transmission (pin DIR) is set to logical 1. Consequently, this module operates unidirectionally. It sends the addresses A0 to A7.

IC9 transmits and receives data (bi-directional transfer). Therefore, the direction of transmission has to be controlled by IC1 via the DIR pin.

Control of ports 4 and 5

These ports perform typical microcontroller functions. Both ports are buffered by a 74LS245 bi-directional 8-bit driver. It supplies sufficient power for controlling the consumer units (e.g. LEDs) (cf. bipolar transistor control: low base current, high collector current).

Control of the two ports is illustrated in an example program at the end of this chapter.

Discrete components

Clock pulse generator

The quartz crystal Q1 operates as a clock pulse generator for the oscillator integrated in the microcontroller. The capacitors C15 and C16 are also required. The clock frequency is 12 MHz.

Reset button

As its name suggests, the reset button S1 resets the microcontroller. After a reset, the microcontroller starts running programs from address 0000_H. That is why the monitoring program has to start at that address in our example.

The controller board has a straightforward power-up reset. When the power is switched on, capacitor C17 first has to be charged via resistor R1. The voltage at C17 – and therefore at the microcontroller's reset pin – increases from 0 V to +5 V. Above a certain threshold level, reset mode is deactivated. This time delay ensures that the power supply voltage at the microcontroller is stable when the program starts after deactivation of the reset mode.

On/Off switch

When the On/Off switch S2 is on, the red LED D1 lights up. The resistor R2 is required in order to limit the current through the LED.

Capacitors

The capacitors C1 ... C9 filter out interference signals. They need to be as close as possible to the relevant component.

Multipin connector strips

The individual signals can be tapped directly from the multipin connectors J1 ... J6 on the circuit board (see circuit diagram).

Connectors

X1: The connector X1 provides the connection with the 19-inch module rack system (Table 1).
X2: Serial communication with the PC takes place via connector X2 (Table 2).

Back-up battery

While the circuit is connected to an external power supply, the 3.6 V battery (BAT) is charged up because the power supply voltage is 5 V. If the external power supply is switched off, the battery takes over power supply to the RAM module. As the 62256 module (IC2) is a low-power module (L or L-SL version), the power supply from the battery is sufficient to prevent data loss. The diode D2 prevents the remaining components from drawing power from the battery.

Table 1
– = Not used

1 Pin assignment, connector X1

Pin	Row A	Row C	Description
1	+5 V	+5 V	5 Volt power supply
2	–	–	
3	–	–	
4	D0	D1	8-bit data bus
5	D2	D3	
6	D4	D5	
7	D6	D7	
8	–	–	Not used
9	IOW\	–	Control signal input/output Write (low active)
10	IOR\	–	Control signal input/output Read (low active)
11	–	–	
12	P5.0	P5.1	Port P5, 8-bit
13	P5.2	–	
14	P5.3	P5.4	
15	P5.5	P5.6	
16	P5.7	A0	Lower byte of address bus A0 to A7
17	A1	A2	
18	A3	A4	
19	A5	A6	
20	A7	P4.0	Port P4, 8-bit
21	P4.1	P4.2	
22	P4.3	P4.4	
23	P4.5	P4.6	
24	P4.7	–	
25	–	–	Not used
26	–	–	
27	P6.7	P6.6	Port P6, A/D converter port
28	P6.5	P6.4	
29	P6.3	P6.2	
30	P6.1	P6.0	
31	–	–	Not used
32	Gnd	Gnd	Ground

Table 2

2 Pin assignment of serial cable for connector X2

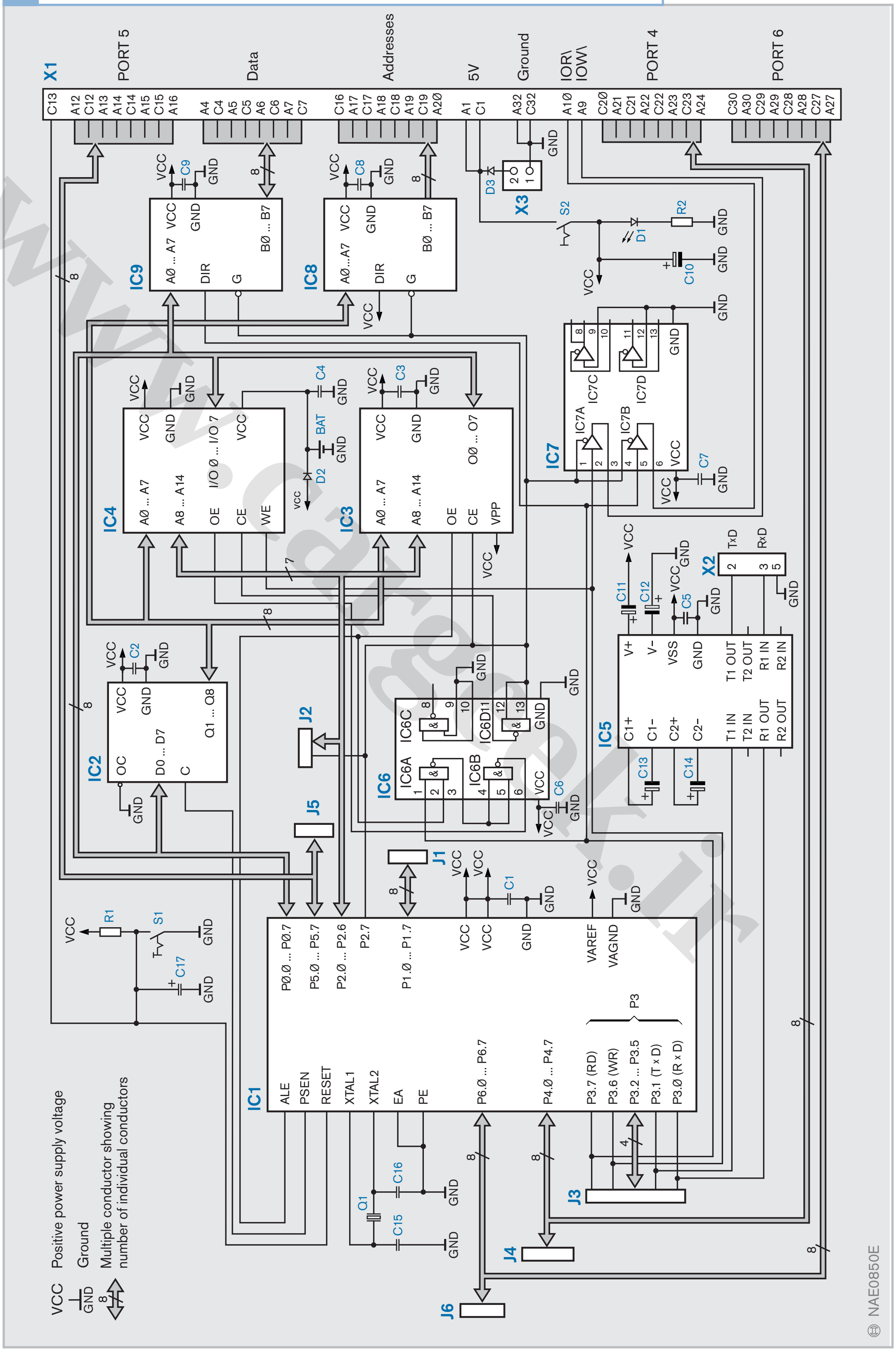
Lead	Controller	PC connector 9-pin	25-pin
TxD	2	2	2
RxD	3	3	3
Ground	5	5	7

3 Component list for circuit example

Name	Designation	Description	Quantity
ICs			
IC1	80C535	CMOS microcontroller	1
IC2	74HCT573	8-bit D flip-flop	1
IC3	27C256	32-kbyte EPROM	1
IC4	62256	32-kbyte RAM	1
IC5	MAX232	RS232 interface module	1
IC6	74HCT132	NAND Schmitt trigger	1
IC7	74HCT125	unidirectional 8-bit bus driver	1
IC8	74HCT245	bi-directional 8-bit bus driver	1
IC9	74HCT245	bi-directional 8-bit bus driver	1
Capacitors			
C1	100 nF (SMD)	SMD interference-suppression capacitor	1
C2...C9	100 nF (ceramic)	interference-suppression capacitor	8
C10	1000 µF (ELKO)	Smoothing capacitor	1
C11...C14	10 µF (Tantal-ELKO)	Storage capacitor	4
C15, C16	22 pF (ceramic)	Capacitor for quartz crystal	2
C17	10 µF (ELKO)	Power-up capacitor	1
Diodes			
D1	LED (red)	"Power supply on" indicator	1
D2	1N4001	Blocking diode for battery	1
D3	1N4001	Polarity-reversal protection diode for X3	1
Resistors			
R1	68 kΩ (0.25 W)	Power-up resistor	1
R2	820 Ω (0.25 W)	Series resistor for LED D1	1
Quartz crystal and battery			
Q1	12-MHz quartz crystal	Quartz crystal for clock pulse generator	1
BAT	3.6-V battery	Back-up battery for the RAM	1
Switches and buttons			
S1	Button	Reset button	1
S2	Switch	On/Off switch	1
Connectors and sockets			
X1	64-pin male connector	DIN 41 612 type C, to module rack	1
X2	9-pin SUB-D socket	Connection to PC (RS232 interface)	1
X3	2-pin connector	For external power supply (optional)	1
J1...J6	8-pin strip connector	Connections for ports	6
B1	PLCC68	68-pin IC socket for IC1	1
B3, B4	DIL28	28-pin IC sockets for IC3 and IC4	2
B2, B8, B9	DIL20	20-pin IC sockets for IC2, IC8 and IC9	3
B5	DIL16	16-pin IC socket for IC5	1
B6, B7	DIL14	14-pin IC sockets for IC6 and IC7	2

Table 3

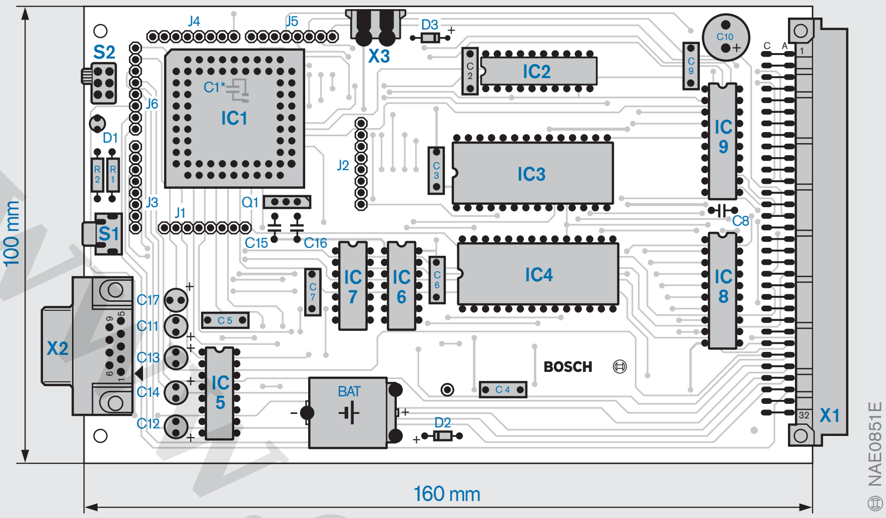
1 Circuit diagram of development circuit with 80C535 microcontroller (circuit example)



2 Component layout and circuit layout of development circuit with 80C535 microcontroller

Component layout

This shows the arrangement of the components on the circuit board. The black dots represent holes in the board. They have tin-plated sleeves inserted in them which provide the electrical connection between the conductor tracks on the two sides of the board.

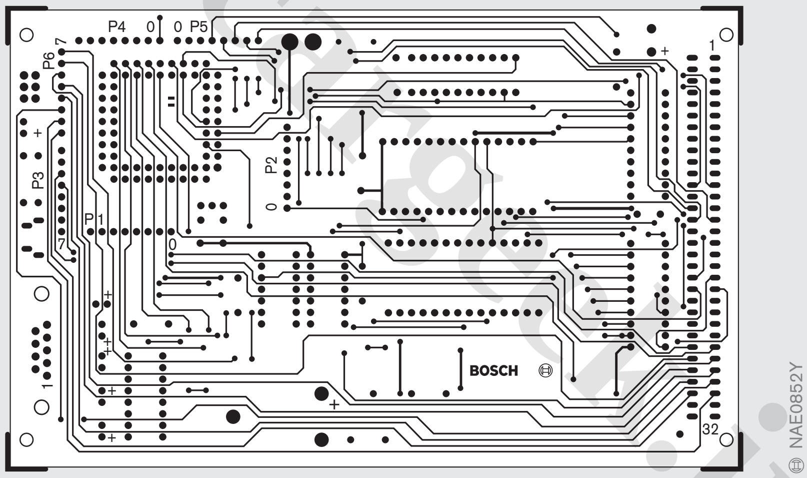


* C1 is fitted on the solder side.

Circuit layout

Component side (top side)

The circuit layout shows the arrangement of conductor tracks. They are created by an etching process. The circuit layout also shows the position of the holes in the board and the soldering points. The soldering points and the holes have to be positioned extremely accurately. For the ICs, the position of the holes relative to each other is also critical.



Solder side (underside)

The solder side is normally represented as if seen through the circuit board. It is thus a mirror image of the true view of the finished circuit board.

The component grid pitch is 2.54 mm.

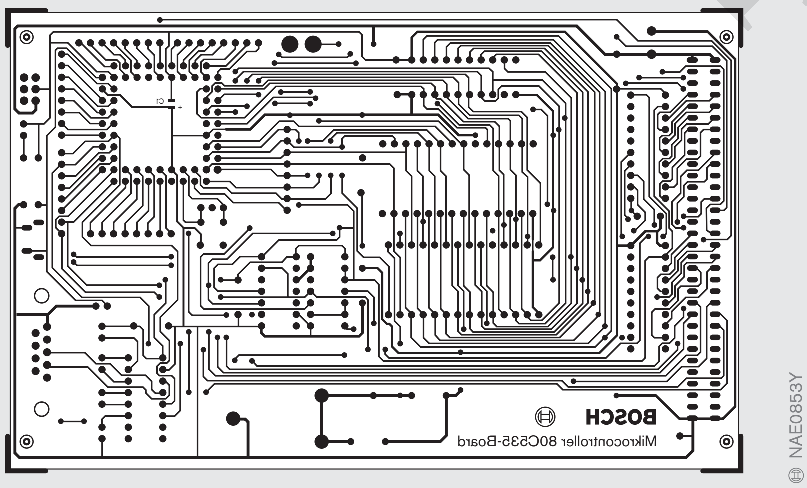


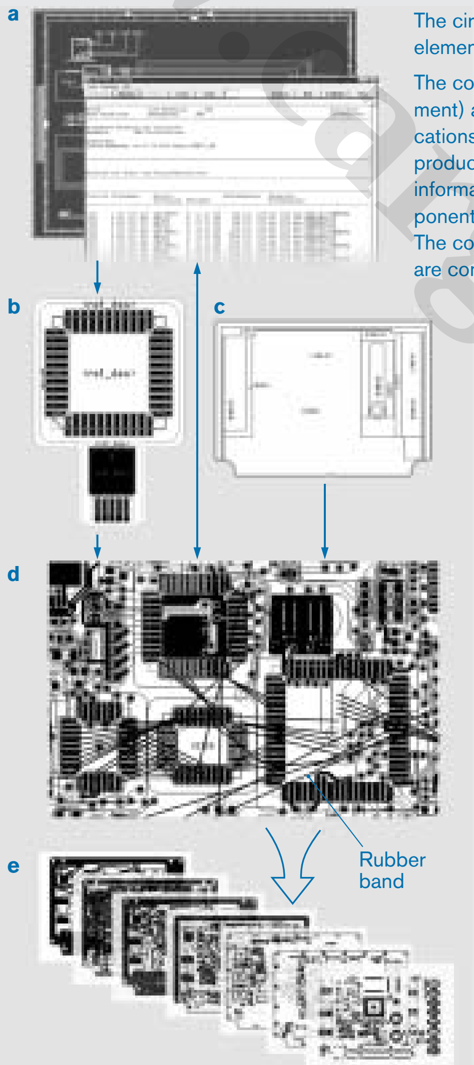
Fig. 2
The layout shown here is a for a circuit board printed on both sides. Only this type of circuit board allows all connections to be made on the plug-in card format without the additional use of copper wiring.

From circuit diagram to layout

Originally, printed circuit boards were simply bases for conductor tracks and components. The ever increasing density of circuits has seen them develop into electronic components, however. Assisted by computer-aided systems, the circuit board designer incorporates the requirements of all disciplines within the circuit board layout during the design process. The considerations involved include circuit development (e.g. circuit diagram), casing design (e.g. dimensions), manufacturing (e.g. testing points, production costs) and last but not least the requirements of the vehicle manufacturer (e.g. connector configuration).

Even in the age of EDA (**E**lectronic **D**esign **A**utomation), only a small part of the design process can be truly automated. A controlled design environment and a controlled design process make it possible to eliminate typical errors and to achieve reproducible results. Design quality is attained by a controlled design process with automated and manual checks.

Stages in the development of a circuit-board layout (design process)



The circuit diagram and the list of components itemising all circuit elements are the starting point for the circuit-board layout (a).

The component details (e.g. dimensions, connector-pin assignment) are entered in a CAD* library (b). Those component specifications define the decisive parameters for circuit-board design, production documents, CAM processing and manufacture. That information and the circuit-diagram information is entered in a component connection schedule on the circuit-layout design system. The component connection schedule details which connections are connected to which.

During this phase, the physical design of the circuit board (c) also has to be considered (e.g. dimensions, cut-outs, connector positions, restricted zones in which no components, or only very small ones, may be placed).

The component positions are then decided upon on the basis of thermal, EMC, production system and physical considerations (d). On the CAE* system, the electrical connections are represented by straight lines (referred to as “rubber bands”) drawn directly between two connected points. Those rubber bands then have to be rearranged as conductor tracks on the circuit board. This involves “untangling” them so as to produce the pattern of conductor tracks. In this process, EMC and circuit design concepts are taken into consideration. Multilayer circuit boards are generally required. Any inconsistencies with design principles or specifications are indicated by the error marker.

Next, the production documents for manufacturing the circuit board are produced (e), i.e. films of the conductor tracks, print films, connection schedules for testing equipment, component mounting diagrams and films for special prints (e.g. solder resist, solder paste).

* CAD: **C**omputer-**A**ided **D**esign;
CAM: **C**omputer-**A**ided **M**anufacturing;
CAE: **C**omputer-**A**ided **E**ngineering.

NAE0854E

Operating concept

Communication

The system bus controls the communication of modules with one another. It consists of the address bus and the data bus.

A backslash in the designation of a control line (e.g. PSEN\) indicates that the active status is represented by the low-level signal. That means that when inactive, the line is set to logical 1.

Address bus

The address bus consists of 16 leads (A0 to A15) via which a physical range of 65,536 addresses can be accessed (2^{16} bits).

As the development circuit has an external data and program memory separate from the microcontroller, the higher addresses (A8 ... A15) are outputted via port P2. Port P0 operates in multiplex mode and alternately outputs the lower addresses (A0...A7) and reads or writes 8-bit data.

The 80C535 (IC1) can address a total of 64 kbyte of program memory (code memory) and 64 kbyte of data memory. The program memory is accessed by a low-level signal on the PSEN\ lead (Pin 49), while the data memory is accessed by the signals RD\ (Read) or WR\ (Write) (Pin 27/Pin 28).

It is a typical characteristic of microcontrollers of the 8051 family that data and program are addressed separately (Harvard architecture).

Data bus

The data bus provides bi-directional data communication since it carries data that is both sent and received by the microcontroller. The data bus width is 8 bits. It is made up of the data leads D0 ... D7 which work with the address leads A0 ... A7 in multiplex mode.

Memory architecture

The memories are subdivided into program memories and data memories. The 80C535 microcontroller (IC1) has an internal program memory and data memory. It is also possible to connect it to external program and data memories (Fig. 1 overleaf). The microcontroller uses appropriate control signals to access the various memories.

The data memory is made up of an internal 256-byte RAM and an external 32-kbyte RAM. They can be accessed by means of MOV commands (see section entitled "Programming"). The SFRs (special function registers) control the special functions of the controller (e. g. A/D converter, timer, serial interface). The external RAM can be accessed by the command MOVX. Accessing the external RAM is slower than accessing the internal RAM.

The external data memory is used by the microcontroller to store data if the capacity of the internal RAM is insufficient for more involved applications.

Memory allocation

The overall addressable memory area of 64 kbyte is split between a 32-kbyte EPROM (IC3) and a 32-kbyte RAM (IC4). Both memories are connected in parallel to the address leads A0 to A14. Lead A15 is used to select the memory module.

Memory area 0000_H to $7FFF_H$

For the range 0000_H ... $7FFF_H$ the address lead A15 is set to low. This lead is connected to the CE\ (Chip Enable) pin of the EPROM. Therefore, this range of addresses accesses the EPROM. The CS\ (Chip Select) pin of the RAM receives the opposite signal. Consequently, the RAM module is deactivated within this address range.

Following a reset, the program starts from address 0000H and thus fetches instructions from the EPROM. This range of addresses contains the code for the monitoring program.

Memory area 8000_H to $FFFF_H$

For the upper 32-kbyte range, the A15 signal is set to high. Thus, within this range, the CS\ pin of the RAM is active. The RAM is accessed both for data retrieval and for reading program code. The program memory function is enabled by the logical combination (IC6A and IC6B) of PSEN\ and RD\. Data can also be written to the RAM.

The user program is stored in this area of memory. The process of program development must ensure that data is not stored in the area occupied by program code. Writing of data to the memory would otherwise overwrite and thereby destroy the program.

Control lines

The microcontroller outputs control signals via the control lines. They ensure that the necessary components are selected at the appropriate times. On the 80C535 (IC1) the control signals are as follows:

ALE (Address Latch Enable)

The microcontroller generates the ALE signal to control the temporary storage memory (latch, IC2) which stores the address low byte. When accessing the program memory, the microcontroller applies the address high byte to port P2 which is directly connected to the program memory. It applies the address low byte to port P0. These eight pins are connected to the temporary storage memory (latch). The latch is "transparent" when the level of the ALE signal is high. That means that the full address is applied to the program memory. The data outputs of the program memory which are connected in parallel with the inputs of the latch are at high resistance so as to prevent data collision.

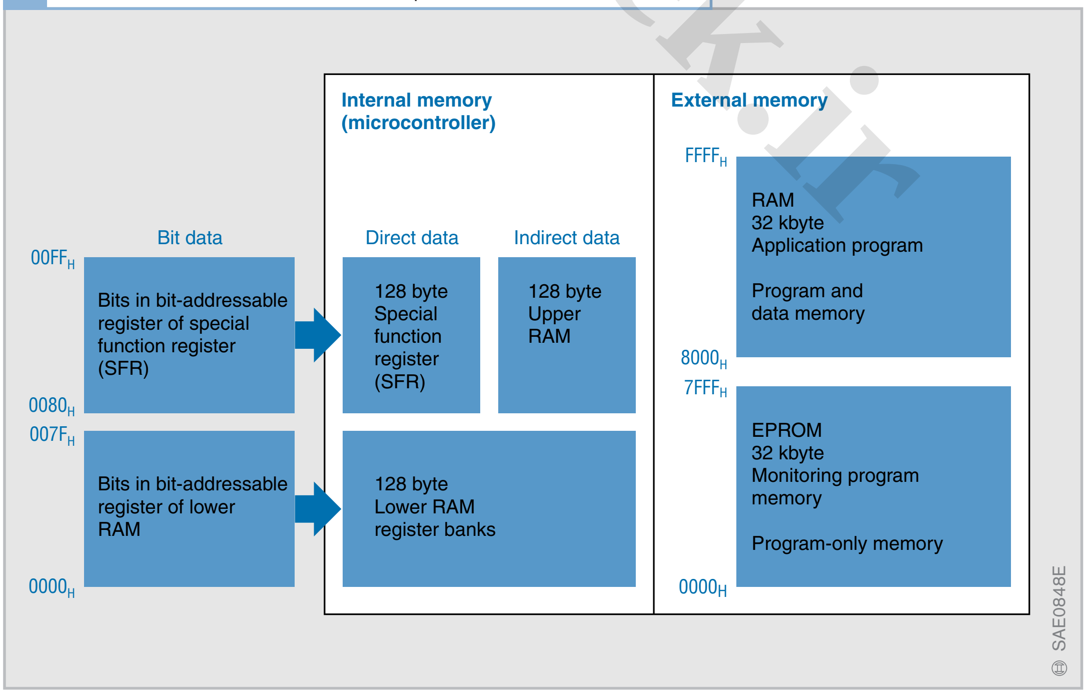
The microcontroller subsequently sets the level of the ALE signal to low. The address low byte then remains stored at the latch outputs. The full address remains applied to the program memory for the entire command cycle.

The ALE signal is also generated for data access to external data memories.

PSEN\(Program Store Enable)

When an address is applied to the program memory, the instruction code stored at that location can be read. In order to do that, the microcontroller generates the PSEN\ signal. This signal controls the OE\ (Output Enable) pin of the program memory. When the signal status is active (low level), the program memory applies the data to the data bus which the microcontroller reads via port

1 Overview of memories in the circuit example with 80C535 microcontroller



P0. The PSEN\ signal then switches back to inactive status and the data lines of the program memory revert to a high resistance level.

RD\

To read the data memory (external RAM), the microcontroller must first access the module. This is done in the same way as when reading the program memory. In this case too, the address low byte is temporarily stored in the latch.

The data is placed on the data bus by the RAM when the OE\ pin switches to the low-level signal. The OE\ pin is controlled by the RD\ signal. During this phase, the microcontroller reads the data bus via port P0.

Theoretically, the OE\ pin can be controlled directly by the RD\ signal from the microcontroller. In this development circuit, however, the RD\ signal is combined with the PSEN\ signal via a logic circuit (IC6A, IC6B) so that the RAM can also operate as a program memory.

WR\

The address to which data is to be written is applied by the microcontroller to the address bus. The ALE signal is used to control copying of the data to the latch. The data is subsequently placed on the address bus. The RAM copies the data on the data bus when a low-level signal is present at the WE\ (**W**rite **E**nable) pin. To that end, the microcontroller activates the WR\ signal.

Reset\

The Reset\ input (Pin 10) is required to reset the microcontroller to a defined status (input active at low signal level). When the development circuit is switched on, this takes place automatically through the connection to capacitor C17 and its series resistor R1.

At the moment the power is switched on, capacitor C17 constitutes a short circuit. This is a physical property of the uncharged capacitor. As a result, the Reset\ input of the microcontroller is briefly connected to ground. The microcontroller performs a reset and reverts to the start of the monitoring program at address 0000_H. After a certain period, the capacitor is sufficiently charged for the Reset\ input to adopt the high-level status. The precise length of time depends on the rating of the capacitor and the series resistor R1.

In contrast to our example, there are also microcontrollers on which the reset is active when the signal is high.

Serial interface

The connection between the PC and the controller is provided by an RS232 serial interface.

The ports COM1 or COM2 can be used at the PC end, using either a 25-pin or a 9-pin SUB-D connector. That and the data transmission rate are set in the PC program during the programming process.

On the development circuit there is a 9-pin SUB-D socket (X2) which is connected to the microcontroller serial interface pins via the level converter MAX232 (IC5) (see description of IC5).

Programming

The object code consists of an 8-digit binary number in the range 00_H to FF_H.

In all, the instruction set of the 80C535 (IC1) extends to 111 instructions, of which

- 49 consist of a single byte,
- 45 consist of a two bytes, and
- 17 consist of a three bytes.

In the single-byte instructions, all information is contained in the "operation code".

In the 2-byte instructions, the operation code is followed by the operand which is either a constant or an 8-bit address.

Such a series of numbers is, of course, difficult to remember. That is why easily memorable abbreviations have been devised to represent the binary operation codes.

Example:
"Move register R0 to accumulator" is a single-byte instruction. It transfers the contents of register R0 to the accumulator (A) – the microcontroller's computation register. In binary notation, that instruction reads thus: $1110\ 1000_B$ (which corresponds to $E8_H$). The abbreviated form of the instruction is "MOV A,R0". This method of representing the commands by mnemonics, i.e. easily memorable abbreviations, is referred to as assembly language. These mnemonics have to be translated into operation code by a program called an assembler.

A particular feature of the mnemonics is the order of the operands. The instruction is followed first by the destination byte and then the source byte.

"MOV R0, #01" is an example of a 2-byte instruction. It instructs the CPU to transfer the numerical value 01_H to the register R0 (direct addressing).

When the CPU reads the operation code "MOV R0" it knows that the next byte read from the program memory is also part of the instruction.

Example program for "moving lights"

The example program in Figure 1 produces two moving-light sequences. A series of LEDs is connected to the output channels of each of the ports P4 and P5. These LEDs light up when the logical 1 signal is present at the corresponding pin. The time loop that determines the speed of the moving light can basically be "set" by means of the registers R2 and R3.

Program sequence

The example program starts at the address 8000_H. That is the lowest memory address in the RAM, which in this case is being addressed as a program memory. In order for the user program to be run, the monitoring program must jump to that location.

Initialisation
The initialisation process writes initial data to the registers. This means that the program starts with defined data.

The first instruction is a transfer instruction. The 8-bit constant with the value $0000\ 0001_B$ (01_H) is loaded into the 8-bit register R0. The hash symbol, #, indicates that the operand is a numerical value and not an address. Since this instruction is a 2-byte instruction, it is stored at the locations 8000_H and 8001_H.

In the next step, the constant $1000\ 0000_B$ (80_H) is copied to the register R1.

Main program
The next instruction is also a transfer instruction. The contents of register R0 are now transferred to the accumulator and from there to the port register P4. The first LED then lights up.

Now the contents of the accumulator are "rotated" one place to the left and reloaded into register R0. Thus, in the next cycle, the illuminated LED "moves" to the next position in the row of LEDs connected to P4.

The program next branches into the subroutine "time" which determines the speed of the moving-light sequence. It can be changed by altering the value of R3 or by multiple calling of the subroutine.

The sequence for the second row of LEDs is identical (register R1 and port P5) except for the fact that the value of the accumulator is "rotated" to the right. The two "moving lights" thus move in opposite directions.

The program concludes with the 3-byte instruction "JMP 8004". This tells the program to jump to the address 8004_H. In other words is repeats itself indefinitely until it is stopped by external intervention.

In contrast to this example, programs on motor-vehicle control units use an internal timer for time-based functions which does not draw on the microcontroller's CPU resources.

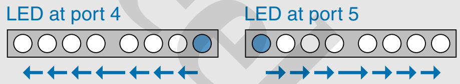
Subroutine
A subroutine is called at the address 8009H. The program then continues from the address 8100_H. This subroutine represents a double-nested time loop.

The two registers R2 and R3 are loaded with the constant FF_H. R2 is then decremented (decreased by 1). If the contents of the register are then not equal to the value 00_H, the subroutine returns to the start of the inner nested loop. Once the value of the register has reached 00_H, the subroutine moves on to its next instruction. The outer loop is then executed until the value of R3 is also 00_H.

The instruction RET (Return) then sends the subroutine back to the main program. The main program is continued with the instruction immediately following the one in which the subroutine was invoked.

The advantage of a subroutine is that it can be invoked from any number of points in the main program.

1 Example program for "moving lights"



Memory address in RAM (Hex)	Assembly language instruction (source code)	Remarks
	;Initialisation	
8000:	MOV R0,#01	;Load register R0 with the constant 01_H
8002:	MOV R1,#80	;Load register R1 with the constant 80_H
	;Main program	
	;1st moving light:	
8004:	MOV A,R0	;Load contents of register R0 into accumulator A
8005:	MOV P4,A	;Output the contents of accumulator A at port 4
8007:	RL A	;Rotate the contents of A one place to the left
		;(e.g. 000 0001_B becomes 000 0010_B)
8008:	MOV R0,A	;Load contents of accumulator A into register R0
8009:	LCALL 8100	;Invoke the subroutine at address 8100_H
	;2nd moving light:	
800C:	MOV A,R1	;Load contents of register R1 into accumulator A
800D:	MOV P5,A	;Output the contents of accumulator A at port 5
800F:	RR A	;Rotate the contents of A one place to the right
		;(e.g. 1000 0000_B becomes 0100 0000_B)
8010:	MOV R1,A	;Load contents of accumulator A into register R1
8011:	LCALL 8100	;Invoke the subroutine at address 8100_H
8014:	JMP 8004	;Jump to address 8004_H
	;Subroutine "time" (approx. 0.13 s)	
8100:	MOV R3,#FF	;Load register R3 with the constant FF_H
8102:	MOV R2,#FF	;Load register R2 with the constant FF_H
8104:	DJNZ R2,8104	;Decrement (reduce by 1) R2,
		;If R2 ≠ 00H go to address 8104H
8106:	DJNZ R3,8102	;Decrement (reduce by 1) R3,
		;If R3 ≠ 00H go to address 8102_H
8108:	RET	;Return to main program

SAE0849E

ECU development

The control unit (ECU) is the central point from which the functions of an electronic system in a motor vehicle are controlled. For that reason, extremely high demands in respect of quality and reliability are placed on ECU development.

Overview

An electronic system consists of sensors and setpoint generators, an ECU and actuators (Fig. 1).

The sensors detect the operating parameters of the electronic system (e.g. wheel speed, engine temperature, ambient pressure). The setpoint generators register the settings that the driver has specified with his/her operating controls (e.g. by means of the air-conditioner switches). The sensors and setpoint generators thus supply the input signals that are analysed and processed by the ECU.

Actuators (e.g. ignition coils, fuel injectors) convert the electrical output signals into physical variables.

The process of developing an electronic system is made up of a number of stages, in

1 Components of an electronic system

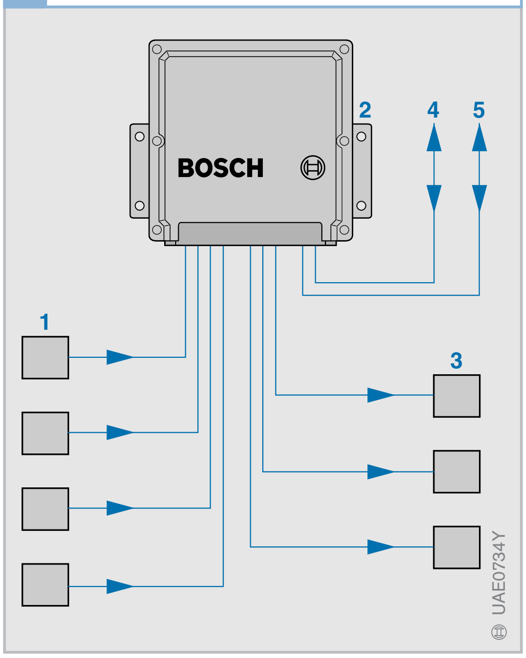


Fig. 1
1 Sensors and setpoint generators (input signals)
2 ECU
3 Actuators
4 Interface with other systems
5 Diagnosis interface

which the ECU development stage plays a decisive role. The following tasks are involved in the development of the ECU (Fig. 2):

- hardware development,
- function development,
- software development, and
- application.

Requirements

The product specifications and the development specifications document the requirements that a particular electronic system has to meet. Those two documents form the basis for the development process.

Product specifications

The product specifications define the requirements from the point of view of the vehicle manufacturer. They describe the functions that the product concerned must perform. They detail all requirements on the part of the vehicle manufacturer with regard to the products and services to be supplied. The requirements specified should be quantifiable and measurable. The product specifications thus define *what means* of performing *what task* is to be provided.

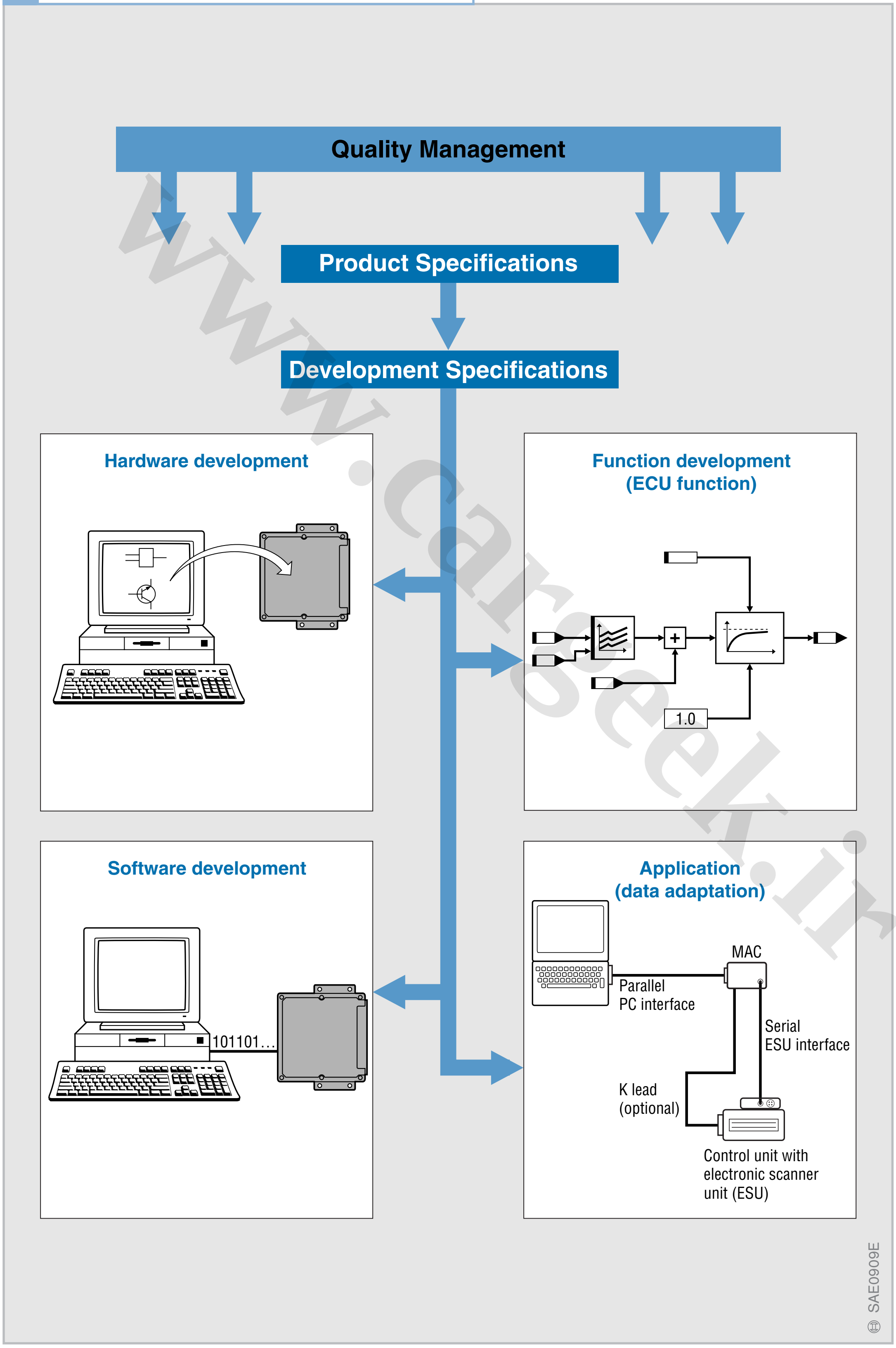
The product specifications are not revised during the course of the development process.

Development specifications

From the requirements set down in the product specifications, the ECU manufacturer draws up the development specifications. The development specifications define *how* and *by what means* the requirements are to be implemented (implementation specifications).

The development specifications are the basis for practical development of the ECU. They have to be regularly reviewed and updated in consultation with the vehicle manufacturer during the course of the development process.

2 ECU development: Interrelationship of development tasks



Hardware

Hardware is a generic term for the physical (i.e. tangible) components of a system whether mechanical (e.g. heat sink, casing) or electronic (e.g. microcontroller, memory modules, output modules). This applies to the components both individually and collectively.

The hardware of the ECU electrically processes the signals received from the sensors and passes them onto the "processor core" of the control unit for further processing. Output modules amplify the control signals so that the actuators can be operated with the required electrical power.

The task of *hardware development* (Fig. 2) is to design and produce an ECU that meets the requirements arising from the development specifications.

ECU functions

The control unit's job is to analyse the input signals and to control the actuators so that the system operates according to its intended purpose. The overall behaviour of the system can be broken down into a number of separate functions (e.g. for gasoline-engine management system: control of engine idling speed, exhaust emission levels, radiator fan, etc.). And even with today's state-of-the-art ECUs, the apparently straightforward control of the radiator fan is dependent upon a whole range of input variables. It is not sufficient merely to switch the fan on when the engine is hot, and off again when it has cooled down. Furthermore, every vehicle manufacturer has its own ideas as to exactly how this unit should operate.

Function development (Fig. 2) involves the implementation of the engine manufacturer's specifications, and the drawing up of the function descriptions which describe the ECU functions down to the very last detail. Those documents form part of the development specifications.

Software

Software is a generic term for the total of the programs and data stored in the memory of a computer-controlled system.

The central element of an ECU that performs a function in a motor vehicle is the microcontroller. It runs the program stored in the program memory. That program represents the functions of the ECU in the form of program code.

The process of *software development* (Fig. 2) converts the requirements arising from the function descriptions into a program. The machine code thus produced is entered in the program memory of the ECU. To simplify the process of writing a program, programming languages are used. The increasing complexity of electronic systems makes the use of high-level programming languages (such as the language C) absolutely essential. Software development is also assisted by simulation tools.

Data adaptation

The behaviour of an electronic system does not depend solely on the ECU program. A decisive role is also played by the data stored in the memory along with the program code. An example of such data in the case of an engine ECU would be the ignition timing map, which specifies the optimum ignition timing from the point of view of fuel consumption and emission levels for a range of engine operating conditions (engine speeds and loads/torques). Such data is engine-specific. For that reason, it has to be calculated and adapted during the process of development of the vehicle. Adaptation of the data to the engine in this way is the task of the *application* stage (Fig. 2).

Quality management

Quality assurance measures accompany the entire development process, and subsequently the production process as well. Only in that way can consistent quality of the end product be guaranteed. The quality requirements placed on safety-related systems (e. g. ABS) are particularly strict.

Quality assurance systems

All elements of a quality management system and all quality assurance measures have to be systematically planned. The various tasks, authorities and responsibilities are defined in writing in the quality management handbook. International standards such as ISO 9001 to 9004 are also adopted.

In order to regularly monitor all elements of a quality management system, quality audits are carried out. Their purpose is to assess the extent to which the requirements of the quality management system are being followed and the effectiveness with which the quality requirements and objectives are being met.

Quality assessment

On completion of specific stages in the development process, all information available up to that point about quality and reliability is subjected to a quality assessment and any necessary remedial action initiated.

FMEA

FMEA (**F**ailure **M**ode and **E**ffects **A**nalysis) is an analytical method for identifying potential weaknesses and assessing their significance. Systematic optimisation results in risk and fault cost reduction and leads to improved reliability. FMEA is suitable for analysing the types of fault occurring on system components and their effects on the system as a whole. The effect of a fault can be described by a causal chain from point of origin (e.g. sensor) to system (e.g. vehicle).

The following types of FMEA are distinguished:

- Design FMEA: assessment of the design of systems for compliance with the specifications. It also tests how the system reacts in the event of design faults.
- Process FMEA: assessment of the production process.
- System FMEA: assessment of the interaction of system components.

FMEA assessments are based on theoretical principles and practical experience.

Example: a direction indicator fails. The effects in terms of road safety are serious. The likelihood of discovery by the driver is small, however, since the indicator is not visible from inside the vehicle. As a means of making the fault obvious, the rate at which the indicators flash must be made to change if an indicator fails. The higher flashing rate is discernible both visually on the instrument cluster and audibly. As a result of this modification, the effect of the fault can be reduced.

Review

The review is an effective quality assurance tool in software development in particular. Reviewers check the compliance of the work produced with the applicable requirements and objectives.

The review can be usefully employed as a means of checking progress made even at early stages of the development process. Its aim is to identify and eliminate any faults at as early a point as possible.

Hardware development

The complexity of electronic systems has seen a continuing increase in the past years. This tendency will continue in the future, and it will only be possible to master such developments by applying highly integrated circuits. The call for small dimensions for all system components places further exacting demands on the hardware development process.

Efficient and cost-effective hardware development is now only possible by using standard modules that are produced in large numbers.

Project starting point

A schematic diagram sets out all the functions that are to be performed by the ECU being developed. From that point, the following aspects can be clarified:

- definition of hardware required,
- cost estimate for the hardware,
- extent of development work required, and
- cost of tools.

Hardware design prototypes

Once the project is underway, hardware design prototypes are produced and subjected to quality tests. These design prototypes may be one of four categories representing successive stages along the road to the final production ECU. Each prototype category is based on its predecessor and is designed for a particular purpose in each case.

'A' prototype

The 'A' prototype is derived from an existing or modified ECU or a development circuit board. Its range of functions is limited. Its technical function is largely in place but the 'A' prototype is not suitable for continuous testing. It is a *function prototype* that is used for initial trials and to confirm the basic viability of the design.

'B' prototype

The 'B' prototype includes all circuit components. It is a *trial prototype* that is used to test out the full range of functions and the technical requirements. At this stage it is ready for continuous testing in prototype vehicles.

The connection sizes and ECU dimensions are as required for final production. It may be, however, that not all of the vehicle manufacturer's specifications are satisfied at this stage, perhaps because different materials have been used, for instance.

'C' prototype

The 'C' prototype is the *approval prototype* on which the vehicle manufacturer's tests for "technical approval" are performed. This version of the ECU must reliably conform to all specifications. On successful approval of the product, the development process is complete. As far as possible, 'C' prototypes are produced using full-production tools and production methods as close as possible to full series production.

'D' prototype

The 'D' prototype is the *pilot-series prototype* which also carries the full-production identification plate showing the version number. 'D' prototypes are fitted in pilot-series vehicles for large-scale vehicle trials. This version of the ECU is produced using volume-production methods and is fitted and tested under volume-production conditions. It is the version with which the reliability of production is verified.

Preparations

The preparations for the 'B' prototype start from the beginning of the project and involve
- definition of connector pin assignment,
- definition of the casing design,
- ordering and developing new circuit modules (function groups),
- producing the circuit diagram,
- defining the components (only approved components may be included, or inclusion must be subject to approval).

When selecting the circuit modules (e.g. knock-sensor analyser circuits integrated in an IC), developers check whether existing circuits – with modifications if necessary – can be used. If not, new modules have to be developed.

Circuit diagram and components list

A CAD (Computer-Aided Design) system is used to create the circuit diagram (Fig. 1 a) and the list of all components used. The components list also details the following information for each component:
- its size,
- its pin assignment,
- its casing design, and
- the supplier and terms of supply.

Layout

A circuit-board layout (Fig. 1 b) is required for production of the circuit board. It shows the positions of the conductor tracks and the connector pins.

The layout is produced on a CAD system. The process starts by taking the circuit-diagram data and converting it. The component connection schedule (list of connections between components) then produced shows how the various components are connected to each other. From this component connection schedule and the CAD data for the components (size and connector pin assignment) the layout can then be produced.

1 Hardware development process

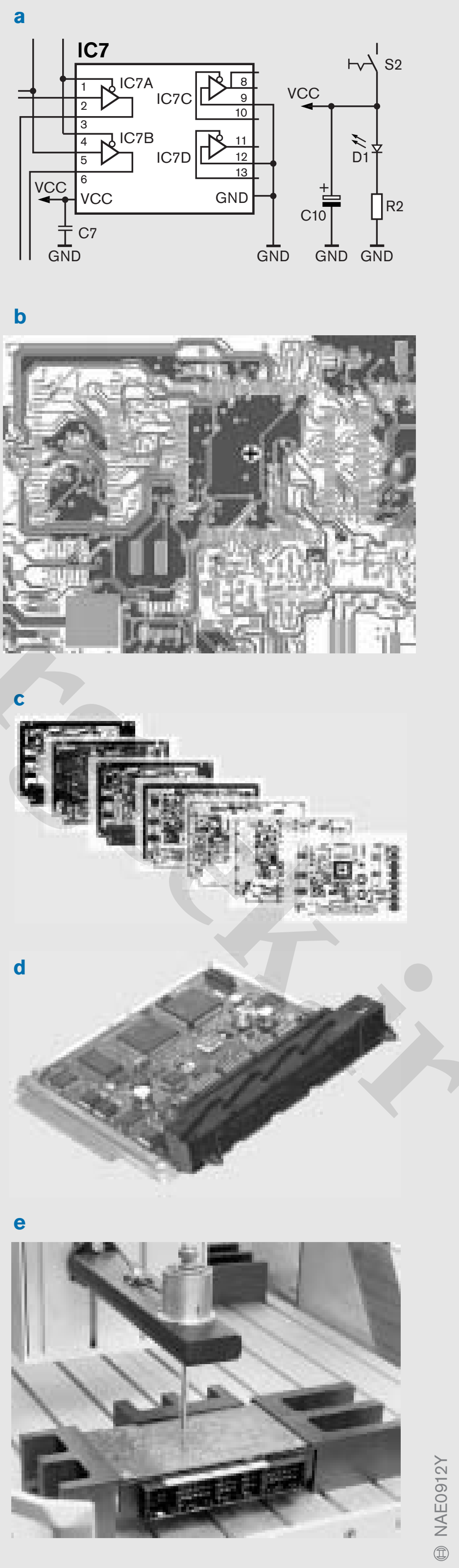


Fig. 1
a Circuit diagram
b Layout
c Circuit-board production
d Prototype construction
e Testing

There are specific criteria that have to be taken into account in the creation of the layout. In connection with the positioning of components these can include the following considerations:
- the power loss from specific components (possibilities for heat dissipation),
- EMC (electromagnetic compatibility) effects,
- convenient positioning of components in relation to connectors,
- observance of restriction zones (component size),
- ease of fitting of components by automated production machinery,
- accessibility of testing points, and
- space requirements for testing adaptors.

Circuit board

The layout data can be used to produce films for manufacture of the printed-circuit board (Fig. 1 c). The films are used to photographically expose the blanks (which are covered with a light-sensitive coating) and then develop and etch them. The individual layers of multilayer printed-circuit boards are placed on top of one another and hardened.

Finally, the component print, the solder resist and a carbon lacquer are applied to the circuit board.

Prototype construction

The finished circuit board has to have the components mounted on it (Fig. 1 d). In the case of prototype products, this is part of the process of prototype manufacture. Because of the miniaturisation of components and the high degree of integration on the printed circuit, even prototypes have to have their components fitted by machine. The machine is controlled by the CAD layout data.

Following fitting of the components, their connections are soldered. There are two alternative soldering methods:
- wave soldering and
- reflow soldering.

Testing the finished circuit board

Electrical testing

Once all components have been fitted and all connections soldered, the circuit board must be tested. To this end, electrical testing sequences are defined which run on a computer. These automatic tests check that all components are fitted and that the circuit functions properly.

Thermographics

Thermographic images of the printed circuit show the heat generated by the components during operation (Fig. 2). Different temperature ranges appear on the film as different colours. In this way, components that are too hot can be identified. The information obtained is incorporated in the list of modifications between the 'B' and 'C' prototypes. Changes to the circuit layout (e.g. heat-dissipation through-contacting) can reduce the amount of heat generated.

2 Thermographic image of a printed-circuit board

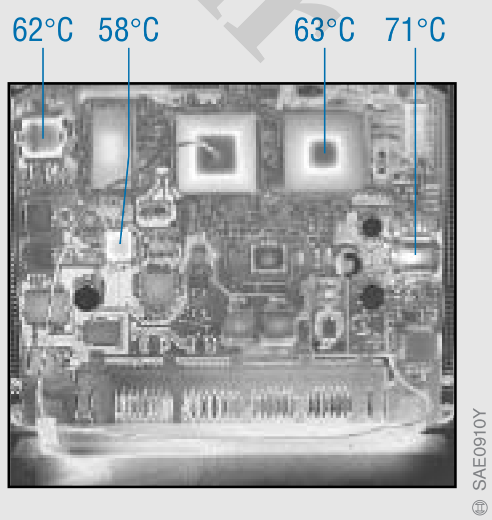


Fig. 2
Engine control unit
Power supply voltage
U = 14 V
Idling speed
n = 1000 rpm

Electromagnetic testing

The electromagnetic fields created on the circuit board can be scanned with a magnetic field detector (Fig. 1 e). The readings are analysed on a PC. Different field strengths show up as different colours. If necessary, modifications to the circuit layout have to be made and additional components fitted that reduce the magnetic field or make the ECU immune to interference.

These tests are performed on the 'B' prototype so that the necessary modifications can be incorporated in the 'C' prototype.

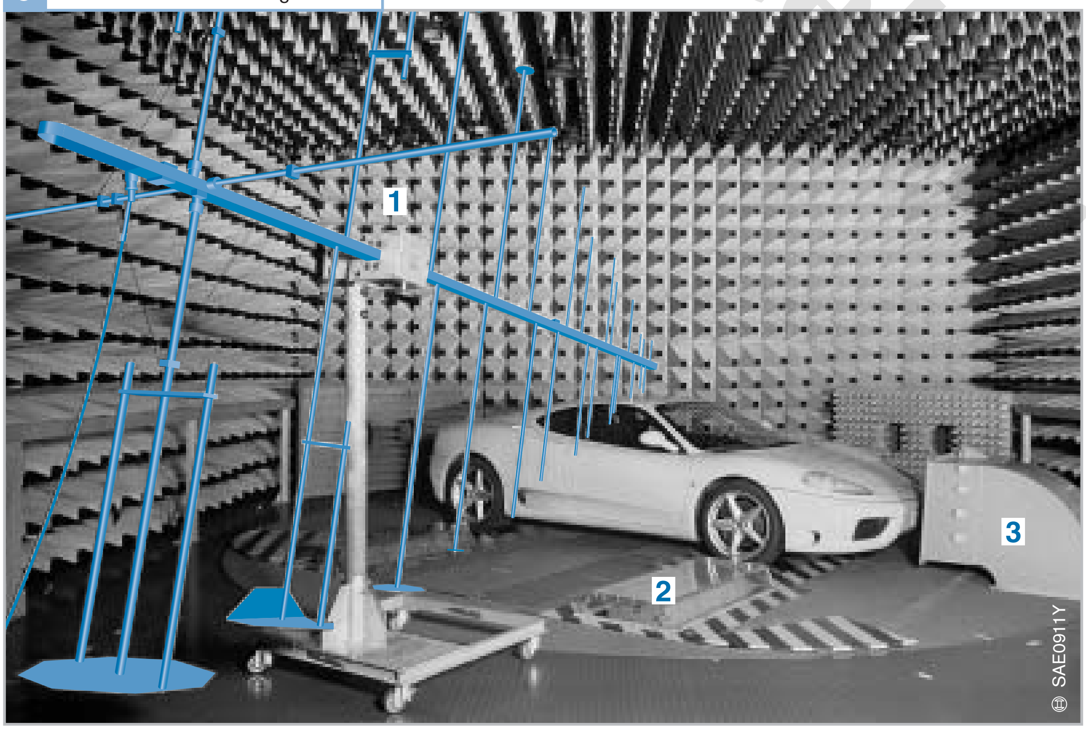
EMC tests

Readings taken in the EMC testing cell or testing chamber (Fig. 3) test the ECU response to external and internal sources of electromagnetic interference. Measurements are taken both with the ECU fitted in the vehicle (in-situ tests) and in the laboratory (e. g. stripline method).

The difficulty with the *in-situ tests* is that they cannot be carried out until development of the vehicle and all its electronic systems is at a very advanced stage. If the EMC characteristics are found to be unsatisfactory at this stage, the scope for modifications is then severely limited. For this reason, *laboratory testing* is extremely important because it permits potential problems to be identified on the hardware prototypes at an early stage in the process.

The EMC readings are taken at a variety of frequencies and with a range of electric field strengths. The immunity of the output signals (e.g. ignition signals, fuel injection signals) to external interference is examined as well as the level of interference generated by the unit itself.

3 Vehicle in the EMC testing chamber



Manufacture of semiconductor components and circuits

The manufacture of microelectronic components and circuits demands ultra-high precision as well as extremely high standards of cleanliness in the various stages of production. Even the minutest speck of dirt can irreparably damage a circuit.

Semiconductor components

Most semiconductor components are made of silicon. The raw material is SiO_2 (quartz). It occurs naturally in large quantities as rock crystal or silica sand. It is used to make monocrystalline cylinders or rods of extremely pure silicon with a diameter of between approx. 50 mm (2″) and 300 mm (12″). The most common diameters are 150 mm (6″) and 200 mm (8″).

During the production of the crystals, they are given a precise conductivity specified by the component manufacturer (basic doping). The cylindrical crystal is sliced into discs roughly 0.3 ... 0.7 mm thick which are then smoothed and polished. These discs are referred to as wafers.

The process of producing the semiconductor components starts with these silicon wafers (Fig. 1). The value-added chain from the raw material to the end product shows an extremely high rate of value growth: silicon in the form of silica sand costs about 1 euro per kg, but by the time it has become a finished microprocessor it can be worth up to 1 million euro per kg.

A large number of identical circuits (chips) is created on each wafer and those circuits are then separated by mechanical means. Before separation, however, the chips are tested to make sure they match the electrical specifications. Any unusable chips are marked and then discarded once separated. Only the usable chips are assembled, i.e. cemented, bonded, encased, sealed and passed on for final testing (Figs. 1 and 3).

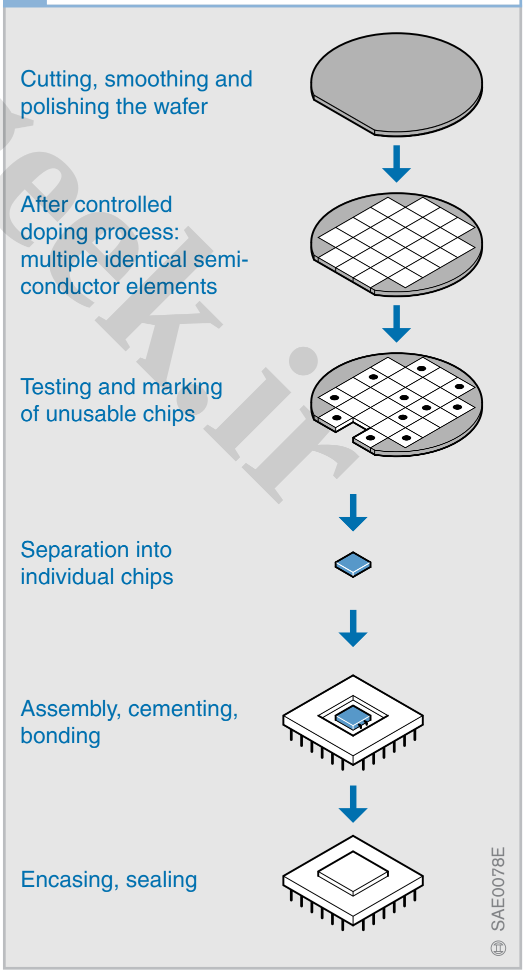
Doping process

Doping involves introducing impurities with a known electrical effect into the semiconductor crystal at specific points and in precisely defined concentration levels using reproducible methods. Doping techniques are the fundamental processes of semiconductor technology. The guiding parameters are concentration profile, penetration depth, surface concentration and planar parallelity of the doping-material fronts.

Doping during crystal growth

While the crystal is being grown, phosphorus can be added to the silicon melt as a doping material. As the monocrystal forms, the phosphorus atoms become fixed in the

1 Stages in the manufacture of semiconductor components



silicon crystal, giving it n-type conducting properties.

Manufacturers of semiconductor components buy their raw material in the form of wafers that have been pre-doped according to precisely defined specifications.

Doping by diffusion

At high temperatures, the doping material can diffuse into the silicon crystal. The effect is achieved by creating a specific concentration level of the impurity atoms at the surface of the wafer. The difference in concentration causes the doping atoms to diffuse into the silicon wafer.

The process involves placing between 50 and 200 wafers in a kiln together and exposing them to boron or phosphorus compound vapours at temperatures of around 1000 °C. Boron produces p-type regions while the effect of phosphorus is to create n-type regions. Surface concentration levels, temperature, and time, are the parameters that determine the penetration depth of the doping material.

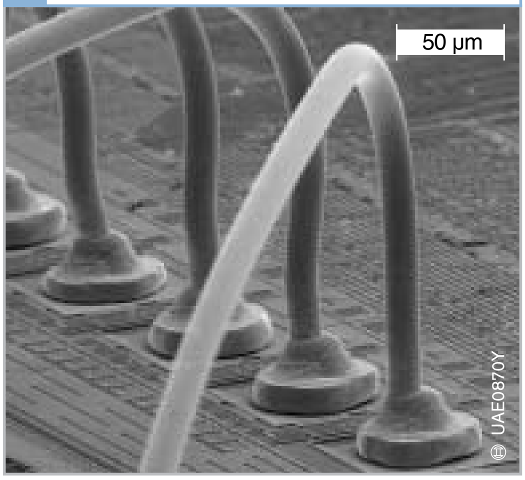
Doping by ion implantation

In this process, the atoms of a gaseous doping material are first of all ionised in a vacuum and then accelerated by a high voltage (up to 300 kV) so as to be “catapulted” into the semiconductor. This allows particularly precise control of concentration level and localisation of the implanted material. Embedding of the implanted atoms and restoration of the crystal lattice requires subsequent thermal treatment.

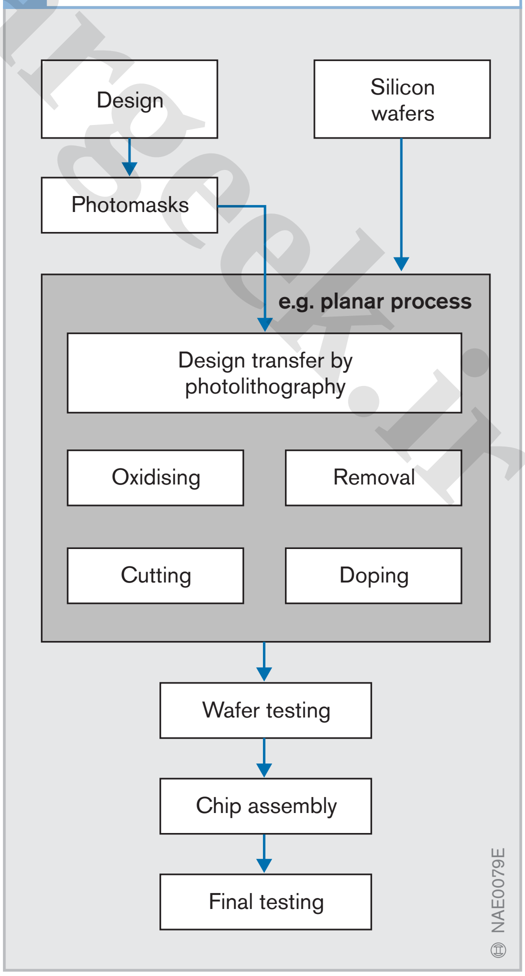
Epitaxy

This doping method creates a monocrystalline doped semiconductor layer a few micrometres thick on a monocrystalline substrate. If gaseous silicon tetrachloride and hydrogen are passed over silicon wafers heated to around 1200 °C in a quartz tube, the vapour breaks down and silicon is deposited at a rate of about 1 µm/min, form-

2 Microstructure of a microchip with bonded connections



3 Process of manufacturing semiconductor components



ing a monocrystalline layer. If a defined quantity of doping impurity is added to the gas flow, an epitaxial layer is created whose electrical conductivity and conduction type differs fundamentally and abruptly from that of the substrate.

Photolithography and the planar process

In the *photolithography* process, the pattern of the component design is transferred to the wafer by using metal screens. When the design has been created, the data for producing the screen is saved onto a storage medium (e.g. magnetic tape or CD). It is then used to control a photographic exposure device that transfers the design pattern to photographic plates. This pattern is subsequently reduced by optical means to the actual size for the application and copied onto metal screens on which it is repeated as many times as will fit onto the size wafer being used.

As this method can be used for structures that are many times smaller than the wavelength of the light used, it will continue to be used in the future. The size of the smallest achievable structures depends on the wavelength of the light source. Using UV lasers, photolithographic methods have already produced structures as small as 0.08 µm in the laboratory (by comparison, a human hair has a diameter of 40 ... 60 µm).

Even smaller structures can be produced using other methods such as X-ray or electron-beam lithography. However, those methods are substantially more expensive as they can only "expose" *one* IC at a time on the wafer. For that reason, they are only used in special cases.

Silicon wafers are easily oxidised by oxygen or water vapour. The oxide layer thus created prevents penetration by the doping atoms in the doping process. In the *planar process*, holes are created in the oxide layer so that when the wafer is doped, localised areas of n-type and p-type material are created.

The wafer is coated with a special lacquer and then exposed to light while covered with a metal screen. Following developing, the areas of the lacquer coating previously covered by the screen and the oxide layer below it can be removed by etching. The position, size and shape of the holes thus created match precisely the specified design. In the subsequent doping process in the diffusion furnace (Fig. 4) or by ion implantation, impurities which have a known electrical effect such as boron or phosphorus pass through those holes in the oxide layer into the silicon, thereby creating n-type or p-type regions in the desired locations. Afterwards, the oxide layer is removed again and the wafer is ready for the next stage in the manufacturing process.

The photolithographic process and the doping process are repeated as many times as required to produce the desired number of layers of different conductivity on the semiconductor component. In the case of complex integrated circuits, this may involve as many as 20 or more separate manufacturing stages. In order to provide the electrical connections between the resulting functional elements, the wafers are coated with aluminum or copper and then the metal conductor track patterns are formed. This process too may involve the formation of multiple layers of metal, one above the other.

Completion of the wafer production process is followed by electrical testing (preliminary testing) of the individual chips on the wafer. Chips that do not meet the specifications are marked with coloured dots. The wafers are then cut up into individual chips using a diamond cutting tool (Fig. 5). The functional chips are subsequently placed in metal or plastic casings and fitted with external connections (Fig. 6). After being hermetically sealed or encased in plastic, they go through a final testing stage.

4 Wafers about to be placed in a diffusion furnace

6 Assembled microchip with connections prior to sealing

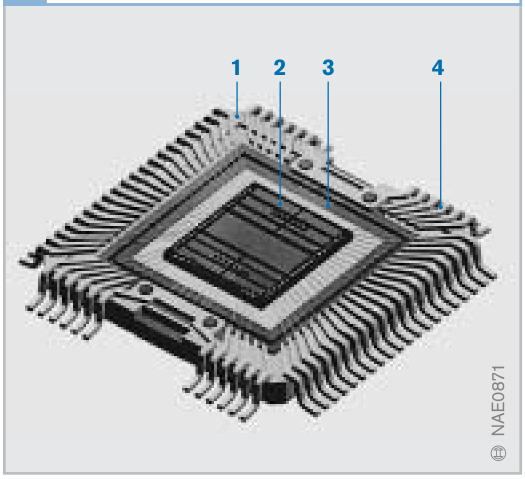
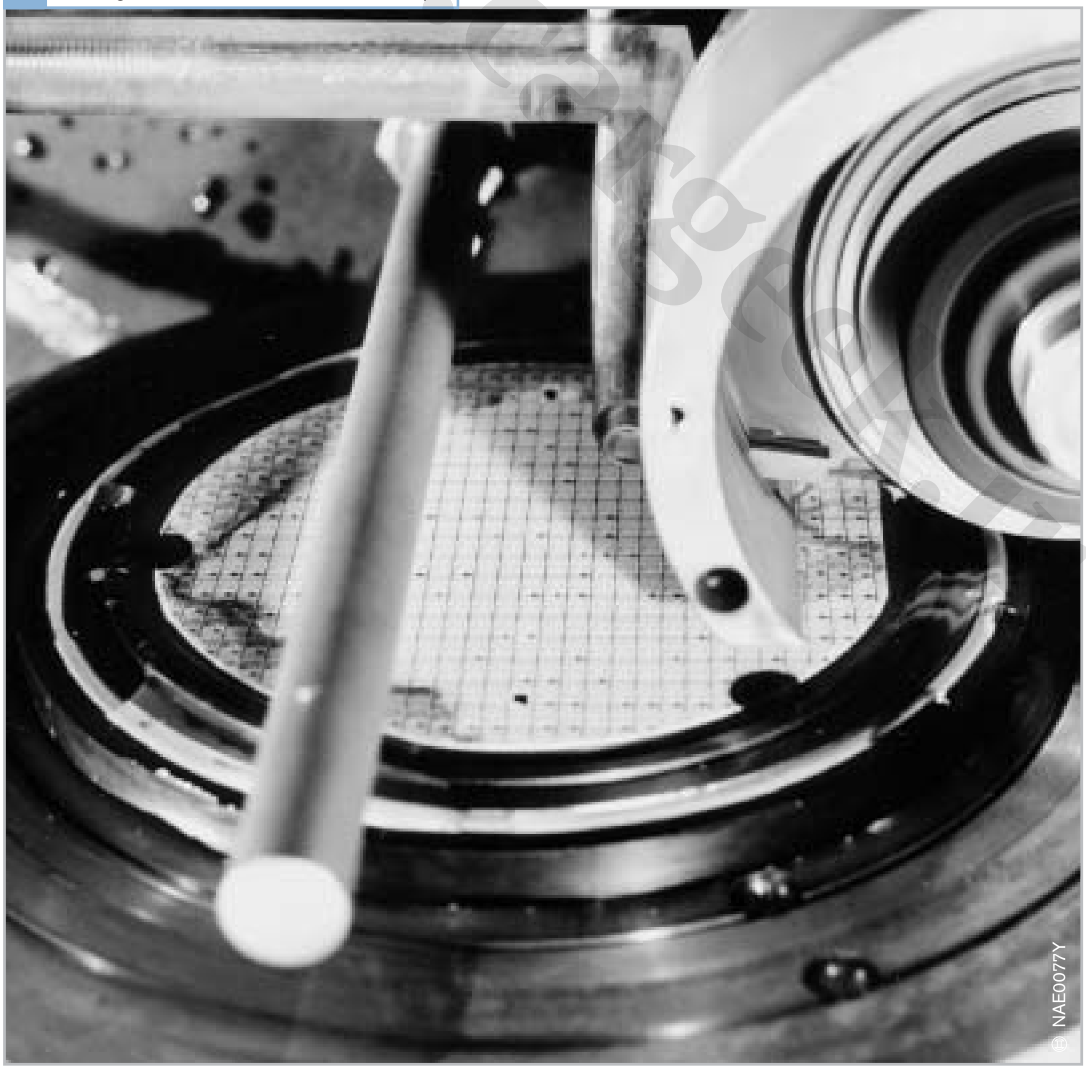


Fig. 6
1 Base
2 Chip
3 Bonded connection wire
4 Connection pin

5 Cutting the silicon wafer into individual chips



Micromechanics

Micromechanics is the name given to the production of mechanical components from semiconductors (generally made of silicon) utilising semiconductor technologies. In addition to its semiconducting properties, the mechanical characteristics of silicon are made use of as well. This makes it possible to create sensors with the most minute of dimensions. The following techniques are used:

Bulk micromechanics

The silicon wafer material is removed up to the full thickness of the wafer by anisotropic (alkaline) etching with or without electrochemical resist. The material is etched away from the reverse side of the silicon layer (Fig. 1, item 2) in those areas where it is not protected by the etching screen (1). This method can be used to create very small diaphragms (a) with typical thicknesses of between 5 and 50 µm, holes (b) and bars and ridges (c), e.g. for pressure or acceleration sensors.

Surface micromechanics

The base material is a silicon wafer on the surface of which miniature mechanical structures are formed (Fig. 2). First of all, a "sacrificial layer" is applied and shaped (A) using semiconductor production processes (e.g. etching). This is deposition-coated with a layer of polysilicon approximately 10 µm thick (B) which is then shaped by vertical etching using a lacquer screen (C). In the final stage, the sacrificial oxide layer below the polysilicon layer is removed using gaseous hydrogen fluoride (D). In this way, structures such as flexible electrodes (Fig. 3) for acceleration sensors can be created.

Wafer bonding

Anodic and seal glass bonding are methods used to join wafers together by the action of electricity and heat or heat and pressure in order, for instance, to hermetically seal a reference vacuum or to protect sensitive structures by placing a cap over them.

1 Structures producible using bulk micromechanics

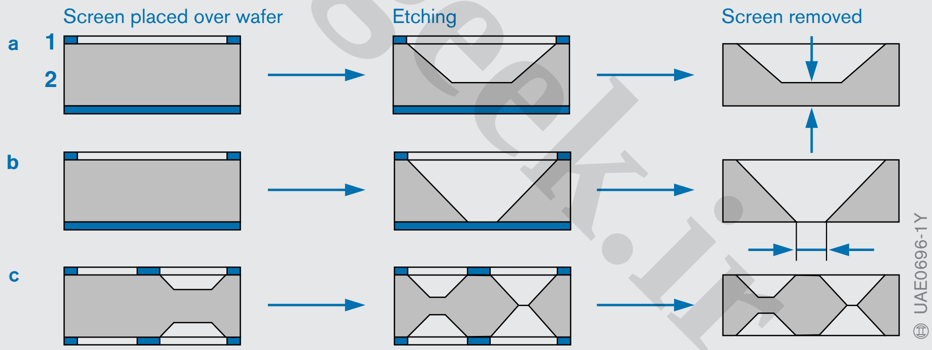


Fig. 1
- **a** Production of a diaphragm
- **b** Production of a hole
- **c** Production of bars and ridges

1 Etching screen
2 Silicon

2 Surface micromechanics (stages of process)

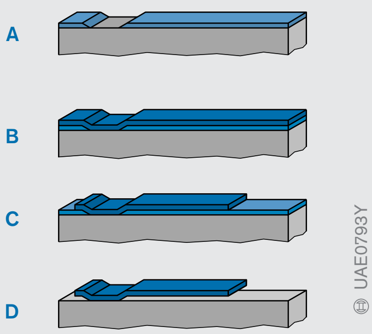


Fig. 2
- **A** Deposition and shaping of the sacrificial layer
- **B** Deposition of the polysilicon
- **C** Shaping the polysilicon
- **D** Removing the sacrificial layer

3 Surface micromechanics (structure details)

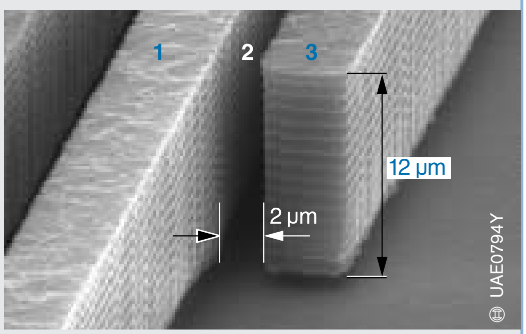


Fig. 3
1 Fixed electrode
2 Gap
3 Oscillating electrode

Conventional printed-circuit boards

The printed-circuit board has become an electronic component in its own right. It has to have precisely defined electrical and mechanical properties. In a motor vehicle, for example it may be required to withstand temperatures ranging from –40 ... +145 °C. The standards required in respect of EMC (electromagnetic compatibility), maximum current capacity and complexity are becoming more and more demanding. And at the same time, the circuit boards are continually expected to become smaller and cheaper – in spite of shorter product life cycles.

The basic material – the baseboard – is glass fibre. It can be rigid or semi-flexible. The conductor tracks on the surface are made from a thin layer of copper 12 ... 70 µm thick (copper base). To protect it against corrosion, the surface of the copper is protected by a coating of tin alloy, gold or an organic surface sealant, depending on the type of application.

Designs

Depending on the complexity of the circuit, printed-circuit boards may be made up of a number of layers (Fig. 1). Circuit boards used in motor vehicles consist of between two and eight layers.

Single-sided circuit board

The pattern of conductor tracks and solder eyes is only on one side of the baseboard (a).

Double-sided, non-interconnected circuit board

There are patterns of conductor tracks on both sides of the circuit board. However, those two circuits are not connected to each other (b).

Double-sided, interconnected circuit board

The conductor track patterns on each side are interconnected by copper linings inside the holes (c).

Multilayer circuit board

In addition to the two outer layers (in this case the 1st and 4th layers), there are additional conductor layers on the inside of the baseboard (internal layers). Those conductor layers may be electrically connected with one another. To that end, the relevant conductor layers are connected to the copper linings inside the holes (d).

1 Types of printed circuit board

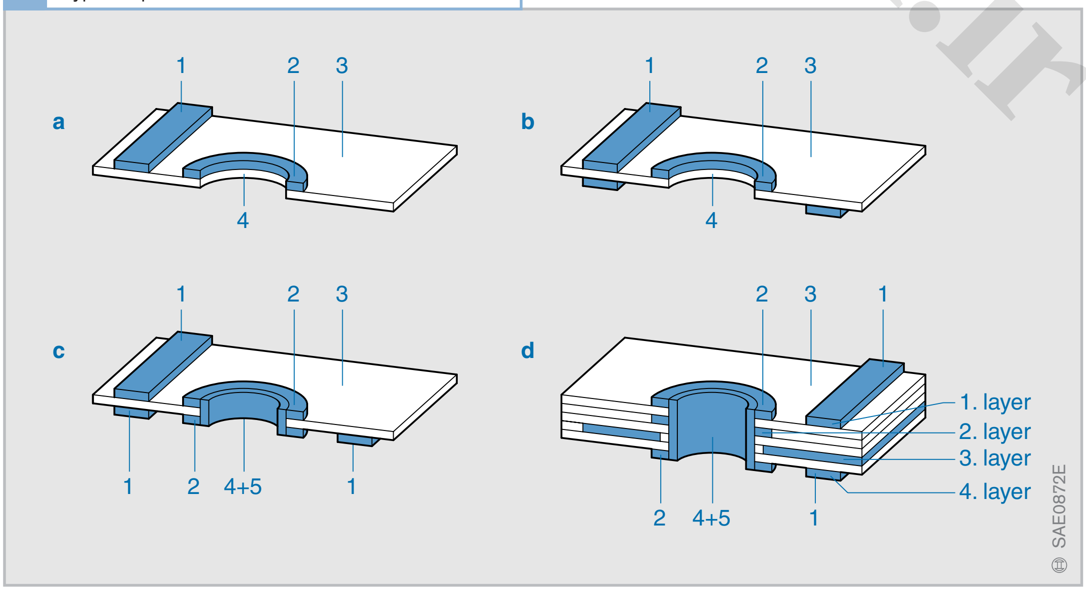


Fig. 1
Schematic diagram: conductor tracks are shown as raised surfaces

a One-sided
b Double-sided, non-interconnected
c Double-sided, interconnected
d Multilayer

1 Conductor track
2 Solder eye
3 Baseboard
4 Hole
5 Copper lining

Production process

Pattern plating and panel plating have become the established methods of manufacturing printed-circuit boards (Fig. 2).

2 Manufacturing process for printed-circuit boards

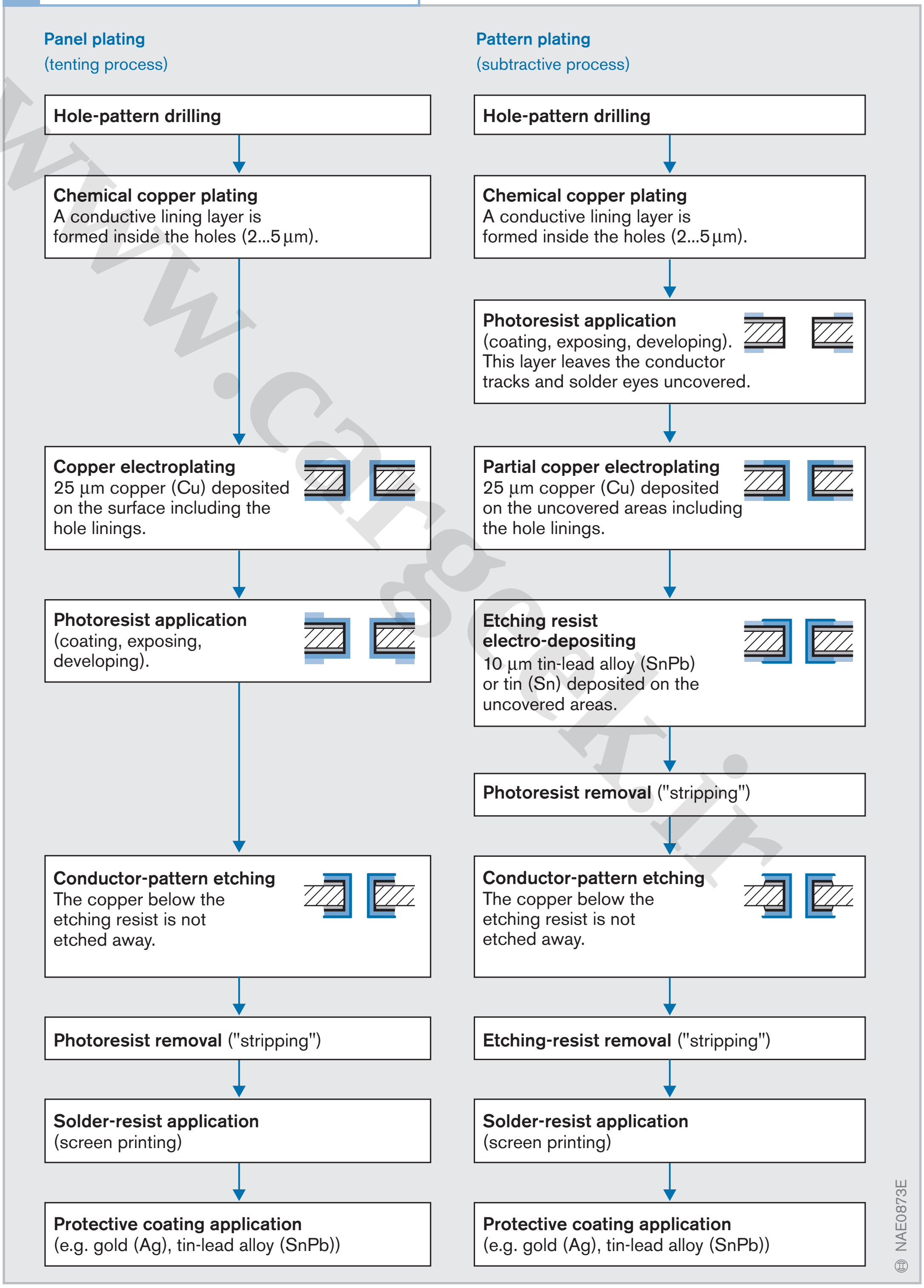


Fig. 2
The base material is coated on both sides with a copper film

Subsequent processing

Once the circuit board has been produced, it has to be fitted with the circuit components. These subsequent processing stages also demand the highest quality standards so as to ensure reliable operation of the finished product in the vehicle. There are two methods of mounting the components on the board – through-fitting and surface mounting.

Through-fitting method

With the through-fitting method, the component connections are passed through the holes in the circuit board and then soldered (Fig. 3a).

Surface-mounting method

The surface-mounting method SMT uses special electronic components whose connections lie flat against the surface of the circuit board (Fig. 3b). Such components are referred to as SMDs (surface **m**ounted **d**evices). Another advantage – in addition to the greater component density achievable – is that the components can be fitted to the circuit board fully automatically. The surface mounting method is therefore becoming more and more widely adopted. The component fitting machines used can achieve work rates of more than 60,000 components an hour, increasing productivity as a result.

Production sequence

There are various production systems that are used in the subsequent processing of circuit boards and they are classified according to the spatial arrangement of the production equipment and the work areas. The categories distinguished are job-shop, pool, flow-shop and flow production.

The process described below is an example of flow production (Fig. 4). According to the standard DIN 33 415, flow production involves "a process of operations organised according to the flow principle in a rigid sequence, aligned to a particular spatial arrangement and tied to a cyclic timing pattern". Furthermore, the flow principle involves "arrangement of the individual work stations according to the order in which the operations are performed". This method is also referred to as line production.

Production of the electronic circuits of an ECU involves the following stages:

Basic materials

The basic materials consist of printed-circuit boards that have not yet been fitted with circuit components. Usually, one or more circuit boards are arranged on a "panel" of standardised dimensions. After being fitted with components, the circuit boards are cut out of the panel.

3 Component fitting methods for printed-circuit boards

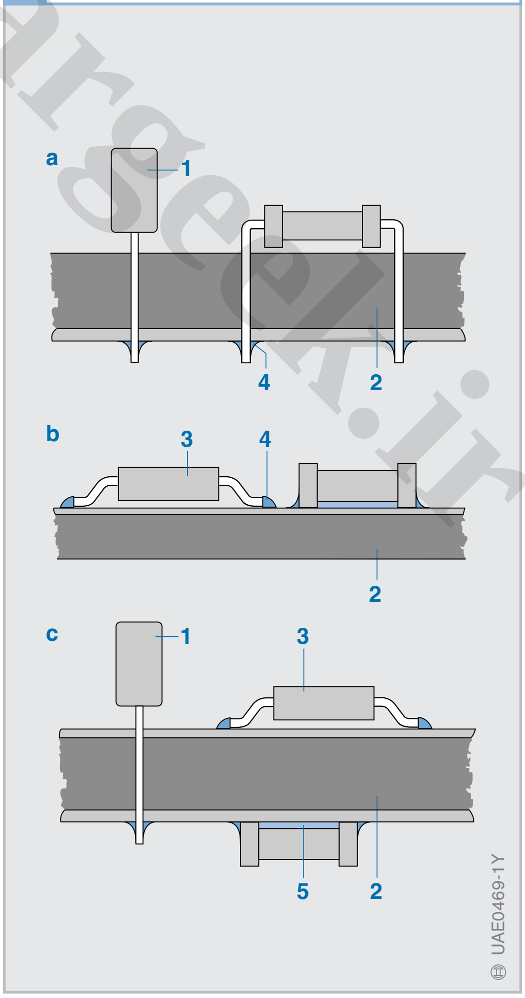


Fig. 3
a Through-fitting method
b Surface-mounting method
c Mixed mounting

1 Wired component
2 Circuit board
3 SMD
4 Soldering point
5 Cement

Solder-paste application
The first stage is the application of solder paste for the SMDs using a screen printing process. The solder paste is a mixture of metal powder, flux and other organic additives. It is pressed through the spaces in the printing screen onto the panel with the aid of a "scraper".

SMD fitting (reflow side)
An automatic component-insertion machine presses the SMDs into the solder paste applied to the panel.

Reflow oven
The panel passes through a reflow oven on a conveyor belt. The heat of the oven melts the solder paste on the panel. This electrically and mechanically connects the components to the pcb.

Position checking
The camera of a vision system examines the soldered connections (position checking system). Depending on the findings of the examination, the panel may be diverted to a repair station or automatically moved on to the next stage of production.

Turning the circuit board
In this stage, the circuit board is rotated through 180 degrees so that the "underside" is facing upwards and can be worked on. The way in which components are fitted to the underside depends on the nature of the second soldering stage.
If the *reflow/reflow method* is used, the process described above is repeated.

If the alternative *reflow/wave method* is used, since the soldering method applied differs to that from the top side, the SMDs are fitted differently and a cementing station is required.

Cementing station (underside)
At the cementing station, cement spots are first of all applied in those positions on the "underside" where SMDs are to be placed. This is done using pipettes or a screen, similar to the procedure for applying the solder paste. The cement holds the SMDs in position until they are soldered.

SMD fitting (underside)
At succeeding stations along the assembly line, the circuit components are placed on the cement spots. On this side of the circuit board, only components that are suitable for attachment by wave soldering can be fitted.

Cement hardening
In order to prevent the components falling off during the soldering process, the cement is hardened in an oven.

Component checking
A visual component checking system checks that all components are present and that they are in the correct positions on the underside of the circuit board. Non-compliant products are indicated and diverted to a repair station.

Turning the circuit board
The circuit board is once again rotated through 180 degrees, returning it to its original orientation.

Wired-component fitting
This stage involves fitting components that have leads attached (e.g. large coils and connectors) into ready-made holes (solder eyes) in the panel using the through-fitting method. They are subsequently soldered on the reverse side at the same time as the SMDs during the wave-soldering stage. The wired-component fitting stage completes the component fitting operations.

Wave soldering
The wave soldering stage solders the connections of all SMDs and "wired components" on the underside. The wave soldering process involves the following three stages:
1. Flux is applied to the underside of the panel.

2. The panel then passes through a pre-heating zone so that the components are not damaged by the sudden increase in temperature in the subsequent stage.
3. The panel passes over a wave of liquid solder created by a jet. As it does so, the solder is deposited on the soldering points (pads) on the panel. The solder resist on the circuit board prevents the solder attaching itself in the wrong places.

Position checking
Another camera checks the finished soldered connections. If faults are found, the panel is passed to a repair station.

In-circuit testing (ICT)
The purpose of the in-circuit test is to check that the electrical circuit is working properly. The testing equipment is connected via a testing adaptor and checks the components for correct function and electrical readings.

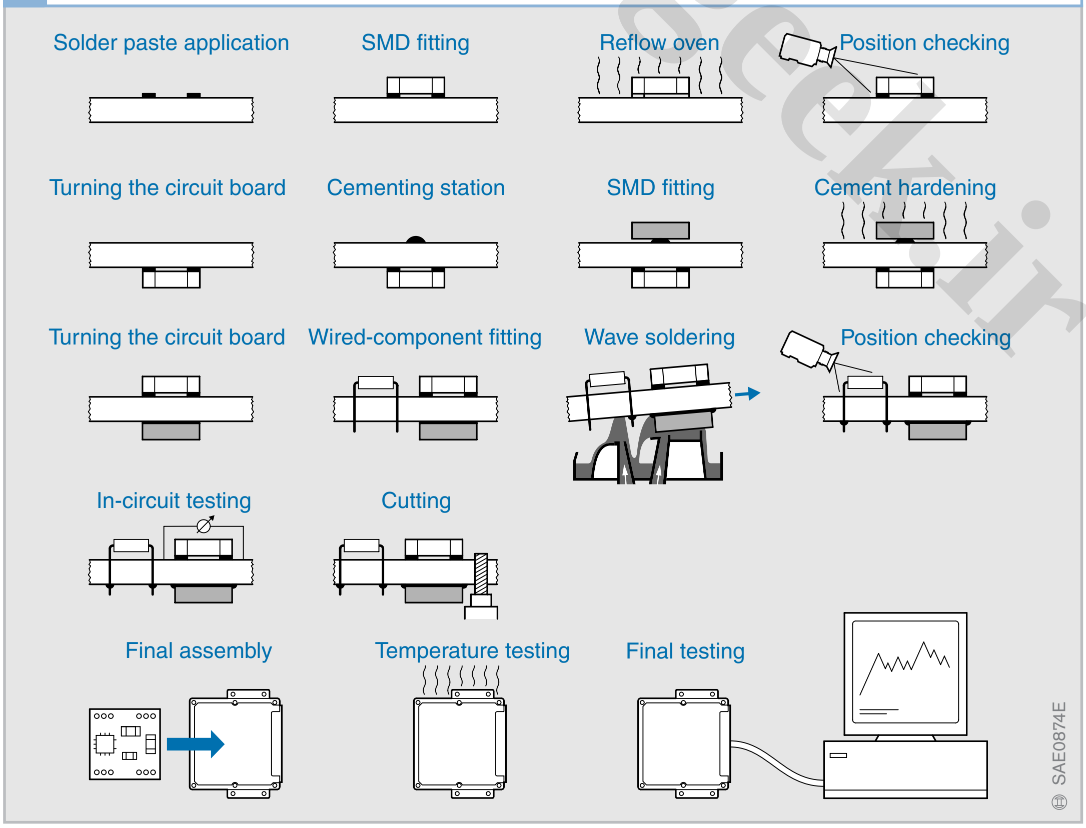
Cutting
The individual circuit boards are now cut out of the panel using computer-controlled machines.

Final assembly
In the final-assembly stage, the circuit board is fitted into the ECU casing.

Temperature testing
The finished ECU is subjected to very high temperatures in order to test its ability to withstand extreme conditions. The tests simulate operation of the ECU in real conditions. In this way, component or soldering faults can be detected.

Final testing
Before the ECU leaves the production line, it is subjected to a final test to make sure it performs the functions required of it in actual use.

4 Stages in pcb manufacture using a flow production process and reflow/wave soldering (example)



Film and hybrid circuits

Film circuits

In film integrated circuits, passive circuit elements (usually conductor tracks, insulation and resistors, but also capacitors and inductors) are applied to a base (substrate) in layers. The advantages of these circuits are:

- fine structures (up to approx. 10 µm) with high element density, and
- good high-frequency characteristics.

Those advantages are counterbalanced by relatively high production costs.

The thickness of the films originally gave rise to the terms "thin-film circuit" and "thick-film circuit". Classification is now based on differences in the method of production.

Thin-film circuits

With thin-film circuits, the films are applied to glass or ceramic substrates using vacuum-coating processes.

Thick-film circuits

For thick-film circuits, the preferred method of manufacture involves application of the films to ceramic substrates by screen printing processes followed by firing.

Ceramic multilayer substrates

Ceramic multilayer substrates are made up of unfired ceramic foils onto which conductor tracks are applied by screen printing methods. Several such foils are then combined to form a multilayer laminate which is fired at 850 ... 1600 °C so that it becomes a solid ceramic body with integral conductor channels. A typical hybrid substrate consists of four or five layers. Particularly high wiring densities can be achieved with LTCC (low-temperature cofired ceramic) line-line substrates.

To make the electrical connections between layers, holes are punched in the individual films and filled with metal paste. Those holes are referred to as so-called "vias". By using suitable material systems, resistors and capacitors can also be integrated. The wiring densities of these circuits is considerably greater than with thick-film circuits.

Hybrid circuits

Hybrid circuits are integrated film circuits with additional discrete components such as capacitors and integrated semiconductor circuits (ICs) that are attached by soldering or cementing. The use of unpacked semiconductor chips, to which the connections are made by bonding, or SMDs makes high component densities possible. Extremely small hybrid ECUs (microhybrids) can be made by using ceramic multilayer substrates. The advantages of these circuits are:

- high permissible installation temperatures due to their good heat dissipation properties,
- compactness of design combined with good shock resistance, and
- good resistance to mediums.

Hybrid circuits are therefore particularly suited to use in telecommunications systems and for automotive applications where they can be found in ABS, traction control, ESP, transmission control and engine-management ECUs (engine-mounted).

Figure 1 shows the basic stages in the production of a hybrid-circuit substrate. The tapes are punched with holes for the vias for each separate wiring layer and the holes are then filled with silver paste (Fig. 2a). Screen-printing stations print the conductor tracks onto the film. The various layers are aligned with each other, laminated and then fired at 890 °C. A specially controlled sintering process keeps the ceramic firing to within tolerance limits of roughly 0.03 %. This is important for the packing density. The circuit's resistors are printed on the reverse side of the substrate and fired (Fig. 2b).

For bonding on the top side, the surfaces are finished using a plating process adapted to the LTCC. The spacing of the microcontroller contacts (bond land grid on the substrate) ranges from 450 to 260 µm. The component connections are bonded using 32-µm gold wire and 200-µm aluminum wire.

Alongside the electrical vias, there are also thermal vias with a diameter of 300 µm for optimum cooling of ICs with high power losses. The effective thermal conductivity of the substrate can thus be increased from approx. 3 W/mK to 20 W/mK.
All components are cemented with conductive cement. There are two methods used for final assembly of the finished hybrid.

Method 1: the finished hybrid is cemented to the steel panel of the casing using heat-conducting cement and connected to the glass feed-through for the connector using bonded 200-µm aluminum wire. The casing is then hermetically sealed.

Method 2: the finished hybrid is cemented to the aluminum casing using heat-conducting cement and connected to the plastic-encased connector pins using bonded gold or 300-µm aluminum wire. Before the cover is cemented in place, a gel is applied to protect the circuit.

1 Production sequence of a microhybrid substrate

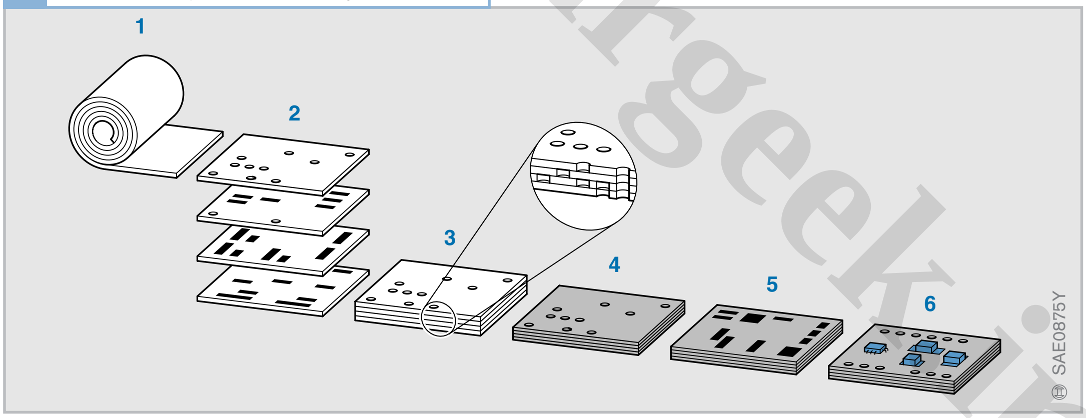


Fig. 1
1 Unfired glass-ceramic film
2 Punching of holes, filling with conductor paste and printing of conductor tracks
3 Aligning and stacking (laminating)
4 Sintering
5 Printing of resistors (reverse side), firing and plating of the bonding pads (top side)
6 Fitting of components and wire bonding

2 Example of a hybrid circuit (sections)

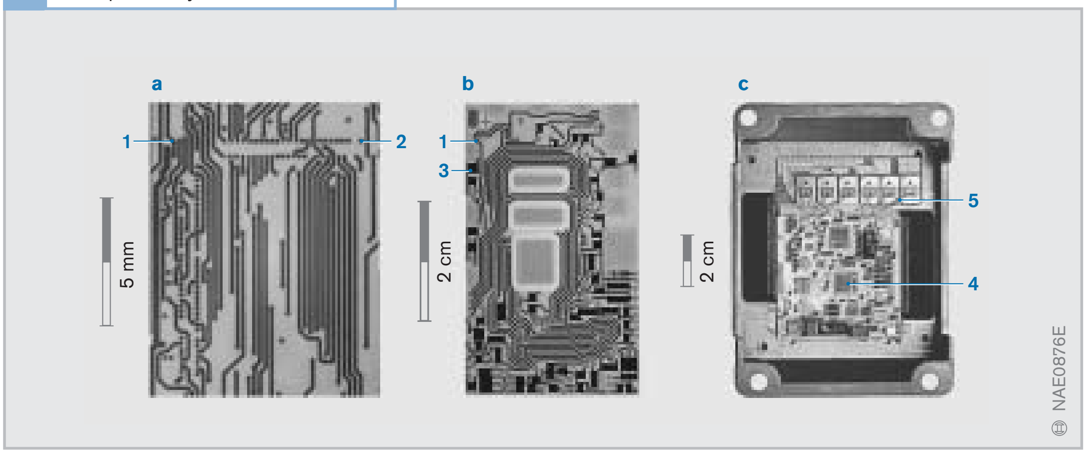


Fig. 2
a Inner layer
b Reverse side with resistors
c Top side inside ECU

1 Conductor track
2 Via
3 Resistor
4 Microcontroller
5 Bonded wire

Glossary and tables for automotive microelectronics

This glossary provides a quick guide to the most important automotive microelectronics terms.

An arrow pointing to a term printed in *italics* (e.g. → *semiconductor*) indicates that this italicised term is also explained in this glossary.

A

Analog circuit → *Electronic circuit* with which analog signals are processed – as opposed to → *digital circuit*. Analog signals are infinitely variable within a specific range. Examples: battery voltage, speedometer display.

Avalanche breakdown: Sudden increase in the number of → *charge carriers* resulting from the release of bound electrons from the crystal lattice of a → *semiconductor* in the → *space-charge zone* caused by a high-strength electric field (→ *avalanche effect*).

Avalanche effect: see → *avalanche breakdown*

B

Binary code: Information encoded in the form of binary numbers and therefore consisting of a sequence of the digits 0 and 1.

Bipolar transistor: → *Transistor* with three regions of differing conductivity (→ *p-type semiconductor* and → *n-type semiconductor*). It is constructed either as a p-n-p or n-p-n transistor. The terminals are called emitter (E), base (B) and collector (C).

Bit: Contraction of **bi**nary digi**t**. A single digit in → *binary code*

Breakdown voltage: reverse voltage at a → *p-n junction* above which a small increase in voltage brings about a large increase in the → *reverse current*.

Bus: Connection linking the elements of a → *digital circuit* and consisting of a single lead (serial bus) or multiple leads (parallel bus). A bus can connect up a large number of separate elements with different functions but with electrically identical interfaces. A parallel bus transmits all bits of a data "word" simultaneously (one lead for each bit). A serial bus transmits the bits sequentially (one after the other).

Byte: Coherent group ("word") of eight → *bits*.

C

Capacitance: Property of a capacitor which indicates how much electrical charge – and therefore electrical energy – it can store when a specific voltage is applied.

CCD: Abbreviation for → *charge-coupled device*.

Central processing unit (**CPU**)**:** Main operating unit of a → *microcontroller*.

Charge-coupled device (**CCD**)**:** Silicon image sensor in which incident light striking a transparent electrode releases → *charge carriers* in proportion to the light intensity and exposure time. The charge carriers collect in a Si-SiO_2 boundary layer ("potential well").

Charge carrier: Carrier of electric charge, usually of a single elementary unit of charge. Free electrons and anions are examples of negative charge carriers. → *Lattice defects* (holes), cations and protons are examples of positive charge carriers.

Chip: Basic building block of microelectronic circuits. Base material is a monocrystalline block of semiconductor material, primarily silicon.

Circuit diagram: Diagram illustrating the manner in which the individual electronic components of a → *circuit* are electrically connected to each other.

Circuit, electrical: Combination of electrical or electronic components forming a self-contained functional unit. Circuits are also referred to, both collectively and individually, as → *hardware*.

Circuit layout: Arrangement of conductor tracks and components on a printed circuit.

Clock-pulse generator: Circuit that generates a timing pulse of a fixed frequency for synchronising all operations in a → *microcomputer*.

Closed-loop control: Method of control characterised by a closed control loop. The actual value of a variable is continually compared with its specified value. As soon as a difference is detected between the two, a readjustment is made to the setting of the actuator (output variable).

Communication: Exchange of data between electronic components.

Conductivity: Reciprocal of resistance ($1/\Omega = A/V$).

Conductor: Material with high electrical → *conductivity* due to its large number of free → *charge carriers*. Metals are the best conductors.

CPU: Abbreviation for → *central processing unit*.

D

Degree of integration: Number of functional elements per → *chip*.

Diffusion: Migration of atoms, molecules or → *charge carriers* from a region of high temperature or concentration to a region of low temperature or concentration. This effect is utilised in the → *doping* of silicon wafers with impurities.

Diffusion voltage: Voltage created by the → *diffusion* of → *free electrons* and → *holes* at the → *p-n junction* of a → *semiconductor*.

Digital circuit: Electronic circuit with which digital signals are processed – as opposed to → *analog circuit*. Digital signals are in the form of a series of discrete pulses rather than infinitely variable characteristics. Example: digital clock.

Diode: Component with a → *p-n junction* and two connections – one to the p-type region and one to the n-type region.

Discrete active component: Active component with fewer than 100 individual functions on a single → *chip*.

Doping: Controlled introduction of impurities with known electrical effects into high-purity semiconductor crystals with the aim of producing → *free electrons* to create an → *n-type semiconductor* or → *holes* to create a → *p-type semiconductor*.

DRAM: Abbreviation for → *dynamic random-access memory*.

Dynamic random-access memory (DRAM): Volatile short-term memory (DRAM = Dynamic → *RAM*) which allows direct access to any memory location and which does not retain its contents when the power supply is switched off.

E

EEPROM: Abbreviation for Electrically Erasable Programmable Read-Only Memory, i.e. a → *read-only memory* that can be electrically erased and reprogrammed. EEPROM is also called E^2PROM.

Epitaxy: Method of → *doping* silicon wafers in which a gaseous mixture of silicon tetrachloride, hydrogen and doping impurities is passed over wafers heated to around 1200 °C. An epitaxial layer is created, the electrical conductivity of which differs significantly from that of the → *substrate.*

EPROM: Abbreviation for Erasable Programmable Read-Only Memory, i.e. a → *read-only memory* that can be erased using UV light and then reprogrammed.

F

FET: Abbreviation for → *field-effect transistor.*

Field-effect transistor (**FET**)**:** → *Transistor* in which the current flowing through a conductive channel is controlled by an electric field that is created by a voltage applied to the control electrode.

Flash EPROM: Programmable → *read-only memory.* Often simply called a "flash".

Free electron: Electron that is unattached to an atom and can therefore move freely.

H

Hardware: Generic term for the physical (tangible) components of an electronic system. It includes the mechanical as well as the electronic components and the combination of them.

Hole: See → *lattice defect.*

Hybrid circuit board: → *Integrated circuit* in which the conductor tracks and the passive elements are printed directly onto a ceramic hybrid substrate. Hybrid circuit boards are particularly suited to uses where there are high operating temperatures.

I

Inductance: Characteristic of a coil which indicates the amount of magnetic energy the coil can store for a given current passing through it.

Insulator: Non-metallic material that, because of its lack of free → *charge carriers* has only infinitesimally small electrical conductivity. Also call non-conductor.

Integrated circuit (IC)**:** Electronic circuit consisting of inseparable interconnected semiconductor components on a single, monolithic → *chip.*

Interface: Defined electronic connection between electronic components or modules.

Intrinsic conduction: Conduction of electricity by a → *semiconductor* made possible by presence of free → *charge carriers* (electron-hole pairs) created by the action of heat or light on undoped silicon which makes it intrinsically conductive.

Ion implantation: High-precision procedure for → *doping* silicon → *wafers* using doping impurities in gaseous form that are first ionised by a high voltage (up to 300 kV), accelerated and then "catapulted" into the semiconductor.

L

Laser diode: → *Light-emitting diode* that emits laser light, i.e. a parallel beam of monochromatic (of a single wavelength) and coherent (the waves are in phase) light.

Lattice defect: Mobile electron gap and crystal lattice vacancy in → *semiconductors* the atoms of which are held together by valency bonds. It carries a positive charge. Lattice defects are also called → *holes*.

LED: Abbreviation for → *light-emitting diode*.

Light-emitting diode (LED): → *Diode* that emits light when a voltage is applied in forward direction. An LED produces the reverse effect to a photodiode.

M

Magnetoresistor: Magnetically controllable → *semiconductor resistor* the resistance of which increases as the magnetic flux density *B* increases.

MCM: Abbreviation for → *multi-chip module*.

Memory: Electronic component capable of storing data.

Microcomputer: Electronic module consisting of the following components: a → *microprocessor* as the CPU (**c**entral **p**rocessing **u**nit), input and output modules, program memory, data memory, clock-pulse generator and power supply system.

Microcontroller: Electronic module in which the → *CPU*, → *RAM*, input/output units, various peripheral modules and, in some cases, the → *ROM* or → *EPROM* are integrated on a single → *chip*. The microcontroller is thus the central module of a microcomputer-controlled circuit.

Microelectronics: Branch of technology that deals with the conception, design, technology, manufacture and use of highly miniaturised electronic circuits.

Microprocessor: Central processing unit of a computer in the form of an integrated circuit on a → *chip*.

Monitoring program: A → *program* by means of which the user can control the actual → *user program*.

Monolithic integration: Highly integrated arrangement of all components of a circuit, including the interconnecting conductors, on a single silicon crystal (→ *chip*).

Monolithic integrated circuit (IC): Active component with ≥100 individual functions on a single → *chip*.

MOSFET: Abbreviation for **m**etal-**o**xide **s**emiconductor → *field-effect transistor*. The terminals of a MOSFET are called gate (G), source (S) and drain (D).

Multi-chip module: Component made up of several interconnected → *chips*.

N

n-type doping: Controlled introduction of impurity atoms with 5 outer electrons (e.g. phosphorus) into a silicon crystal lattice. Those foreign atoms introduce → *free electrons* because only four outer electrons are required to bind each atom within the silicon crystal lattice.

n-type semiconductor: → *Doped* → *semiconductor* that due to the presence of → *free electrons* is negatively conductive (n-type).

NTC thermistor: → *Semiconductor resistor* the resistance of which decreases as temperature increases (NTC stands for **n**egative **t**emperature **c**oefficient).

O

Open-loop control: Method of control in which the actuators are controlled (output variables) on the basis of input variables, specified data (e.g. stored data maps) and algorithms (computing procedures) in the ECU. The effect of control operations is not checked (open control loop).

P

p-n junction: Boundary between a → *p-type region* and → *n-type region* of the same semiconductor crystal.

p-type doping: Controlled introduction of impurity atoms with 3 outer electrons (e.g. boron) into a silicon crystal lattice. These foreign atoms create electron vacancies (→ *lattice defects* or → *holes*) because they are one outer electron short of the four required to properly bind each atom within the silicon crystal lattice.

p-type semiconductor: → *Doped* → *semiconductor* that due to the presence of → *lattice defects* (→ *holes*) is positively conductive (p-type).

PC: Abbreviation for **p**ersonal **c**omputer. Small, powerful computer for home or office use.

Photodiode: → *Diode* with reverse voltage in which incident light increases the → *reverse current* (photoelectric current) in proportion to the intensity of the light.

Photolithography: Process used to transfer a component design to a → *wafer* using photomasks.

Photoresistor: → *Semiconductor resistor* the resistance of which decreases when exposed to light.

Phototransistor → *Transistor* in which the conductivity between the emitter (E) and the collector (C) can be altered by light striking the base (B). Another type of phototransistor can be used like a → *light-emitting diode*. As this type of phototransistor can be switched on and off very quickly, it is ideally suited to use in the manufacture of flat screens.

Photovoltaic cell: Component in which, in contrast to the → *photodiode*, there is no external voltage applied to the → *p-n junction*. When exposed to light, free → *charge carriers* are produced which are separated at the → *p-n junction* and produce a photo-electric voltage.

Planar process: Method of manufacturing → *monolithic integrated circuits*. A silicon dioxide masking layer that repels doping impurities is removed in precisely defined areas using photolithographic methods. → *Doping* is then applied to produce localised p-type and n-type regions in the areas uncovered by the masking layer.

Program: Logically arranged sequence of instructions that are processed by the → *microcontroller*. Programs are also referred to, both collectively and individually, as → *software*.

PROM: Abbreviation for **p**rogrammable **r**ead-**o**nly **m**emory.

PTC thermistor: → *Semiconductor resistor* the resistance of which decreases as temperature decreases (PTC stands for **p**ositive **t**emperature **c**oefficient).

R

RAM: Abbreviation for **r**andom-**a**ccess **m**emory.

Read-only memory (ROM): Semiconductor memory that holds the program for the → *microcontroller*. It is called a read-only memory because data can be read from it but not written into it. The data in the ROM is "indelibly imprinted" in it during the process of manufacturing the semiconductor chip. A PROM (Programmable **ROM**) can be programmed once by the user. An EPROM (Erasable **PROM**) can be completely erased using UV light and reprogrammed. A flash EPROM can have its entire contents or individual sections of it electrically erased. An EEPROM (Electrically Erasable **PROM**) can have every memory cell individually overwritten, which is why this type of "permanent" memory is primarily used for non-volatile data memories.

Rectifier diode: → *Diode* which allows current to pass in one direction (positive terminal on p-type region) but not in the other (positive terminal on n-type region) and thus operates as a "flow control valve".

Register: A memory area integrated in the → *CPU* or a peripheral module that allows fast reading and writing of data.

Reverse current: Small residual current arising from minority charge carriers in the extended → *space-charge zone* at a → *p-n junction* under external voltage.

ROM: Abbreviation for **r**ead-**o**nly **m**emory.

S

Schottky diode: → *Diode* with a → *p-n junction* in the form of a metal-semiconductor junction. The electrons transfer more easily from the n-type silicon to the metal film than in the opposite direction, thereby creating an electron-depleted boundary layer known as the Schottky barrier.

Semiconductor: A material whose electrical conductivity is less than that of a → *conductor* (metal) but greater than that of an → *insulator* (non-conductor). Its conducting properties can be altered by controlled addition of impurities (→ *doping*). Silicon and germanium are the most important semiconductors.

Semiconductor memory: Component for storing digital information by means of the status alternatives conducting/not conducting or charged/not charged. They are made from → *semiconductors.*

Semiconductor resistor: Component whose resistance, in contrast with ohmic resistors, is voltage, current and temperature-dependent.

Sensor: Component for detecting physical or chemical variables.

Software: The → *program(s)* and data required for the operation of a data-processing system.

Space-charge zone: Poorly conductive region depleted of mobile → *charge carriers* at a → *p-n junction* that is created by → *diffusion* of free → *charge carriers* in either direction across the junction without the presence of an external voltage.

SRAM: Abbreviation for **s**tatic **r**andom-**a**ccess **m**emory.

Substrate: Base material in the manufacture of electronic and microelectronic modules (e.g. silicon crystal, circuit board, ceramic).

T

Thyristor: Component with at least three → *p-n junctions* that can be switched from a reverse-bias condition to a forward-bias condition (or vice versa).

Transistor: Component in which a large current between the emitter (E) and the collector (C) can be controlled by a small base current (B).

Tunnel diode: → *Diode* with a highly doped p-n junction. It produces the "tunnel effect" in forward direction.

U

User program: Microcontroller → *program* that can be written and modified by the user using special → *software* (high-level language or assembly language).

V

Variable-capacitance diode: → *Diode* with a → *space-charge zone* at the → *p-n junction* which functions as a capacitor; the dielectric is formed by the semiconductor material depleted of → *charge carriers.*

Varistor: → *Semiconductor resistor* the resistance of which decreases as voltage increases, regardless of the polarity of the voltage.

W

Wafer: A thin slice of a monocrystalline → *semiconductor.* It is the basic material for the manufacture of semiconductor components.

Z

Zener diode: → *Diode* in which an increase in the reverse voltage above a certain level brings about a sudden rise in current due to → *Zener breakdown* and/or → *avalanche breakdown.*

Zener breakdown: Electric breakdown of the non-conducting layer of a → *diode* operated in reverse direction due to the abrupt increase in the → *intrinsic conductivity* as a result of a high-strength electric field at the → *p-n junction.*

Tables

Table 1

1 Decimal multiples and sub-multiples of SI units

Prefix	Symbol	Factor	Factor name Scientific[1])	UK colloquial
atto	a	10^{-18}	quintillionth	
femto	f	10^{-15}	quadrillionth	
pico	p	10^{-12}	trillionth	
nano	n	10^{-9}	billionth	
micro	µ	10^{-6}	millionth	
milli	m	10^{-3}	thousandth	
centi	c	10^{-2}	hundredth	
deci	d	10^{-1}	tenth	
deda	da	10^{1}	ten	
hecto	h	10^{2}	hundred	
kilo	k	10^{3}	thousand	
mega	M	10^{6}	million	
giga	G	10^{9}	billion	thousand million
tera	T	10^{12}	trillion	billion
peta	P	10^{15}	quadrillion	
exa	E	10^{18}	quintillion	trillion

[1]) The factor names are those now generally used for scientific purposes both in Britain and the USA. The colloquial terms for the numbers concerned may differ in Britain (see last column).

Table 2

2 Electrical quantities

Quantity	Symbol	Unit	Formula
Electric current	I	A (ampere)	
Electric potential	U	V (volt)	1 V= 1 W/A
Electric conductivity	G	S (siemens)	1 S= 1 A/V= 1/Ω
Electric resistance	R	Ω (ohm)	1 Ω= 1/S= 1 A/V
Electric charge	Q	C (coulomb)	1 C= 1 A·S= $\frac{1}{3600}$ A·h
Electric capacitance	C	F (farad)	1 F= 1 C/V
Electric flux density, displacement	D	C/m²	
Electric field strength	E	V/m	

Where does the term "electronics" come from?

This term really originates from the ancient Greeks. They used the word electron for "amber" whose forces of attraction for wool and similar materials had already been described by Thales von Milet 2,500 years ago.

The term "electronics" originates directly from the word "electrons". The electrons, and therefore electronics as such, are extremely fast due to their very small mass and their electrical charge.

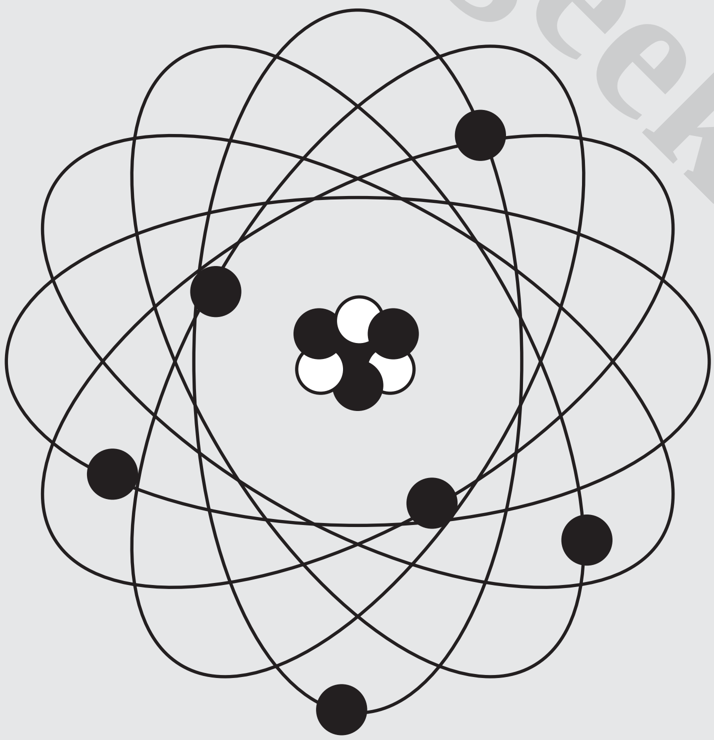
The mass of an electron has as little effect on a gram of any given substance as a 5 gram weight has on the total mass of our earth.

Incidentally, the word "electronics" is a product of the 20th century. There is no evidence available as to when the word was used for the first time. Sir John Ambrose Fleming, one of the inventors of the electron tube could have used it around 1902.

The first "Electronic Engineer" though goes back to the 19th century. He was listed in the 1888 Edition of a form of "Who's Who", published during the reign of Queen Victoria. The official title was "Kelly's Handbook of Titled, Landed and Official Classes". The Electronic Engineer is to be found under the heading "Royal Warrant Holders", that is the list of persons who had been awarded a Royal Warrant.

And what was this Electronic Engineer's job? He was responsible for the correct function and cleanliness of the gas lamps at court. And why did he have such a splendid title? Because he knew that "Electrons" in ancient Greece stood for glitter, shine, and sparkle.

Source:
"Basic Electronic Terms" ("Grundbegriffe der Elektronik") – Bosch publication (reprint from the "Bosch-Zünder" (Bosch Company Newspaper)).



LAE0047Y

Index of technical terms

An arrow pointing to a term printed in italics (e.g. → *conductivity*) indicates a synonym or related term.

Technical Terms

Abbreviations

A

ABS: Antilock braking system
ADC: Analog-digital converter
ALE: Address Latch Enable
ALU: Arithmetic logic unit
ASIC: Application-specific integrated circuit
ASR: "Antriebsschlupfregelung", i.e. traction control
ASSP: Application-specific standard product

B

BKS: "Bauteilkontrollsystem", i.e. component checking system (→ *printed-circuit-board manufacture*, production process)

C

CAD: Computer-aided design
CAE: Computer-aided engineering
CAM: Computer-aided manufacturing
CCD: Charge-coupled device
CISC: Complex instruction-set computer
CPU: Central processing unit (of a → *microcontroller*)

D

DAC: Digital-analog converter
DMA: Direct-memory access

E

EEPROM: Electrically erasable programmable read-only memory
EMC: Electromagnetic compatibility
EPROM: Erasable programmable read-only memory
ESP: Electronic stability program

F

FET: Field-effect transistor
Flash-EPROM: Flash erasable programmable read-only memory

H

HDL: Hardware Description Language

I

I/O ports: Input/output ports
IC: Integrated circuit
ICT: In-circuit test (→ *printed-circuit-board manufacture*, production process)

L

LED: Light-emitting diode
LKS: "Lagekontrollsystem", i.e. position checking system (→ *printed-circuit-board manufacture*, production process)
LSI: Large-scale integration
LTCC: Low-temperature cofired ceramic

M

MOS: Metal-oxide semiconductor
MSI: Medium-scale integration

P

PC: Personal computer
PROM: Programmable read-only memory

R

RAM: Random-access memory
RISC: Reduced instruction-set computer
ROM: Read-only memory

S

SiO_2: Silicon dioxide
SMD: Surface-mounted device
SMT: Surface-mounting technology
SoC: System on a chip
SSI: Small-scale integration

T

TCS: Traction control system

U

UV light: Ultraviolet light

V

VHDL: Visual Hardware Description Language
VLSI: Very-large-scale integration

µC: Microcontroller