KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

Politechnika Wrocławska

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

ROZWÓJ POTENCJAŁU I OFERTY DYDAKTYCZNEJ POLITECHNIKI WROCŁAWSKIEJ

Wrocław University of Technology

Automotive Engineering

Radosław Wróbel

TRENDS IN VEHICLE ELECTRONICS

Wrocław 2011

Projekt współfinansowany ze środków Unii Europejskiej w ramach
Europejskiego Funduszu Społecznego

Wrocław University of Technology

Automotive Engineering

Radosław Wróbel

TRENDS IN VEHICLE ELECTRONICS

Wrocław 2011

Reviewer: Tomasz Długosz

ISBN 978-83-62098-12-5

Published by PRINTPAP Łódź, www.printpap.pl

Contents

1. Introduction

Nowadays car vehicle is a piece of mechatronical art. Electronics system in the car constitutes a large part of the vehicle s price. And it will increase. This script contains a summary of main E-E (electronic - electric) systems which occurs in cars.

The script consists of 11 chapters (with Introduction). Second chapter Automotive Electronic Architecture, EOBD system contains information about routing of information on sensors' buses and characterization of European On Board Diagnosis System. Third chapter includes overview of the fuel controls for diesel and spark ignition engines. In fourth chapter Basic of microcontrollers the history of electronics chips, digital signal processing, processor architecture in light of vehicle’s development is described. Fifth chapter IP networks presents a view on the future. The day, in which all car vehicles will have a specific IP (Internet Protocol) address and access to constant on line communication with the services, is approaching. Next four chapters (Controlled Area Network, Local Interconnect Network, MOST and FlexRay) include topologies, communications type and physical implementation of popular buses which are used in a modern cars. Next chapter “Vehicles Sensoring” is a full overview of sensors and their principles of working. Last chapter Lighting presents light systems from halogen, krypton to LED bulbs with explanations of a physical phenomena that is connected with the electromagnetic waves.

I believe that this book will be a valuable source of information for students (but maybe not only) to achieve new knowledge and inspire them to follow the news in trends of vehicle electronics.

Special thanks to my language corrector Joanna Staszeczka for her hard work and reviewer Tomasz Długosz for valuables comments.

2. Automotive Electronic Architecture, EOBD system

A vehicle is a mechatronic product. It means that its price depends on the amount and quality of electronic systems that the vehicle has (sensors, conditioning systems, microprocessors, bus's drivers etc.). There are three main groups of diagnostic systems in modern vehicles:

1. Function systems, for regulating and steering tasks,
2. Theft protection systems,
3. Systems overseeing engine's performance factors (EOBD, OBD).

Consortiums are not copying sensors of the same parameters if they belong to different systems.

The sensor processes examines physical or chemical properties into electrical signal. Every sensor is described by two formulas input and output:

$$\Phi = f(E, Y) \tag{1}$$

$$E = g(\Phi, Y) \tag{2}$$

where: Φ – examined property, E – electrical exit signal, Y – interference.

Electrical exit sensor signal is proportional to examined property. The influence of interference is noticeable and is an element of a function (that we can't ignore).

Vehicle's electronic system can be described in a form of closed system: sensor – conditioning system – microcontroller – executing unit. Frequently we can find systems working in the scheme of closed loop: sensor – executing system – microcontroller – steering system – executing system. It applies to steering systems and systems overseeing the performance of an engine.

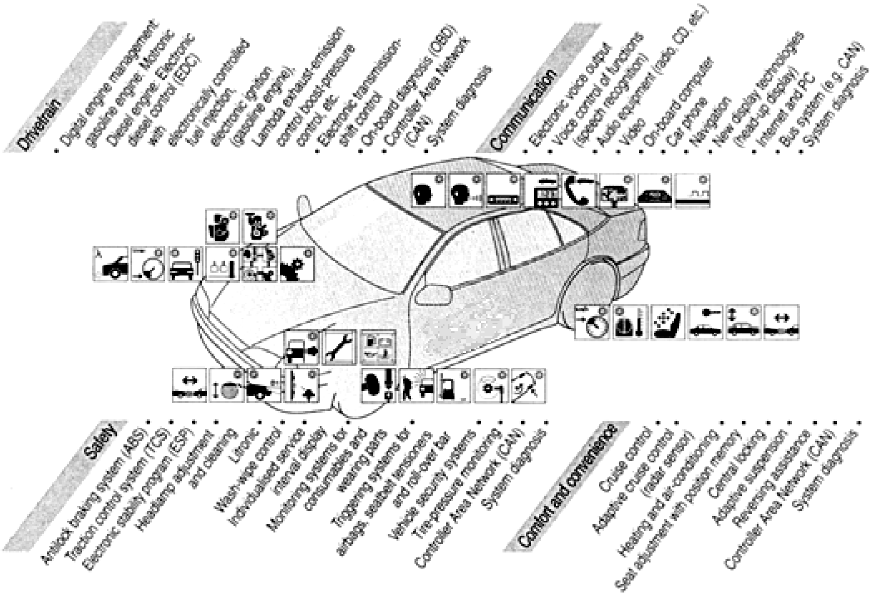


Fig. 2.1. Bosch electronic systems in a vehicle [7]

2.1. Buses

Depending on the type of a system, it is necessary to connect all of its components providing reliability and bandwidth (maximum frequency). There are a few types of buses used in vehicles: CAN, MOST, LIN and FlexRay. Buses are divided into classes (A, B, C, and D) depending on the bandwidth. The classification is shown in table below.

Tab. 2.1. Required bandwidth for vehicles' networks

Class	Communication Speed	Applications	LIN	CAN	FlexRay	MOST/1394
Class A	10K to 125Kbps (body)	Lamps, lights, power w/s, door locks, power seats, etc.	↕			
Class B	125K to 1Mbps (status information)	Electronic indicators, driving information, automatic air conditioner, failure diagnosis, etc.		↕		
Class C	1M to 10Mbps (real-time control)	Engine control, ABS, transmission control, break control, suspension control, etc.			↕	
Class D	10Mbps or faster (multimedia)	Car navigation system, audio system, etc.				↕

The first functional model of bus was introduced at the SAE (*Society of Automotive Engineers*) conference in 1986. It was a product of Robert Bosch's laboratories. In the following years CAN bus based system revolutionized the automotive industry. Despite the definite CAN bus's monopoly on the automotive market, new, revolutionary, competitive systems were introduced, i.e.: InterBus-S, Profibus, FIP, EIB, LON. The CAN bus is described in details in a separate chapter.

Word FlexRay™ comes from a combination of words *flexible* and *ray*. This name represents a consortium created in 2000 between DaimlerChrysler, BMW, Philips and Motorola. FlexRey's goal is to work on fast, reliable, and universal informatics network used in vehicles. The consortium, that in 2005 had 130 members, has a hierarchy structure. To the core of FlexRay consortium that makes all the decisions belong: BMW, Bosch, DaimlerChrysler, Freescale Semiconductor, General Motors, Philips and Volkswagen. The effect of cooperation between mentioned companies is the new version 2.1. The system has 10Mbps bandwidth in two channels, and has a collision avoidance system. FlexRay is suppose to become CAN's competition. But since it is a work in progress, it is still too expensive. FlexRay is described in a separate chapter.

In the late 1990s the Local Interconnect Network (LIN) Consortium was founded by five European automakers, Volcano Automotive Group and Freescale. The first fully implemented version of the new LIN specification was published in November 2002 as LIN version 1.3. In September 2003 version 2.0 was introduced to expand configuration capabilities and make provisions for significant additional diagnostics features and tool interfaces. LIN bus has a smallest bandwidth but is a very easy in implementation.

Improvements in the automotive industry demand larger bandwidth in bus in order to process more data. First bus with larger throughput that were implemented in large scale production were in Class S Mercedes. It was the D2B (digital data BUS). After few years, D2B had new competitors; IEEE1394 (Firewire) and MOST (Media Oriented System Transport).

First version of MOST was finished in 1998 as a result of cooperation between Audi, BMW, Deimler Chrystler, Harmann-Becker and Oasis Silicon System (the last two companies are leaders in the automotive electronic industry). We'll talk more about MOST in a different chapter.

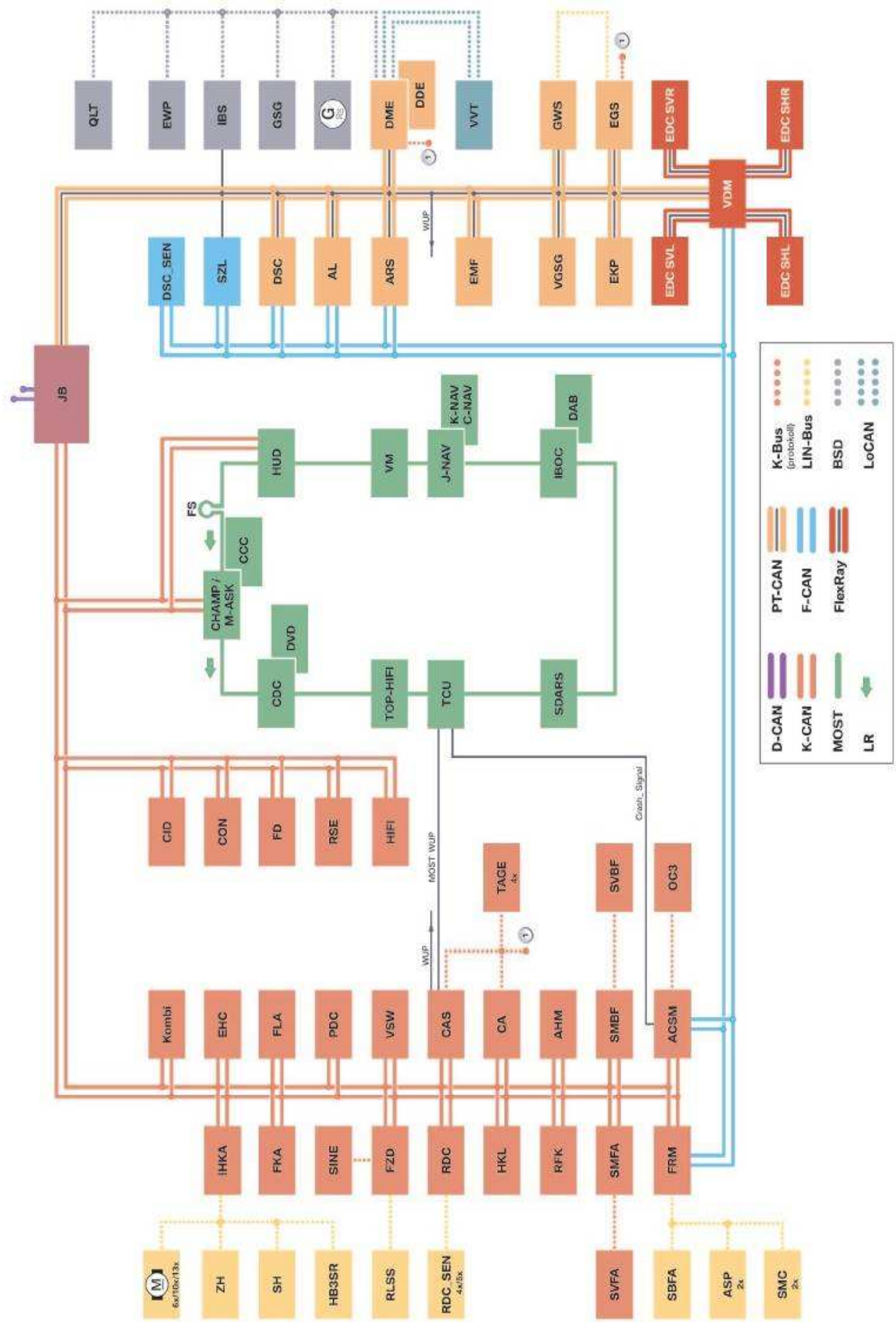


Fig. 2.2. Logic scheme of networks in vehicles [6]

Tab. 2.2. The abbreviations in vehicle's networking

Abbreviation	Explanation in German	Explanation in English
ACSM	ACSM-Steuergerät	ACSM (Advanced Crash Safety Management) control unit
AHM	Anhängermodul	Trailer module
AL	Aktivlenkung	Active steering
ARS	Aktive Rollstabilisierung	Active roll stabilising
ASP	Außenspiegel	Exterior mirrors
CA	Comfort Access	Comfort access
CAS	Car Access System	Car access system
CCC	Car Communication Computer	Car communication computer
CDC	CD Wechsler	CD-changer
CHAMP/M-ASK	Central Head Unit and Multimedia Platform/Multi-Audiosystem-Kontroller	Central head unit and multimedia platform/multi-audiosystem controller
CID	Central Information Display	Central information display
C-NAV	China Navigation	Navigation system China
CON	Controller	Controller
DAB	Digital Audio Broadcasting	Digital audio broadcasting
DDE	Digitale Diesel Elektronik	Digital diesel electronics
DME	Digitale Motor Elektronik	Digital engine electronics
DSC	Dynamische Stabilitäts-Control	Dynamic stability control
DSC_SEN	DSC-Sensor	DSC-sensor
DVD	DVD-Wechsler	DVD-changer
EDC SHL	Elektronische Dämpfer-Control Satellit hinten links	Electronic damping control rear left satellite
EDC SHR	Elektronische Dämpfer-Control Satellit hinten rechts	Electronic damping control rear right satellite
EDC SVL	Elektronische Dämpfer-Control Satellit vorn links	Electronic damping control front left satellite
EDC SVR	Elektronische Dämpfer-Control	Electronic damping control front

	Satellit vorn rechts	right satellite
EGS	Elektronische Getriebesteuerung	Electronic transmission control
EHC	Elektronische Höhenstands-Control	Electronic height control
EKP	Elektrische Kraftstoffpumpe	Electric fuel pump
EMF	Elektromechanische Feststellbremse	Electromechanical parking brake
EWP	Elektrische Wasserpumpe	Electric water pump
FD	Fond-Display	Rear display
FKA	Fond-Heiz-Klimaanlage	Rear temperature control
FLA	Fernlichtassistent	Headlamp assistant
FRM	Fußraummodul	Legroom module
FZD	Funktionszentrum Dach	Functional centre roof
GSG	Glühsteuergerät	Preheater control unit
GWS	Gangwahlschalter	Selection switch
HB3SR	Heizung Belüftung 3. Sitzreihe	Heating and ventilation 3rd row
HIFI	HIFI Verstärker	HIFI amplifier
HKL	Heckklappe	Back-cover lift
HUD	Head-Up-Display	Head-Up-Display
IBOC	High Definition Radio	High definition radio
IBS	Intelligenter Batteriesensor	Intelligent battery sensor
IHKA	Integrierte Heiz-Klima-Automatik	Integrated heating and airco control unit
JB	Junction Box-Steuergerät	Junction box control unit
J-NAV	Japan Navigation	Navigation system Japan
K-NAV	Korea Navigation	Navigation system Korea
Kombi	Instrumentenkombination	Instrumentation panel
OC3	Sitzbelegungsmatte US	Seat occupation
PDC	Park Distance Control	Park distance control
QLT	Quality-Level-Temperature Sensor	Quality level temperature sensor
RDC	Reifendruck-Control	Tire pressure control
RDC_SEN	Reifendruck-Control Sensor	Tire pressure control sensor
RFK	Rückfahrkamera	Backwards driving camera
RLSS	Regen-Fahrlicht-Solarsensor	Rain-light-sun sensor

RSE	Rear Seat Entertainment	Rear seat entertainment
SBFA	Schalterblock Fahrer	Switching block driver
SDARS	Satellitentuner	Satellite tuner
SH	Standheizung	Interior preheater
SINE	Sirene Neigungsgeber	Siren slope angle sensor
SMBF	Sitzmodul Beifahrer	Seat module passenger
SMC	Stepper Motor Controller	Stepper Motor Controller
SMFA	Sitzmodul Fahrer	Seat module driver
SVBF	Sitzverstellung Beifahrer	Seat adjustment passenger
SVFA	Sitzverstellung Fahrer	Seat adjustment driver
SZL	Schaltzentrum Lenksäule	Switching centre steering wheel
TAGE	Türaußengriffelektronik	Exterior door grip electronics
TCU	Telematic Control Unit	Telematic control Unit
TÖNS	Thermischer Ölniveausensor	Thermal oil level sensor
TOP-HIFI	Top-HiFi-Verstärker	Top HIFI amplifier
VDM	Vertikaldynamikmanagement (zentrales Steuergerät für die elektronische Dämpfer-Control)	Vertical dynamic management (central control unit for electronic damping congrol)
VGSG	Verteilergetriebe-Steuergerät	Transfer-gearbox control unit
VM	Videomodul	Video module
VSW	Videoswitch	Videoswitch
VVT	Variabler Ventiltrieb	Variable valve drive
ZH	Elektrischer Zuheizer	Electrical additional heating
Corresponding to buses		
BSD	Bitserielle Datenschnittstelle	Bit Serial Data interface
Crash_Signal	Crash-Signal	Crash signal
D-CAN	Diagnose-CAN	Diagnosis-CAN
F-CAN	Fahrwerks-CAN	Chassis-CAN
FlexRay	FlexRay-Bus-System	FlexRay bus
K-Bus	Karosserie-Bus	Body-bus
K-CAN	Karosserie-CAN	Body-CAN

LIN-Bus	Local Interconnection Network-Bus	Local Interconnection Network-bus
LoCAN	Local-CAN	Local-CAN
MOST	Media-Oriented-System-Transport	Media-Oriented-System-Transport
MOST WUP	MOST-Wake-up-Leitung	MOST-wake-up wire
PT-CAN	Powertrain-CAN	Powertrain-CAN
WUP	Wake-up-Leitung	Wake-up wire
1	CAS Bus-Verbindung	CAS bus connection

2.2. EOBD

OBD *(On Board Diagnosis*) standard was created in 1988 in California, USA. OBD was the effect of over 20 years of research on lowering the emission of exhaust gases to the atmosphere. The system's characteristics are:

1. Necessity of informing about vehicle's problems that could cause the exhaust gases to be polluted with toxins,
2. The need for every vehicle to have a self-diagnostic system,
3. Access to the vehicle's state of health information by its owner, and certainty that the information will be stored.

In 1994 ODB II standard was introduced. In 2000 it was brought to Europe, and named EOBD (*European On Board Diagnosis*). In literature EOBD is often falsely compared to OBD, where in fact EOBD is the equivalent of OBD II.

The main function of vehicles equipped with EOBD/OBDII is protection of the environment by constant monitoring of its emissions. Every vehicle that has EOBD/OBDII standard can self-diagnose 849 problems.

Companies are constantly working on introducing EOBDII standard. EOBDII is supposed to diagnose more problems and assist service technicians in diagnosing problems. In few years an on board diagnose system is not only going to be diagnosing problem, but also predict them, and send information to service stations. It is going to make an appointment with repair technicians however, first asking the vehicle's owner if the appointment date and hour are convenient for him or her. The work on EOBD II has a goal of introducing digital signal processing and implementation possibilities in modern auto-diagnostic systems. Diagnostic channel equipped in this type of systems will diagnose the problems in much shorter time (quasi real time) and because of that will be able to diagnose more malfunctions.

The measure of the malfunction's code is done by a specialized computer in the diagnostic stations. The computer can also erase the error's code into computer.

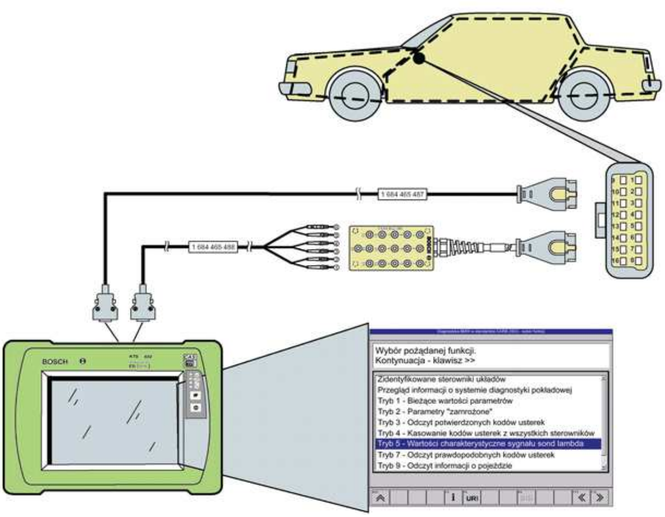


Fig. 2.3. The measure of errors from EOBD [7]

In 1991 a DIN ISO 9141-2 standard was introduced, that is the equivalent of American standard. The DIN ISO 9141-2 describes a standard plug of EODB, diagnostic tester (scanner), data transmission protocols, and definitions of error codes. It only allowed for unique system of communication and also included it in ODB II standard. This system assured independence of systems belonging to EODB from ODB II.
The standard also defines the location of a plug. It says that the plug should be reachable from the drivers seat. It is usually located between steering wheel's center line and the steering column. Sometimes the plug is installed in the glove compartment.

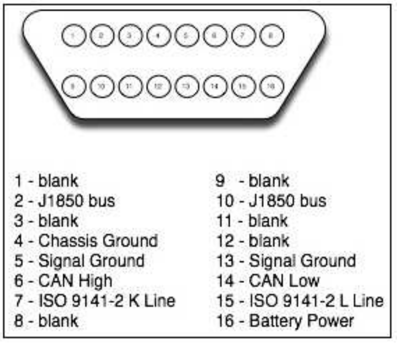


Fig. 2.4. Standard plug of EOBD [6]

The first diagnostic standards were *Keyword Protocol 2000* bus and CAN. Since 2009 CAN is the only standard. Error code has 5 fields:

1. Field 1 describes the system that the code defines,
2. Field 2 describes the type of the error code,
3. Field 3 describes the subsystem where the failure has occurred,
4. Field 4 and 5 describe the type of problem or point to the malfunctioning circuit.

Tab. 2.3. The EOBD code for idle control system

Code	**Amount**
P0500	Vehicle Speed Sensor Malfunction
P0501	Vehicle Speed Sensor Range/Performance
P0502	Vehicle Speed Sensor Circuit Low Input
P0503	Vehicle Speed Sensor Intermittent/Erratic/High
P0505	Idle Control System Malfunction
P0506	Idle Control System RPM lower Than Expected
P0507	Idle Control System RPM higher Than Expected
P0510	Closed Throttle Position Switch Malfunction
P0520	Engine Oil Pressure Sensor/Switch Circuit Malfunction
P0521	Engine Oil Pressure Sensor/Switch Range/Performance
P0522	Engine Oil Pressure Sensor/Switch Low Voltage
P0523	Engine Oil Pressure Sensor/Switch High Voltage
P0530	A/C Refrigerant Pressure Sensor Circuit Malfunction
P0531	A/C Refrigerant Pressure Sensor Circuit Range/Performance
P0532	A/C Refrigerant Pressure Sensor Circuit Low Input
P0533	A/C Refrigerant pressure Sensor Circuit High Input
P0534	Air Conditioner Refrigerant Charge Loss
P0550	Power Steering Pressure Sensor Circuit Malfunction
P0551	Power Steering Pressure Sensor Circuit Range/Performance
P0552	Power Steering Pressure Sensor Circuit Low Input
P0553	Power Steering Pressure Sensor Circuit High Input
P0554	Power Steering Pressure sensor Circuit Intermittent
P0560	System Voltage Malfunction
P0561	System Voltage Unstable
P0562	System Voltage Low
P0563	System Voltage High
P0565	Cruise Control On Signal Malfunction
P0566	Cruise Control Off Signal Malfunction
P0567	Cruise Control Resume Signal Malfunction

P0568	Cruise Control Set Signal Malfunction
P0569	Cruise Control Coast Signal Malfunction
P0570	Cruise Control Accel Signal Malfunction
P0571	Cruise Control/Brake Switch A Circuit Malfunction
P0572	Cruise Control/Brake Switch A Circuit Low
P0573	Cruise Control/Brake Switch A Circuit High
P0574	Cruise Control System - Vehicle Speed Too High
P0575	Cruise Control Input Circuit
P0576	Cruise Control Input Circuit Low
P0577	Cruise Control Input Circuit High
P0578	through P0580 Reserved for Cruise Control Codes

PROBLEM

Work in groups. Find the error code for indicator: "Engine Speed Input Circuit No Signal". Use the information: http://www.bba-reman.com/uk/index.aspx

The driver, who does not possess diagnostic equipment, can also notice a malfunction which is detected by the EOBD. Vehicles are equipment with diagnostic signal lights MIL that are located on the dashboard. Their shapes are reserved by the standard.

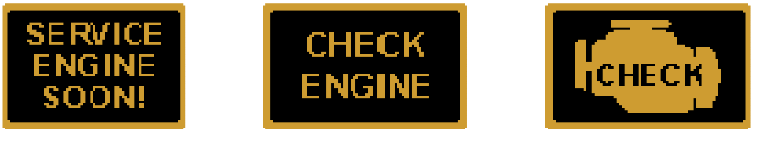


Fig. 2.5. MIL EOBD indicators

If one of the lights is on, it means that in the last three driving cycles there was a malfunction and that the vehicle should be inspected by a service technician.

2.3. EOBD II

According to major service centers, EOBD II should have been in use since 2009. However, it is still in the design phase. It is a result of very high goals, such as:

1. Increase of diagnosed malfunctions (from 849 to over 2000),
2. Increase in sensor sensitivity so that they are able to detect ever increasing emission standards,
3. Ability to predict a failure,
4. Increase in the speed in the processing of diagnostic signals (as well as the bandwidth) to the points where diagnosis in made in quasi real time,
5. Creation of an interface that would allow the service station to continuously communicate with the vehicle through an IP based network,
6. Increasing the ability self-diagnose problems.

We can predict what full digitalization of vehicles will do; driver unaware of the multitude of diagnostic systems working simultaneously will be notified about a possibility of a system failure. Then the system, having a continuous contact with the service station will ask for a convenient date to fix the problem. If the driver does not specify a date, the system will go into an emergency state and limit the vehicles capabilities, such as maximum speed.

References

[1] Merkisz J., Mazurek S.: *Pokładowe systemy diagnostyczne pojazdów samochodowych, W*ydawnictwa Komunikacji i Łączności, Warsaw, 2002.

[2] „Engine Technology International” June 1999.
[3] *The Thrill of Solution: Dynamometers and Drivers,* AVL, update 12.4.2001.
[4] A. Janicka, A. Kaźmierczak, M. Tkaczyk, R. Wróbel: Contour of vibroacoustic map in diagnoses of engine's failure. Journal of KONES. 2009, vol. 16, nr 2, s. 217-224.
[5] Elektor Electronics *CAN bus*, Segment B.V., the Netherlands 1998.
[6] Rokosch U., *Układy oczyszczania spalin i pokładowe systemy diagnostyczne samochodów OBD,* Wyd. WKŁ, Warszawa 2007.
[7] Informatory techniczny BOSCH, *Czujniki w pojazdach samochodowych*, Wyd. WKŁ, Warszawa 2007.

3. The electronic fuel control system

Direct injection system delivers fuel directly to the combustion chamber. Earlier injection systems delivered fuel into a carborator or into a mixing chamber in diesel engines, and into intake manifold or directly into the head. First models in both types of engines were mechanical. In 1970's, as a result of higher emission standards, electronic gasoline injection system started to be produced. In 1988 FIAT introduced a first direct injection system in a diesel engine that is used as a passanger vehicle (TD-ID). A year later, Volkswagen intorduced its own version in Audi 100 (TDI).

However, in 1997 Mitsubishi started mass producing GDI engine with direct, electronic gasoline injector. Electronic direct injector is widely used today in both gasoline (GDI, FSI, IDE, HPi, D4, JTS etc.) and in diesel engines (HDi, JTD, CDI, dCi, TDI, D4D etc.). Engines that are supplied with fuel by an electronic direct injection system have a very precise pressure control, as well as of the timing and quantity of fuel delivered to the combustion chamber. In gasoline engines, the greatest advantage of the direct injection system is its ability to achieve layered combustion.

In case of diesel engines, direct injection allows full control over the propulsion system due to the precise amount of fuel delivered, its timing and its duration, as well as the degree of atomization which allows for more economical operation of the engine and lowering the toxicity of its emissions. In the most modern systems of this type, it is possible to control the combustion temperature and control the flow of fuel into the combustion chamber, as well as precision injection of pilot quantity of fuel. In order to take full advantage of direct injection in diesel engines, the injection pressure must be extremely high. The disadvantages of direct injection is rough and noisy operation of the engine and lower maximum speed, are more than compensated by its advantages such as its responsiveness, lower fuel consumption and lower toxicity of its emissions.

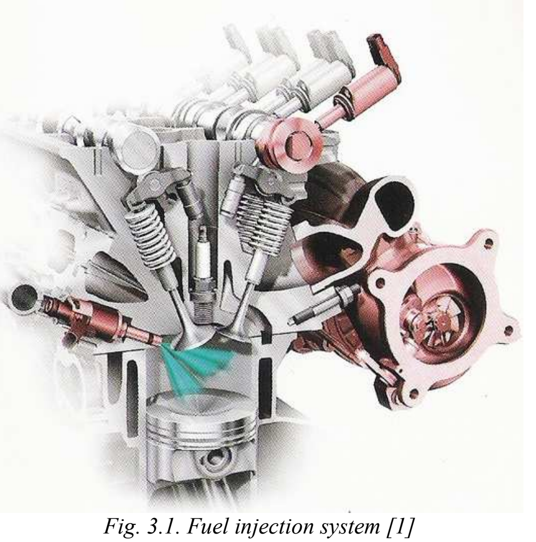


Fig. 3.1. Fuel injection system [1]

Engines using spark plus as ignition source

To consume one kilogram of fuel, 15 kilograms of air are necessary. In order for the amount of fuel injected to be appropriate, the amount of air entering the combustion chamber must be measured, and based on this amount a proper amount of fuel should be injected into the chamber. This work is done by electronic – microprocessing units.

In order to measure the amount of air – the easiest way is to check the angle of the throttle and the speed of the engine (everything will work as long as the air pressure remains the same). At lower air pressures, and at the same throttle angle and the same engine speed, the mass of intake air will be much lower.

One of the solutions is to measure pressure at the intake manifold (and its temperature) and the speed of the engine. On this basis, we can determine the density of air at the intake manifold, and thanks to the known speed of the engine and its volume we can determine the volumetric air flow and appropriately determine the amount of fuel. The problem with this solution is that depending on the speed of the engine, the cylinder is irregularly filled with air, it intakes more or less air. To eliminate this problem, a map with given time of injection and measured engine speeds and intake manifold pressures is written into the memory of the driver (the throttle angle may also be entered). In this case, the driver which determines the amount of injection regulates it through the duration of the injection.

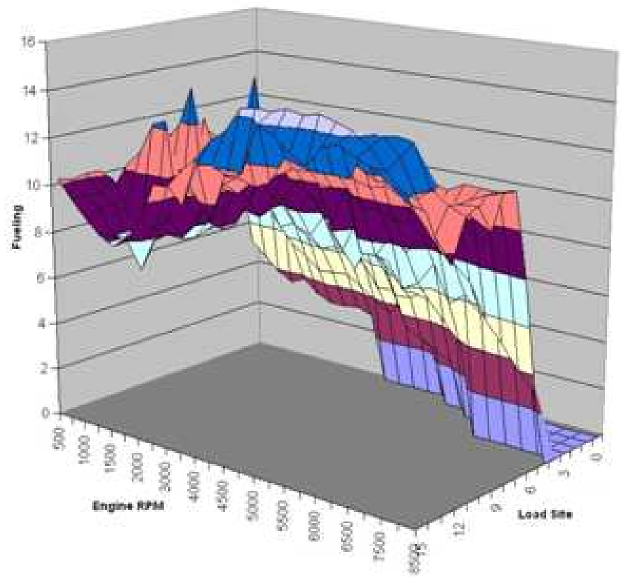


Fig. 3.2. The example of fuelling graph

The most accurate method of measuring the volumetric air flow is through the use of heated wire or heated plate. The wire is heated to 100 degrees C or more, while the flowing air is cooling it. The measurement is based on the determination of the voulume of air needed by the wire to keep a constant temparature, even though the air is cooling it.

Injection system use a very important piece of equipment: oxygen sensor. The sensor determines the amount of oxygen in emissions (in most cases it simply determines the presence or lack of oxygen). If there is too much oxygen, this means that the fuel mixture is too lean. However, if there is no oxygen, then the mixture is too rich (the oxygen sensor is unable to recognize an ideal mixture level). On the basis of the sensor's measurements, the driver corrects the pre-determined timing of the injection in order to bring the emission levels to normal.

This correction works only when the vehicle is moving at a constant speed or when it is in a neutral gear (only at pre-determined states of engine's work). This happens because correction is not in one of the pre-determined states of the engine, for example during acceleration, it would generate a lagging effect – the mixture is made at the intake manifold, whereas the emissions are measured at the exhaust manifold.

Fig. 3.3. Oxygen sensor (lambda sensor)

Off course, other than the volumetric air sensor, it is necessary to use other systems to support decisions made by the microprocessor that controls electronic fuel injection. They are presented in a figure below.

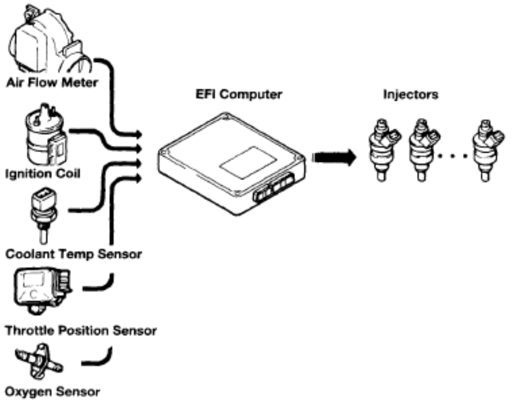


Fig. 3.4. The elements of electronic fuel injection system [2]

There are a few advantages for EFI (e.g.):

1. Each cylinder has its own injector which delivers fuel directly to the intake valve. This eliminates the need for fuel to travel through the intake manifold, improving cylinder to cylinder distribution,
2. Supplies a continuously accurate AFR to the engine no matter what operating conditions are encountered. This provides better drivability, fuel economy and emission control,
3. By delivering fuel directly at back of the intake valve, the intake manifold design can be optimized to improve air velocity at the intake valve. This improves torque and throttle response,
4. Cold engine and wide open throttle enrichment can be reduced with an EFI engine because fuel paddling in the intake manifold is not a problem. This results in better overall fuel economy and improved emission control,
5. The combination of better fuel atomization and injection directly at the intake valve improves ability to start and run a cold engine,
6. The EFI doesn't rely on any major adjustments for cold enrichment or fuel metering. Because the system is mechanically simple, maintenance requirements are reduced.

3.1. The Lambda sensor

There are two types of Lambda sensors:

1. Narrowband lambda sensor,
2. Wideband lambda sensor.

Narrowband do not produce a linear output for a given AFR (Air / Fuel Ratio) and therefore, they are only accurate for a narrow range around 14.7:1 AFR, which is also called the stoichiometric reading. At 14.7:1, the oxygen atoms react completely with the fuel atoms and therefore, ideally, there are no left over air or fuel atoms. This is where cars are designed to idle and operate at low throttle. This is why narrowband only target this value and its immediate range. When the mixture becomes rich or lean, it is virtually impossible for a narrowband to estimate the AFR.

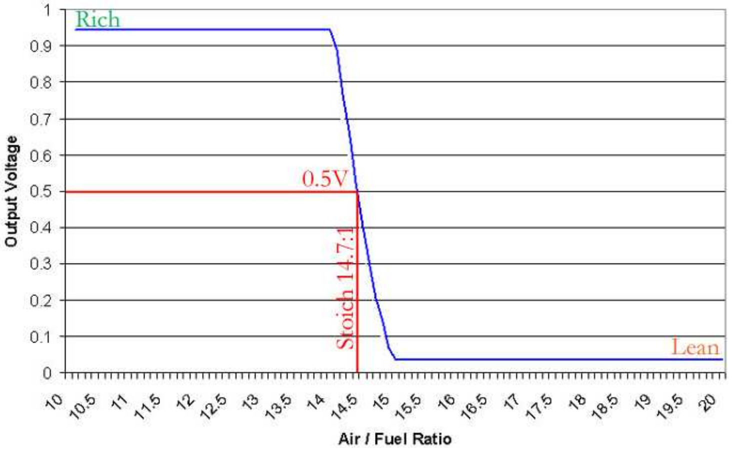


Fig. 3.5. Narrowband output: AFR vs. voltage

Wideband O2 sensors on the other hand work slightly different. They incorporate electrochemical gas pump, which delivers oxygen into a measuring chamber, instead of relying on diffusive flow. This way, measuring the air concentration becomes very much direct and the sensor does not need to communicate with the ECU in order to estimate an AFR. In addition, wideband are also therefore able to produce a specific output for a given oxygen concentration, which makes it possible to obtain an AFR reading for a range much wider than the one of a narrowband.

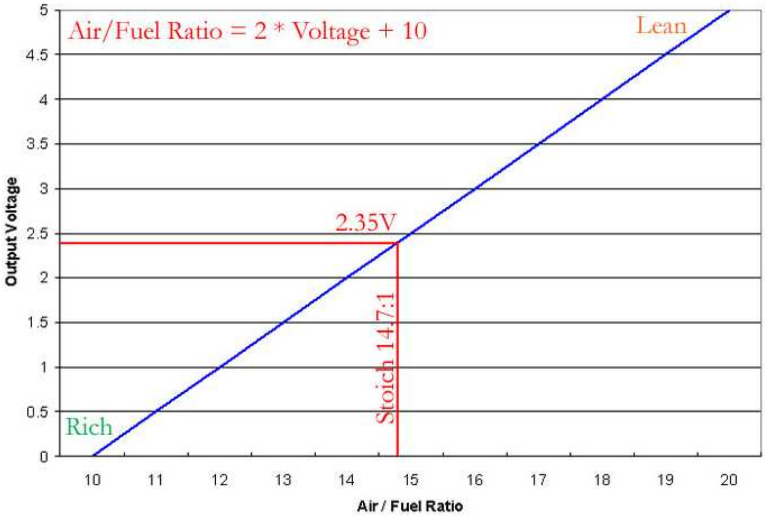
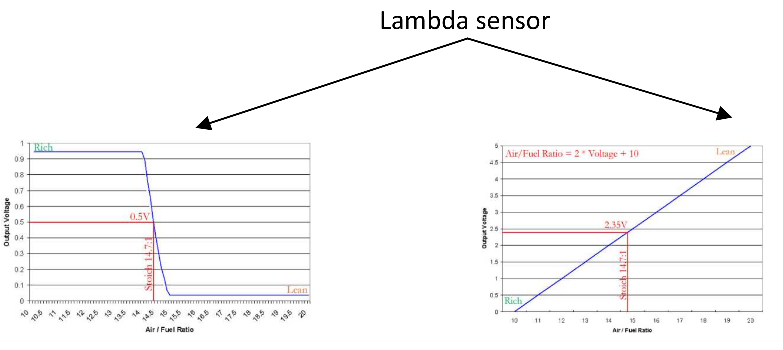


Fig. 3.6. Wideband output: AFR vs. voltage

PROBLEM
Work in pairs. What are the main differences between narrowband and wideband lambda sensor?



. . . .

. . . .

. . . .

3.2. Common Rail

Governments and car manufacturing plants have been under increasing pressure to produce cars with substantially less harmful smoke emissions. Upgrades on previous diesel engines have allowed for this. Fuel efficiency, reduced engine noise and improved fuel injection systems are being included as a non-negotiable standard for many diesel production cars nowadays. The Common Rail is the best facilitator for these increasing standards of production.

Because a build up of pressure is required in the Common Rail Accumulator Injection system, the actual build up of injection pressure is separate from the injection action itself. Pressure is generated in The Rail (also known as an accumulator), from a high pressure pump. The pressure build up is determined by the injection pressure settings in the ECU (Engine Control Unit) and is usually around 1600 bar. It should be noted here that the ECU and high pressure pump function totally independently from the quantity of fuel injected and the engine speed of the car. The fuel is fed through a piping system and eventually ends up at the injectors which then inject the exact amount of fuel into the combustion chambers of the engine. The pressure in the rail, the duration and the timing of the injection as well as a list of other injection parameters and engine functions are controlled by the EDC (Electronic Diesel Control).

The Common Rail has undergone a series of development over the ages. In the First and Second Generation Common Rails, the hydraulic force build up was previously controlled by a magnetic solenoid on the injectors and the force was then transmitted to the injector needles by a piston rod. Nowadays, injection actuators consist of thousand of thin crystal-like wafers that have an ability to rapidly expand when an electric current is applied to them. In this crystal wafer like injector structure, the actuator is positioned close to the jet needle and no mechanical parts are needed to facilitate its friction free transmission to the jet needles.

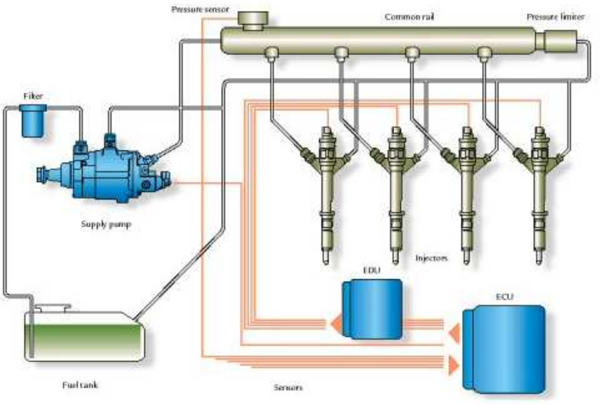


Fig. 3.7. The Common Rail system [4]

Electronic diesel injection system consists of sensors, ECU (computer) and the implementing agencies, and other components. Its mission is to electronically controlled fuel injection system, fuel injection volume, and the realization of the injection timing with the operating status of the real-time control. Using speed, temperature, pressure sensors, real-time detection parameters will be entered into the computer simultaneously, and had been stored values of the parameters of comparison, has been treated in accordance with the calculation of the best value on the fuel pump, exhaust gas recirculation valve, such as preheating Cypriot implementing agencies control, driven fuel injection system to achieve the best diesel engine operating status.

3.3. The experimental system

The central part of ECU is the microcontroller (MCU). Motorola’s serious MCU has become an industrial standard in engine control. Both GM Inc. and Caterpillar Inc. (the biggest producers of Common Rail System) adopted Motorola's MCU in their electronically controlled engine products and made remarkable results. In this paper, the high-performance 8-bit MC68HC908 GP32 MCU is employed to accomplish control functions, and this chip has the following advantages:

1. 32 kB of FLASH memory with in-chip programming capabilities and in-chip programming firmware for use with host personal computer which does not require high voltage for entry,
2. 8 MHz internal bus frequency and clock generator module with 32 kHz crystal compatible Phase Lock Loop (PLL),
3. Two 16-bit, 2-channel timer interface modules with selectable input capture, output compare, and Pulse Width Modulation (PWM) capability on each channel,
4. 8-channel, 8-bit successive approximation analog-to-digital converter (ADC) to satisfy multi-input requirements,
5. System protection features such as optional computer operating properly (COP) rest, low-voltage detection with optional rest and illegal address detection.

The top dead center (TDC) signal of cylinder must be detected first for precise timing control. A signal plate is installed on the rim of flywheel. When the engine piston arrives at the top dead center position, the signal plate moves faced to a magnetoelectric sensor, and then the TDC signal becomes active. A reshaping circuit is used to regular this signal to supply MCU for input capture. A rotary timing wheel is attached to the engine crank shaft, and produces 720 standard square waves every cycle. The engine speed can be calculated in MCU using these signals, and coordinating with TDC signal, the exact moment of HSV operation is determined with a tolerance of less than 0.5 crankshaft angle. The analog signals such as oil pressure, oil temperature and pedal signal are induced to ECU’s AD channel through an amplify circuit and a modulate circuit.

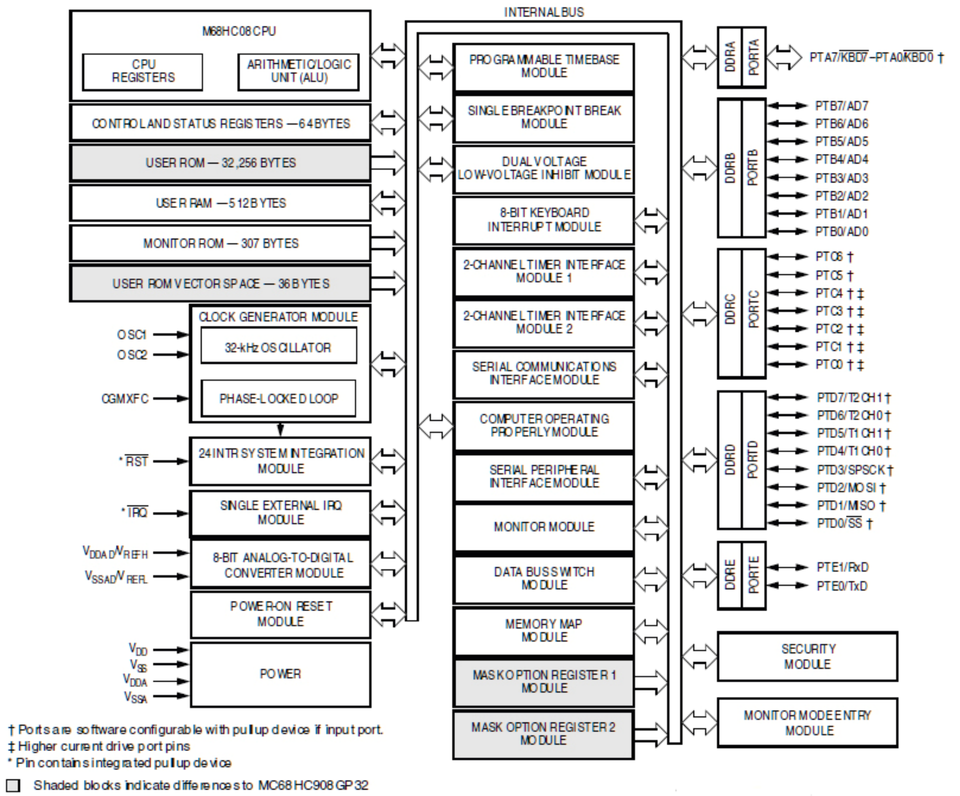


Fig. 3.8. MC68HC908GP32 microcontroller block diagram scheme [1]

PROBLEM
What are main parts of MC68HC908 family circuit?

The HSV in injector undertakes the task of injection control. The injector must be able to quickly open to time the delivery of fuel into the combustion chamber. It also must be able to raise pressure quickly in order to stop the delivery of fuel. The dynamic response characteristics of the injector directly affect the performance of the injection system. In order to create strong force to overcome the resistance of return spring, a 60V is adopted for the solenoid, and the Pulse Width Modulation (PWM) control method is used to keep the holding current. When the MCU gives an injection control pulse, the injector gets 60V, and the current in solenoid rises quickly according to the system's resistance and inductance. 1ms later, the circuit turns to PWM mode, and the average current of PWM pulse keeps the valve open with low power dissipation. The maximum open current is 4 ampere and the average holding current is 1.5 ampere. This operation mode ensures the injector to run reliably for extended periods of time.

In order to acquire engine real-time status, DM-162, a character liquid crystal diode (LCD) display chip is used to monitor the engine speed, injection timing and injection pulse width. The small-size DM-162 uses low power, and stores 160 different array character graphs in its internal character generator. It has convenient interface to MCU. It is easy to configure the display function by giving the corresponding character an address in the software.

The control software is edited by Motorola assembly language, and it includes the following modules:

1. Injection control module: On the condition where the common rail is acquiring stable pressure from pump, the MCU calculates the injection quantity on the basis of engine speed and pedal position signal. After modification, the MCU produces a corresponding pulse to the HSV drive circuit. A closed loop control algorithm is used to adjust the injection quantity to stabilize engine speed. The injection timing in different engine operation modes is acquired by table-look-up method,
2. Data acquisition and process module: This module scans the signals from various sensors, and uses software filtering to eliminate false signals due to engine disturbance,

3. Display module: According to the display demand, the engine speed, the crankshaft angle of starting injection and injection pulse width are showed in different position of the LCD screen. In every software cycle, the data refreshes to keep the real-time engine status.
4. Malfunction process module: This module judges whether the signals from sensors exceed its limited value. When a serious error occurs, the MCU turns off for safety.

3.4. EGR system

Since the 1960s, when rapid degeneration of the environment was first noticed, ways of reducing the emission of dangerous substances by automotive vehicles have been sought. The dangerous compounds are hydrocarbons (HC), nitrogen oxides (NO_x) and carbon dioxide (CO_2). The dangerous emissions can be reduced by:

1. Improving the fuel,
2. Improving lubrication,
3. Maintaining higher thermal stability,
4. Perfectly insulating the system,
5. Using catalytic systems.

Among the systems which have a bearing on engine thermal economics by reducing lean mixture temperature there is the Exhaust Gas Recirculation (EGR) system whose principle of operation consists in forcing some of the exhaust gas back into the combustion chamber.

The first experiences with EGR systems date back to the 1970s. The operation of the system was then limited to steady feeding the exhaust gas to the combustion chamber when the vehicle user turned on the system by means of a proper switch. That system only partially fulfilled its functions which included:

1. Lowering the combustion temperature of the lean mixture,
2. Oxidizing harmful substances,
3. Accelerating fuel vaporization.

As one can easily guess, the first vehicles with the EGR, which, thanks to General Motors, appeared on the American market in 1973 contributed little to emission reduction. It was mainly the fault of the operator who would decide when the EGR was and was not to work. In the late 1970s the system was improved by introducing a primitive diagnostic system whose integral part was a temperature sensor located on the cooler. It would turn on the EGR (provided the decision unit, i.e. the operator had switched on the whole system) only at specified engine (coolant) temperatures. The next generations of the system were equipped with timing circuits which would switch off the EGR for a few seconds after the throttle was fully opened.

In 1983 research on the 4th generation EGR which is driver-independent, i.e. it takes its own decisions about switching the system on or off, started. Besides taking the right switching decision, the system also decides what percentage of the exhaust gas can be turned back to the combustion chamber. The modern EGR takes a decision to switch on the system only if the following conditions are fulfilled:

1. The engine temperature is higher than 77˚C,
2. The temperature under the bonnet is above -6˚C,
3. The engine has been working for at least 3 minutes at the above temperatures,
4. The crankshaft rotational speed is 1952-2400 rpm for the manual gearbox,
5. The crankshaft rotational speed is 2248-2688 rpm for the automatic gearbox,
6. The exhaust gas overpressure is 667-2667 Pa,

7. The fuel temperature does not differ from the one specified by the vehicle manufacturer (T_o) by -8% to +7%,
8. The voltage generated by the throttle opening sensor is in a range of 0.6-1.8V,
9. The driving speed is higher than 40 km/h.

EGR systems are highly complex, particularly with regards to their electronic and sensor system. However, when one examines the evolution of the EGR, one can notice similarities to other emission reducing system. This is illustrated in fig. below.

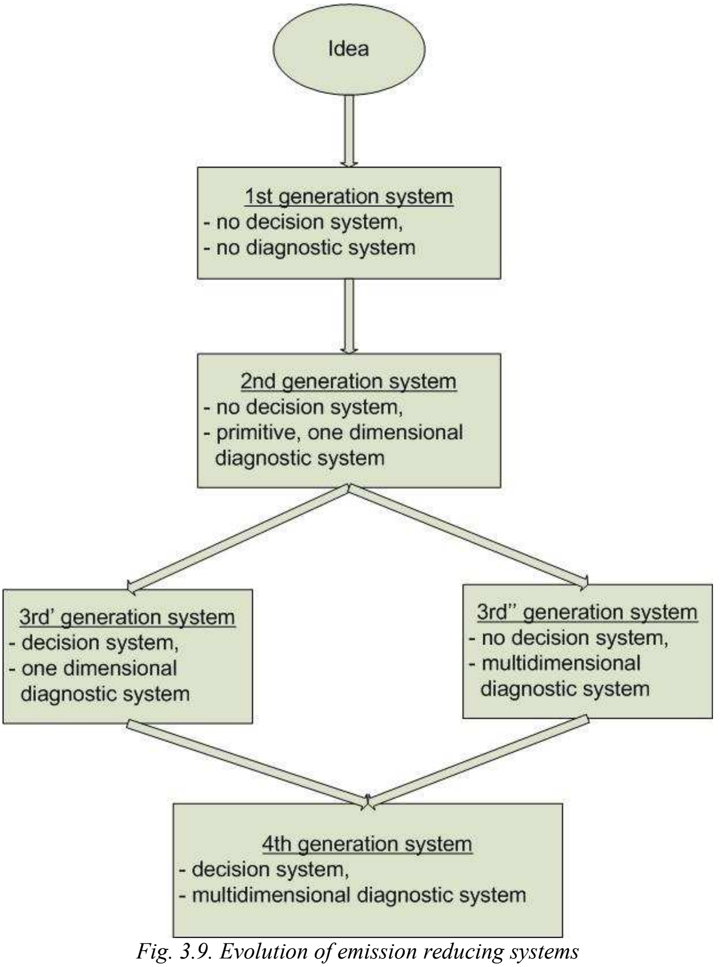


Fig. 3.9. Evolution of emission reducing systems

The emission reduction idea led to the development of the 1st generation EGR which is a simple implementation of this idea. Its control is limited to switching the system on by the driver at a proper moment, similarly as in other systems, e.g. in the 1st generation LPG system. The 2nd generation systems are equipped with a temperature sensor to aid the operator in decision making.

The systems installed in cars have evolved in two ways, depending on system complexity and popularity. If the system is simple and there is a demand for it, a hybrid is usually created by adding a processor diagnostic system which eliminates the human from decision making. The diagnostic system takes decisions on the basis of one- or two-dimensional information coming from sensors (e.g. temperature and crank sensors). This is the case in LPG systems. The EGR has evolved in the other way, i.e. the decision still belongs to the driver who can switch the system off, but he/she is aided by systems consisting of many sensors (the 3rd generation system).

The synthesis is the 4th generation system. Here the driver does not make any decisions. Everything proceeds automatically, but so as not to disturb the operation of the engine or adversely affect driving safety and ergonomics.

Although the idea seems to be simple, since it consists of pumping some of the exhaust gases back to the combustion chamber, its implementation is not so easy. Two groups of systems are distinguished:

1. Pneumatically controlled,
2. Electronically (processor) controlled.

As a rule, the operation of electronically controlled systems is based on a decision unit in the form of a (micro-) processor system. The decision whether to switch the system on and about the amount of exhaust gas which is to be forced back into the combustion chamber belongs solely to the integrated circuit which makes the decision on the basis of the information coming from the sensor CAN bus.

The EGR, belonging to the group of negative pressure-controlled sensors, is employed in both supercharged and not-supercharged engines. Generally speaking, in all the EGR systems the valve is opened by a negative-pressure servomotor and closed by an elastic element. Sub-pressure is produced by a double-purpose pump called a tandem pump, which is a combination of a fuel pump and a sub-pressure pump in one housing. The sub-pressure value is adjusted by an electrovalve controlled by the information contained in the rectangular signal. In this case, this is pulse-duty factor k_w :

$$k_w = \frac{t_i}{T} \tag{1}$$

where: t_i – pulse duration, T– the period.

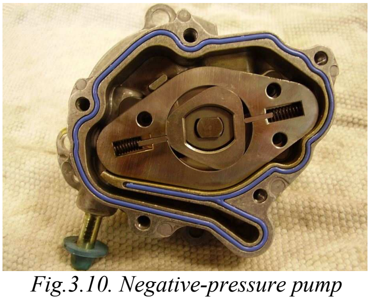


Fig.3.10. Negative-pressure pump (fuel pump is on other side)

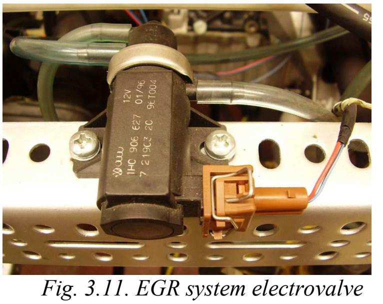


Fig. 3.11. EGR system electrovalve

When the pulse-duty factor is close to 0, the electrovalve is closed. The degree of valve opening is directly proportional to the value of the factor and the maximum opening is reached at k_w=1.

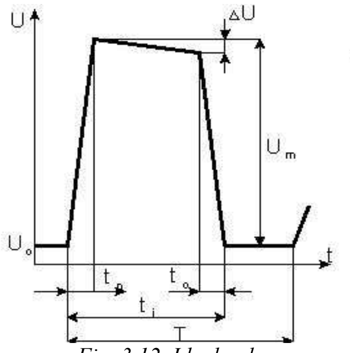


Fig. 3.12. Ideal pulse

The control system allows switching of the ERG only in certain engine operating conditions. But never are all of the exhaust gases forced back into the combustion chamber. The amount of exhaust gas which is forced depends on:

1. The mass of the air sucked in by the engine,
2. The volume of the air sucked in by the engine,
3. The throttle opening angle,
4. The absolute pressure of the intake manifold,
5. The exhaust gas overpressure in the exhaust system.

In the literature on the subject and also on Internet forums the EGR and the need for its existence are often discussed. Unfortunately, many users who know how the EGR system works switch it off to improve the performance of their vehicles and prevent failures in the intake systems. As a result, the EGR valve in the supercharged engine gets gummed up, as shown in fig. below. Carbon deposit accumulation may totally block the EGR electrovalve and thereby damage the lambda probe, which leads to serious consequences.

Fig. 3.13. Gummed up intake valve of EGR

The EGR valve can also be blocked when the negative-pressure system is released or the pulse length control system is damaged. Negative-pressure system leakage is a typical failure of systems controlled this way. The problem here is that the pipes feeding negative pressure to the EGR system are located close to the hot engine parts. Damage to a negative pressure feeding pipe results in, among other things, engine stalling both when the engine is cold and hot.

Besides the failures mentioned above, the sensors which aid the processor unit in making the ERG switch decision that may fail include:

1. A coolant temperature sensor (CTS), which has an NTC (Negative Temperature Coefficient) thermistor. The typical resistances of this sensor are shown in the table below,

Tab. 3.1. Resistance values against coolant temperature sensor temperature

No.	R[Ω]	t[˚C]
1.	177	99
2.	467	70
3.	1459	40
4.	3520	20
5.	9420	0
6.	28680	-20
7.	100700	-40

2. A throttle position sensor (TPS), which is a potentiometer measuring voltage drop

3. An oxygen sensor (OS), which is located in the throttle system and measures air volume. It
4. usually works in tandem with a rate generator coupled with an a/d converter
5. An intake air temperature (IAT) sensor, which, similarly as the coolant temperature sensor,
6. An NTC thermistor. As opposed to the thermistor which measures coolant temperature,
7. A semiconductor thermistor with a different dopant is used in this case. This affects the resistance values versus temperature,

Tab. 3.2.Resistance values versus intake air temperature sensor temperature

No.	R[Ω]	t[˚C]
1.	185	99
2.	450	70
3.	1800	40
4.	3400	20
5.	7500	0
6.	25000	-20
7.	100700	-40

8. A crank sensor (CS),
9. A manifold absolute pressure (MAP) sensor.

A failure of any of the systems mentioned above will cause a malfunction of the EGR system. EGR faults have a special symbol in the EOBD (OBDII) code, i.e. the value 4 in the 3rd field of the fault code.

PROBLEM
Work in groups. What are the advantages of an EGR system? What are the dangers of an EGR failure (or switching it off)?

References

[1] Free on line researchers papers.
http://www.freeonlineresearchpapers.com

[2] Glassey S.F, Stockner A. R., Flinn M. A. *HEUI-A New Direction for Diesel Engine Fuel Systems. SAE paper 930270.*
[3] Stockner A. R., Flinn M. A., Camplin F. A. *Development of the HEUI Fuel System Integration of Design, Simulation, Test, and Manufacturing*. SAE paper 930271.
[4] Uchida N., Shimokawa K., Kudo Y., Shimoda M. *Combustion Optimization by Means of Common Rail Injection System for Heavy-duty Diesel Engines*. SAE paper 982679
[5] Merkisz J., Mazurek S.: *Pokładowe systemy diagnostyczne pojazdów samochodowych, W*ydawnictwa Komunikacji i Łączności, Warsaw, 2002.

4. Basic of microcontrollers

In 1952 Geoffrey Dummer, one of the people working for British Ministry of Defense propose integrating few electronics in one small housing as an independent system which would perform a certain function using an electrical signal. This is how the first integrated circuit came to be. It is worth mentioning that the first patent of an integrated circuit was issued to Texas Instruments in 1958, who today is one of the largest producers of integrated circuits, as well as signal processors. Producing of integrated systems led to the development of the first microprocessor (as a result of a mistake done by Intel) – that is a classical programmable system.

There is a multitude of integrated circuits. They are classified depending on their use. At the beginning, the SSI (Small Scale Integration) were produced, and they contained up to 100 transistors. Today, ULSI (Ultra Large Scale of Integration) systems are produced where the elements are smaller than 20nm and can work at 4GHz frequency.

Tab. 4.1. Development of processors

April 1972 Name of Processor: 8008 Clock speed: 200 kilohertz Number of transistors: 3,500	**May 1997** Name of Processor: Pentium II Clock speed: 300 MHz Number of transistors: 3.3 million
December 1974 Name of Processor: 8080 Clock speed: 2 MHz Number of transistors: 6,000	**October 1999** Name of Processor: Pentium III Clock speed: 733 MHz Number of transistors: 28 million
August 1976 Name of Processor: 8085	**November 2000** Name of Processor: Pentium 4

Clock speed: 5 MHz Number of transistors: 6,500	Clock speed: 1.5 GHz Number of transistors: 42 million
September 1978 Name of Processor: 8086 Clock speed: 10 MHz Number of transistors: 29,000	**July 2006** Name of Processor: Core 2 Duo Level 2 cache 4 MB Number of transistors: 253 million
October 1985 Name of Processor: 386 Clock speed: 16 MHz Number of transistors: 275,000	**Nov 2006** Name of Processor: Core 2 Extreme QX6700 Level 2 cache 8 MB Number of transistors: 582 million
March 1993 Name of Processor: Pentium Clock speed: 60 MHz Number of transistors: 3.1 million	**Dec 2009** Name of Processor: Larrabee Level 2 cache SECRET Number of transistors: SECRET

Signal processors are becoming very popular. Their function is to quickly process on a digital signals. The main difference between signal processor and a microcontroller is in the processing speed of information. Classical processors and signal processors operate on the discrete signal using the time and value domains. However, the processing speed of the signal processor enables it to perform mathematical calculations (like using a time frame or transformation of frequencies), based on which it is possible to make decisions in quasi real-time. Classical processors are able to make decisions only based on received bit or a word in a digital sense. From this derives another difference; conventional systems work using deterministic signals; however, signal processors are able to work with stochastic signals as well.

Fig. 4.1. Atmega8: most popular Atmel's microcontroller

4.1. Signal

The signal describes variation of any physical value, which may be described by using a mathematical function of one or several variables. The signal can be generated by biological, or technical objects. Signal usually carries information about the objects which generates it.
As a result, signal is determined by digital processing. Signals can be described by a mathematical model, but not through a physical model. The figure below shows several signal classifications.

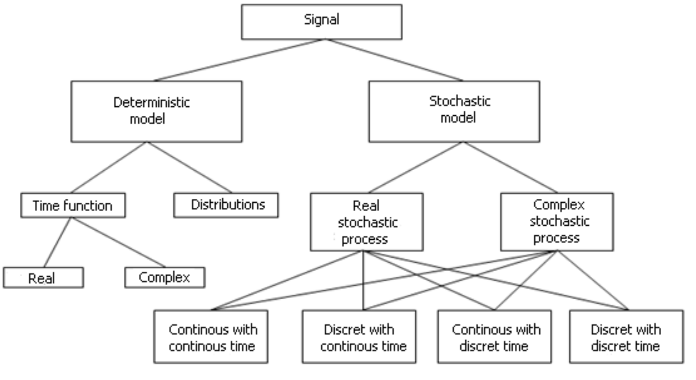


Fig. 4.2. Classification of signals [4]

Analog Signal

Analog signals are signals in the continuous time domain. These signals change their value with time and can have an infinite number of values.

Analog signal can be represented also in the frequency domain. In the case of periodic signals, it is possible to transform from the time domain to frequency domain. Signals that are made up by several harmonic functions are described in the frequency domain as harmonics.

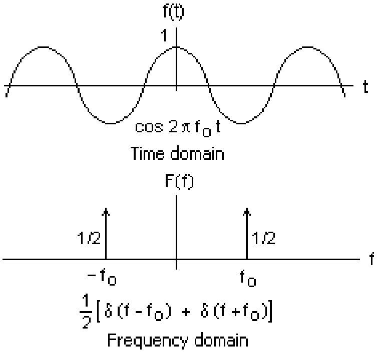


Fig. 4.3. Signal in time and frequency domain [4]

The transformation method from time to frequency domain was described by Jean Baptiste Joseph Fourier in 1822 in his work titled *„Théorie analytique de la chaleur"*. He proposed that stochastic signals should be described by a series, called Fourier series:

$$x(t) = a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)] \tag{1}$$

where: a_0 − constant coefficient,

$$a_0 = \frac{1}{T} \int_0^T x(t)dt \tag{2}$$

where: T – time period.

Coefficients (odd and even) are described as:

$$a_n = \frac{2}{T} \int_o^T x(t) \cos(n\omega_0 t)dt \tag{3}$$

$$b_n = \frac{2}{T}\int_{o}^{T} x(t)\sin(n\omega_0 t)dt \qquad (4)$$

Analog signal, being an energy source is described by deterministic models; however, signal being an information source is usually described by using stochastic models. Stochastic signal theory is harder to understand and analyze. This is due to the character of the signal, whose values one cannot predict.

Signals that are received by sensors in a vehicle are source of a deterministic signal in a properly operating object, whose emission characteristics are known (in the case of engine diagnostics). However, they may be a source of stochastic signals when one of the systems malfunctions, because it is impossible to predict engine vibrations that caused the malfunction.

Deterministic signals

Any integrated, real or time distributed function is a deterministic function if it has a limited energy or power.

First group of deterministic signals are time framed signals, which are part of signals that are not limited by time. Signal value that has an infinite time value is described by the following formula:

$$W_{\acute{S}R} = \lim_{\tau \to \infty} \frac{1}{2\tau}\int_{-\tau}^{\tau} x(t)dt \qquad (5)$$

where: τ - time frame.

In the case of periodic signals, the average value is described by:

$$W_{\acute{S}R} = \frac{1}{T}\int_{\tau}^{\tau+T} x(t)dt \qquad (6)$$

A large group of deterministic signals are impulse signals, which can have non-zero values only in certain periods of time. The average value of this type of signal, which occurs in the time frame from t_1 to t_2 is described by:

$$W_{ŚR} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t)dt \tag{7}$$

where: t_2 – time value at the end of measurement, t_1 – time value at the beginning of measurement.

Energy of deterministic signals is described by:

$$E = \int_{-\infty}^{\infty} x^2(t)dt \tag{8}$$

Average power of the deterministic signal is described by:

$$P_{ŚR} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x^2(t)dt \tag{9}$$

where: t_2 – time value at the end of measurement, t_1 – time value at the beginning of measurement.

The figure below shows an example of deterministic signal Sa^2.

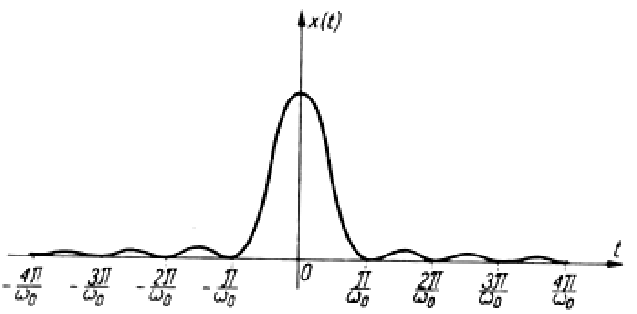


Fig. 4.4. Deterministic Signal Sa^2 *[4]*

Stochastic signal

Stochastic signal is a source of diagnostic information. Experience is the base of diagnostics. Probability is the mathematical model of experiments. It is characterized by such values as probability space, random variable, and distribution function, etc.

The classification of stochastic signals is based on time structure. If the signal is a set of unpredictable values limited by time, then it is called continuous time signal. However, if the signal is a set of predictable values then it is called a discrete time signal. All other signals, that at times can be limited by time as discrete, and others as continuous are called mixed signals.

If in the time frame, the value of the stochastic signal becomes a series of random variables, then the signal is defined as a continuous signal. If the value becomes discrete random variables, then the signal is defined as discrete.

The definitions mentioned above allow us to place stochastic signals into four categories:

1. Continuous signal of continuous time. An example of this type of signal can be sound generated by the engine,

Fig. 4.5. The example of analog signal [2]

2. Discrete signal of continuous time. An example of this signal is a telegraph. It is possible to describe these signals, that processing is based on signal processing,

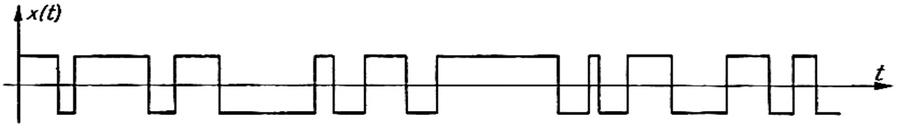


Fig. 4.6. The example of discrete signal of continuous time [2]

3. Continuous signal of discrete time. These are signals that are a result of sampling of continuous signals,

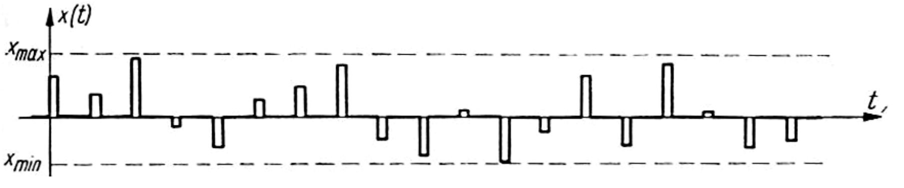


Fig. 4.7. The example of continuous signal of discrete time [2]

4. Discrete signals of discrete time. They are described as digital signals, a result of quantization of continuous signals of discrete time. This group of signals as a source of digital diagnostic information coming from the engine, will be described in the next chapter.

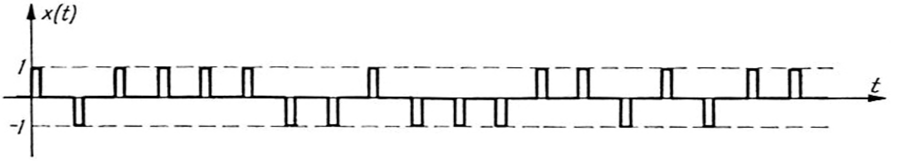


Fig. 4.8.The example of digital signal [2]

Digital Signal

Digital signal is described as discrete signal of discrete time. The transformation process from an analog signal to a digital signal is based on sampling of the analog signal, and then its quantization. The sum of these operations is described as an analog to digital transformation. Below is the schematic of the transformation (10).

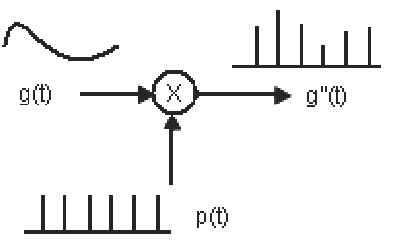


Fig. 4.9. Sampling as multiply [2]

$$g^{\#}(t) = g(t) \cdot p(t) = \sum_{n=-\infty}^{\infty} g(nT) \cdot \delta(t - nT) \qquad (10)$$

where: $g^{\#}(t)$ − sampled signal, g(t) – sampling signal, p(t) – continuous sampling.

Sampling is purely determined by its speed, which synonym is sampling frequency. It is described by the following:

$$f_S = \frac{\omega_S}{2\pi} = \frac{1}{T} \qquad (11)$$

where: f_S – sampling frequency.

There is no necessity to sample in the same time frames. Sometimes it is even required that the time frame is changing or randomly chased. There are also limits of the sampling method. It is not possible to sample in such a fashion that a reverse operation would return the same signal because this would require an infinite number of samples. Also, sampling causes an infinite number of duplicates of the signal, moved by $n\omega_{S.}$ By transforming a number of impulses into a Fourier series, by using formula 11, we were able to determine:

$$g^{\#}(t) = \frac{g(t)}{T}\sum_{n=-\infty}^{\infty}\exp(jn\frac{2\pi}{T}t) \qquad (12)$$

where: $^{\#}$(t) − sampled signal, g(t) – sampling signal, n – sample number.

The Laplace transform of the sampled signal, after using the transformation properties and after the transformation about the axis s=jω becomes:

$$G^{\#}(j\omega) = \frac{1}{T}\sum_{n=-\infty}^{\infty}G[j(\omega + n\omega_S)] \qquad (13)$$

where $G^{\#}$(jω) –Laplace transform of the sampled signal.

If the sampling frequency is too small compared to the bandwidth of the sampled signal, then it is possible for the signal to become indistinguishable (number of duplicates of the original signal cover one another). This is called aliasing. The figure below shows continuous spectrum (a) with a saw characteristic with coefficient of bandwidth <−B, B>. The figure (b) shows sampled spectrum, according to Whittaker – Nyquist – Kotelnikov – Shannon theorem. The theorem states that in order to properly reconstruct the signal from a discrete form, then it must come from a sampling frequency twice as large as half of Nyquist frequency (that is frequency on the edge of the bandwidth). The figure shows spectrum multiplication, with sampling frequency that is too small $\frac{1}{2}f_S<B$ (16).

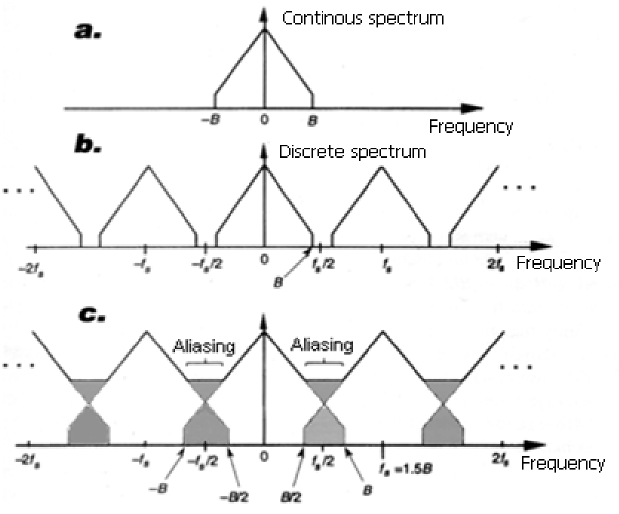


Fig. 4.10. The aliasing phenomenon [4]

The sampling process allows getting a discrete signal in time continuous signal. Also, with properly predetermined parameters, the sampling process is reversible, that is it allows us to get a continuous signal from a discrete signal. In order to receive a digital signal, that is discrete signal in time and value, we must quantize discrete signal in time. Quantization is not reversible. The definition of quantization is a division of signal values into predetermined ranges. These ranges are divided into:

1. Equally, linear quantization,
2. Not equally, non-linear quantization,
3. Not equally, logarithmically, logarithmic quantization.

Every range has a predetermined numeric value. Today, the binary code is used most commonly. In this case, the number of ranges is equal to:

$$N=2^n \qquad (14)$$

where: n – the word length in the binary code.

Longer words mean larger quantization resolution, which is determined by the dynamic properties of the process (every next bit enlarges the number of ranges by 2, thus the dynamic limits enlarges by 6dB when linear quantization is used).

Quantization range has normalized bandwidth, determined by:

$$Q = \frac{1}{N} \tag{15}$$

Thus, binary number, assigned to the range represents all values of the signal from the bandwidth ±*Q/2*. This means that various values from the continuous spectrum are represented by the same binary value. This is called quantization error, and it is semi random because is it correlated to the entry signal, and equally distributed by the probability distribution and has the value of:

$$p(\varepsilon) = \begin{cases} \frac{1}{Q} & dla \quad -\frac{Q}{2} < \varepsilon < \frac{Q}{2} \\ 0 & dla \quad \varepsilon \le -\frac{Q}{2} \;\wedge\; \varepsilon \ge \frac{Q}{2} \end{cases}$$

where: Q – quantization range, p(ε) – quantization error distribution.

Quantization error signal, which sums the original signal, is called quantization noise. The value of the quantization noise is the proportional inverse of the resolution, and is described by:

$$W_{RMS} = \sqrt{\int_{-Q/2}^{Q/2} \varepsilon^2 p(\varepsilon) d\varepsilon} = \frac{Q}{\sqrt{12}} = \frac{2^{-n}}{\sqrt{12}} \tag{17}$$

where: W_{RMS} – value of the quantization noise, Q – quantization range.

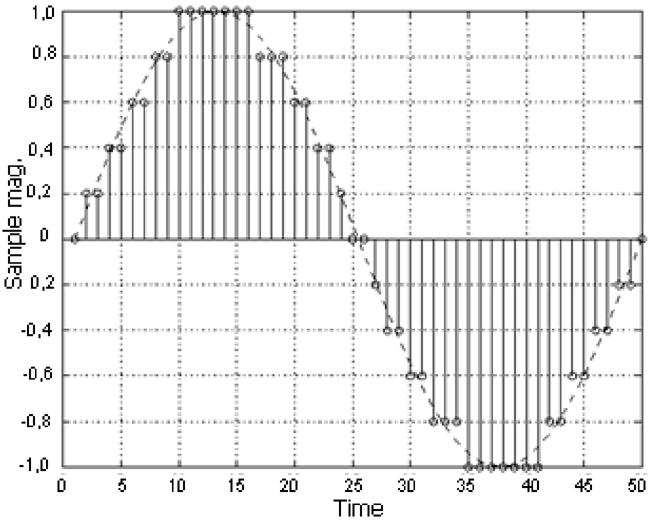


Fig. 4.11. Signal after sampling (with source)

PROBLEM

Independent work. Define the term "signal" as information

PROBLEM

Independent work. What are the differences between deterministic and stochastic signals?

PROBLEM

Group Work. List major differences between analog and digital signals.

4.2. Processor classification

Processors are classified based on several criteria. One was already mentioned at the beginning of the chapter. It was based on the number of microsystems inside the microprocessor. Another, popular classification deals with the length of the processed work (ex. 16, 32, 64, 128 bit). We

can also talk about classification based on the number of instructions from lower computer language class, which is used to program the given system. There are couple of main architectures: RISC (Reduced Instruction Set), CISC (Complex Instructions Set), VLIW (Very Long Instruction Word) and ZISC (Zero Instruction Set).
The RISC architecture is differentiated by a reduced number of instructions to minimum, simplified processor instructions and reduced address scheme. Most operations are performed based on the following:

$$\text{Register}_C = \text{Register}_A \textit{ (operation) } \text{Register}_B$$

Furthermore, RISC architecture is characterized by limited communication between memory and the processor. To send information between memory and registers instructions LOAD or STORE are used. Other instructions function only on registers according to the following scheme:

LOAD from a memory register

→ perform the function

→ write the result from register to memory

The number of registers has also been enlarged, thus the number of communications has been reduced.
Recently RISC microprocessors have been substantially developed. Their main advantage is easy configuration and faster speeds at the cost of larger architecture and more computing power. These processors are called SRP (Small RISC Processor)

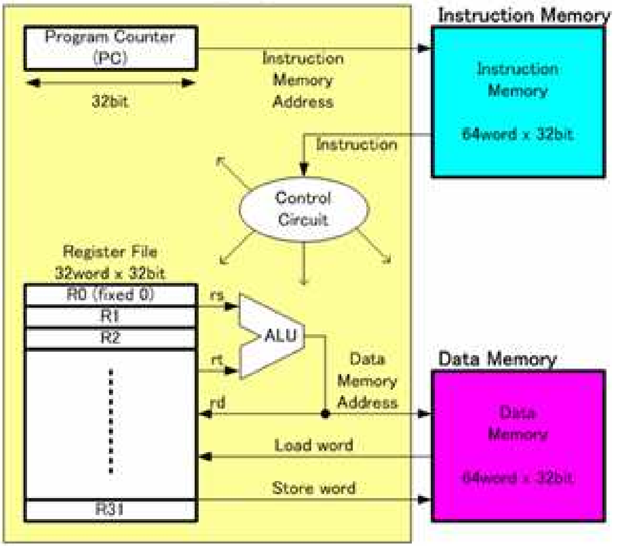


Fig. 4.12. The build of Small RISC Processor [1]

CISC architecture is reserved for computers with large computing power (such as PCs). Some of the characteristics of the architecture are:

1. Large number of instructions,
2. Lesser optimization than in RISC – which means that most of the instructions will not be completed in one cycle,
3. A long addressing scheme,
4. Specialized, complicated instructions set,
5. Ability to access memory by multiple instruction sets,
6. Lower (then RISC) takt processor core frequency,
7. Slower instructions decoding.

VLIW is much less popular architecture. It is characterized by:

1. Simplified drivers,
2. Increased number of functional units,
3. Out of order execution,
4. Compiler instructing the processor.

The functionality of VLIW processor is based on RISC (Except for the above mentioned differences). Compiler divides program into groups independent from each other and assigns them to specific units that it controls.

One of the first ZISC processors contained 36 independent cells. Each of them can be compared to an entry vector of 64 byte length similar to vectors stored in memory cells. If the entry vector is the same as vector stored in memory, the cell then “burns”. ZISC processors are produced to be used in neural networks.

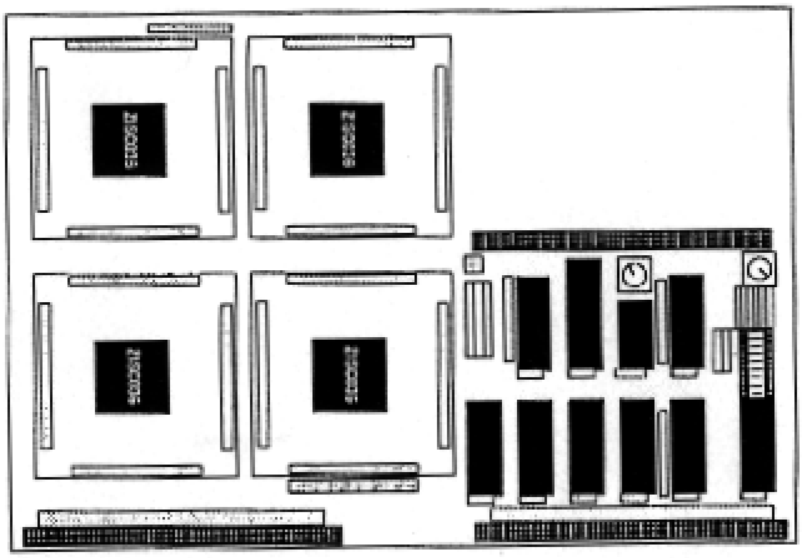
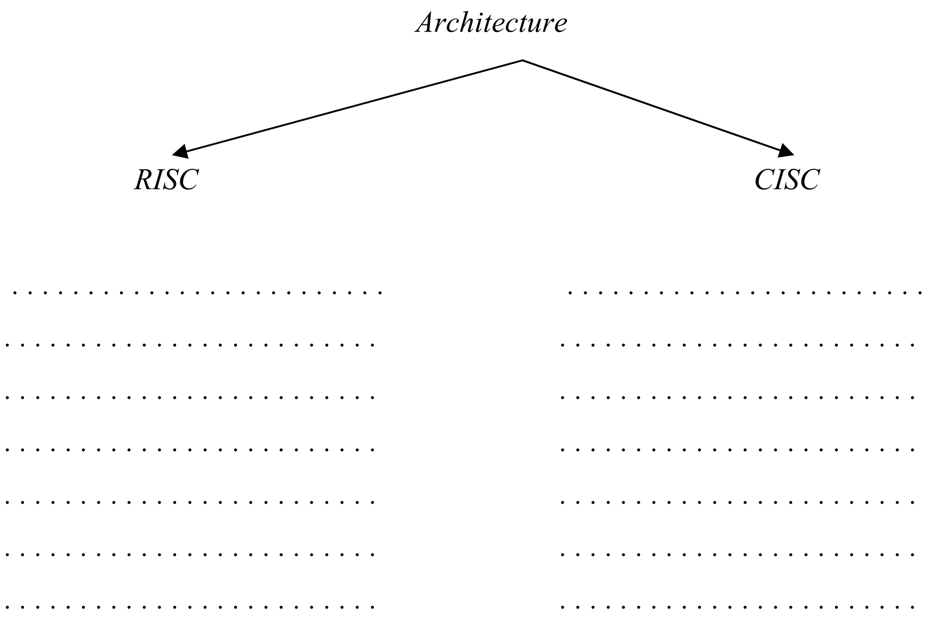


Fig. 4.13. The IBM ZISC036architecture [2]

PROBLEM

Independent Work. List differences between RISC and CISC architecture



Frequently processors are classified depending on the organization of their architecture. Two main ones are Von Neumann and Harvard architectures.

The von Neumann architecture is the classical computer design model. It is composed of three main blocks: CPU, memory and Input/Output devices. All three blocks are linked by means of the buses that act as paths for information interchange.

Its main characteristic is that it uses a single storage structure to hold both instructions and data, becoming the common referential model for sequential architectures. Due to its intrinsic organization it is not possible for the CPU to access simultaneously data and program information. The term "stored-program computer" is generally used to describe a computer of this design.

The consequence of the impossibility to apply parallelism to access memory is the low throughput (data transfer rate) between the CPU and memory, and this effect is known as the von Neumann "bottleneck".

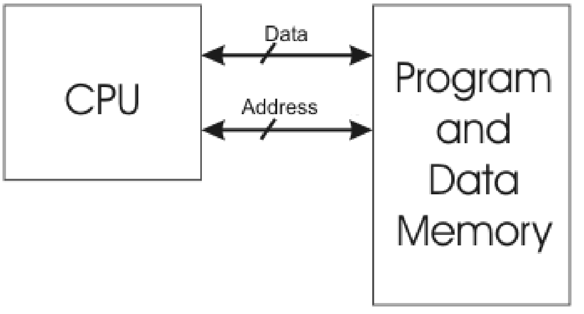


Fig. 4.14. Von Neumann architecture [2]

Harvard architecture physically separates data and program memory, providing different signal pathways for instructions and data. With this improvement it becomes possible for the CPU to access simultaneously program and data information and thus to implement parallelism.

In Harvard architecture, there is no need to make the two memories share characteristics. In particular, the word width, timing, implementation technology, and memory address structure can differ. Instruction memory is often wider than data memory. In some systems, instructions can be stored in read-only memory while data memory generally requires random-access memory. In some systems, there is much more instruction memory than data memory, so instruction addresses are much wider than data addresses.

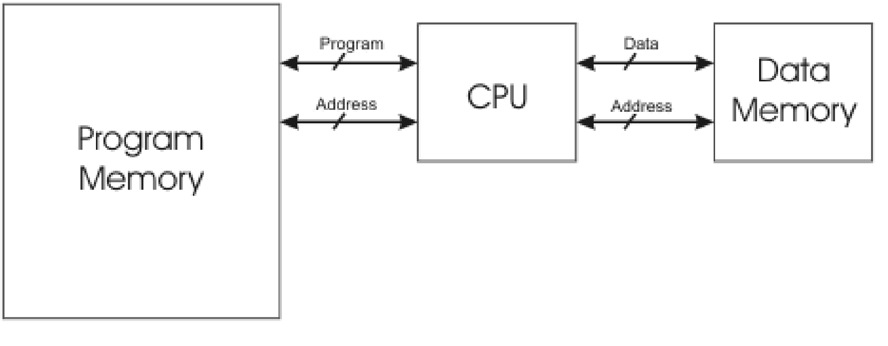
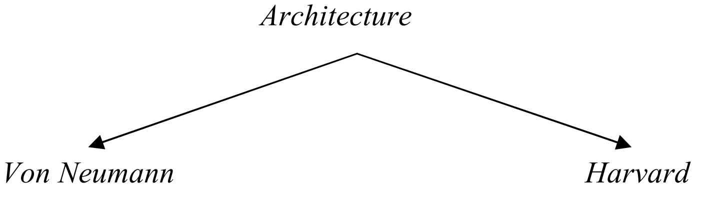


Fig. 4.15. Harvard architecture [2]

PROBLEM

Independent work. List differences between Harvard and Von Nuemann architecture.



Von Neumann	Harvard
. . . .	. . . .
. . . .	. . . .
. . . .	. . . .
. . . .	. . . .

4.3. Processor structure

The Central Processing Unit (CPU), or sometimes simply processor, is the component in a digital computer that interprets computer program instructions and processes data. CPUs provide the fundamental digital computer trait of programmability, and are one of the necessary components found in computers of any era, along with primary storage and input/output facilities.

A CPU that is manufactured as a single integrated circuit is usually known as a microprocessor. Beginning in the mid-1970s, microprocessors of ever-increasing complexity and power gradually supplanted other designs, and today the term "CPU" is usually applied to some type of microprocessor.

The fundamental operation of most CPUs, regardless of the physical form they take, is to execute a sequence of stored instructions called a program.

The program is represented by a series of numbers that are kept in some kind of computer memory. There are four steps that nearly all CPUs use in their operation: fetch, decode, execute, and write back.

Internally a CPU consists basically of a Control Unit, an Arithmetic ALU and a set of Registers. Control Unit is a finite state machine responsible to manage all the operations inside CPU. The arithmetic logic unit (ALU) is a digital circuit that calculates an arithmetic operation (like an addition, subtraction, etc.) and logic operations (like an Exclusive OR) between two numbers. The registers are a small set of memory used for temporary storage.

By far, the most complex electronic circuits are those that are built inside the chip of modern microprocessors like the Pentium. Therefore, these processors have inside them a powerful and very complex ALU. In fact, a modern microprocessor (or mainframe) may have multiple cores, each core with multiple execution units, each with multiple ALUs.

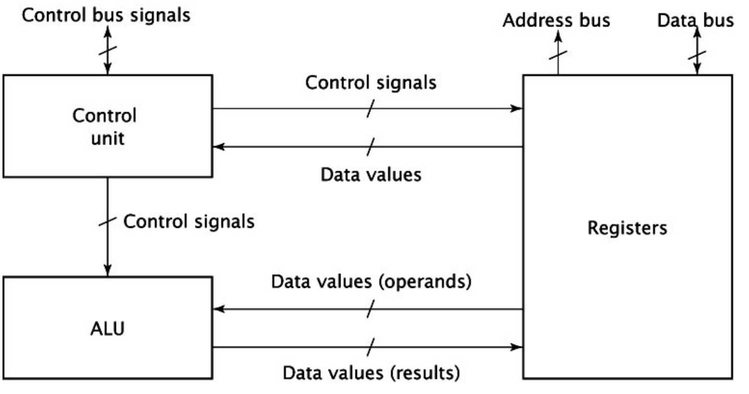


Fig. 4.16. CPU's organization [2]

PROBLEM

Independent work. List the most important elements of the CPU.

The function of the ALU is to perform simple operations on whole numbers (unlike Floating processing unit that performs operations on numbers with various decimal places). Other than strictly mathematical operations (performed on digital words) logical operations such as AND, OR, XOR are possible.

OR	0	1
0	0	1
1	1	1

AND	0	1
0	0	0
1	0	1

XOR	0	1
0	0	1
1	1	0

Fig. 4.17. Logical operations OR, AND and XOR

Modern processors contain several ALU units which can simultaneously perform several operations. It is worth mentioning that the basis of an operation, performed by an ALU, is addition; other operations derived from addition (for example, multiplication can be interpreted as addition several times over) require larger computing power.

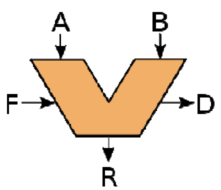


Fig. 4.18. ALU with operands (A, B), output (R), input (F) and output's status (D)

Registers are interpreted as a reserved part of memory dedicated for a specific task. There are registers for special and common tasks.

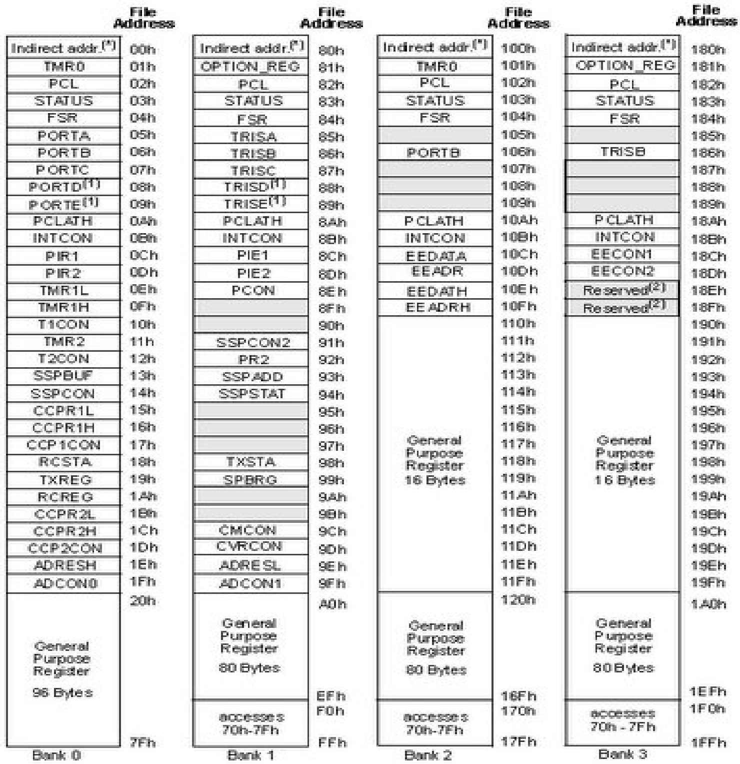


Fig. 4.19. The list of registers of PIC16F876A/877A processor

Registers located in memory have a predetermined length (8 / 16/ 32 / 64/ 128 bytes). There are upper and lower parts of the registers (frequently these parts function as independent registers). Specific bits in the registers called flags perform specific functions (for example, high or low bit value can execute writing to memory).

R/W-0	R/W-0	R/W-0	R-1	R-1	R/W-x	R/W-x	R/W-x
IRP	RP1	RP0	$\overline{TO}$	$\overline{PD}$	Z	DC	C
bit 7							bit 0

Fig. 4.20. Status register of PIC16F876A/877A processor

Registers are special memory banks where operational instructions or information dealing with the status of the processor are stored. Processor (or computer) can have several types of memory.

Random access memory (RAM) chips come in a wide variety of organizations and types. Computer main memories are organized into random addressable words in which the word length is fixed to some power-of-2 bits (for example, 4, 8, 16, 32, or 64 bits). The main memory design consists of a stack of chips in parallel with the number of chips corresponding to that machine's word length. The volatile memory stores the non permanent data that results from the algorithms operations. On the other hand, the non-volatile memories (ROM, EPROM, FLASH) usually are used to permanently store the instructions and fixed parametric data.

The storage capacity of the memory system depends on the width of the address and data bus. For instance a system with an address bus of 16 bits wide could address up to $2^{16} = 65536$ memory locations. In every memory location can store as many bits as the width of data bus. Usual values for the data bus width are 8, 16 or 32.

The communication between CPU and memory system are performed by means of read and write operations. The process of reading data from memory is usually referred to as "loading" data from memory into the processor, meanwhile the process of writing data to memory is usually referred to as storing data to memory from the processor.

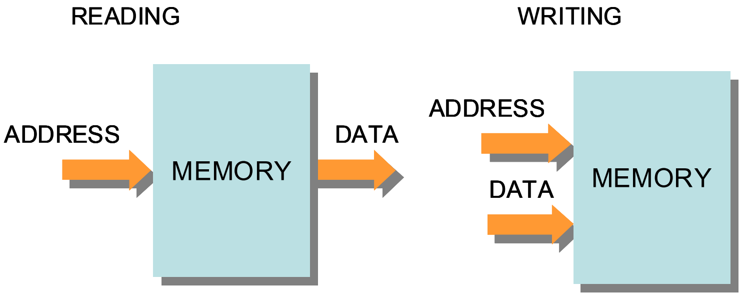


Fig. 4.21. Reading and writing to memory [1]

4.4. Programming Languages

A programming language is a formalized set of rules used to describe functions that a program must perform. Formalization means both context and definitions of a certain set of instructions with a defined purpose.

The list of instructions of a typical microprocessor is very short and contains basic mathematical and logical functions. Because writing instructions with a limited list of instructions would take up a lot of time, programming languages were developed (known as higher level languages), containing instructions that combine several microprocessor directives. They allow people not familiar with computer architecture using predetermined structures, as well as offer several other simplifications. Languages written in other forms than assembler (lower level language) are translated into machine's code (complied), which is understood by the processor.

Higher level languages are useful in writing programs for everyday use, where the speed of processing is not important and the program is used for many functions. There are new programming languages written for the purpose of education or for their own sake, such as Pascal or Delphi. Due to their simplicity, these languages, such as BASIC, are very attractive for people trying to become familiar with computer technology; however, they are not particularly useful for practical uses. Also, these languages are used in small microcontrollers, whose function is to perform simple operations (for example, lighting a light on the dashboard or reacting to an entry signal).

Source code written in assemblers language contains:

1. Constants (signals or numbers),
2. Instructions written in a proper format,
3. Directions,
4. Etiquette,
5. Comments.

Language format is very specific. Every series of signs containing letters, numbers, and the underline "_" starting with a letter or underline is considered a statement. Those statements are commonly used to differentiate various parts of the program (variable names, procedures, etiquette). Numbers are statements containing digits (as well as letters from "a" do "f" in the case of hexadecimal systems). They can be written in decimal, hexadecimal or binary systems.

Constants are values that are used during the performance of the program to execute operations (for example, they define pi as 3.14 when it is needed).

Instructions contain symbolic statements with their arguments. Generally they can be written as:

Etiquette: instructions [argument] ; comment

Instruction (with its possible arguments) is a symbolic statement that must be performed by a microcontroller. In case of use of predefined macros, compiler translates them into series of instructions for the microprocessor according to its definition. Etiquette is a symbolic statement defining location in program (for example, as a location where to skip), and an argument is a value on which a certain operation must be performed (they can be variables) and correlated with the instructions.

Directives in the assembler's language perform two functions; they are statements controlling the work of the compiler and define other elements of the program (such as constants, macors, etc). For example directive BYTE:

Etiquette: .BYTE statement ; comment

Tab. 4.2. The list of arithmetic and logic instructions for ATMEL microprocessors

Mnemonics	Operands	Description	Operation	Flags	#Clock Note
ARITHMETIC AND LOGIC INSTRUCTIONS					
ADD	Rd, Rr	Add without Carry	Rd ← Rd + Rr	Z,C,N,V,H	1
ADC	Rd, Rr	Add with Carry	Rd ← Rd + Rr + C	Z,C,N,V,H	1
ADIW	Rd, K	Add Immediate to Word	Rd+1:Rd ← Rd+1:Rd + K	Z,C,N,V	2
SUB	Rd, Rr	Subtract without Carry	Rd ← Rd - Rr	Z,C,N,V,H	1
SUBI	Rd, K	Subtract Immediate	Rd ← Rd - K	Z,C,N,V,H	1
SBC	Rd, Rr	Subtract with Carry	Rd ← Rd - Rr - C	Z,C,N,V,H	1
SBCI	Rd, K	Subtract Immediate with Carry	Rd ← Rd - K - C	Z,C,N,V,H	1
SBIW	Rd, K	Subtract Immediate from Word	Rd+1:Rd ← Rd+1:Rd - K	Z,C,N,V	2
AND	Rd, Rr	Logical AND	Rd ← Rd • Rr	Z,N,V	1
ANDI	Rd, K	Logical AND with Immediate	Rd ← Rd • K	Z,N,V	1
OR	Rd, Rr	Logical OR	Rd ← Rd v Rr	Z,N,V	1
ORI	Rd, K	Logical OR with Immediate	Rd ← Rd v K	Z,N,V	1
EOR	Rd, Rr	Exclusive OR	Rd ← Rd ⊕ Rr	Z,N,V	1
COM	Rd	One's Complement	Rd ← $FF - Rd	Z,C,N,V	1
NEG	Rd	Two's Complement	Rd ← $00 - Rd	Z,C,N,V,H	1
SBR	Rd,K	Set Bit(s) in Register	Rd ← Rd v K	Z,N,V	1
CBR	Rd,K	Clear Bit(s) in Register	Rd ← Rd • ($FFh - K)	Z,N,V	1
INC	Rd	Increment	Rd ← Rd + 1	Z,N,V	1
DEC	Rd	Decrement	Rd ← Rd - 1	Z,N,V	1
TST	Rd	Test for Zero or Minus	Rd ← Rd • Rd	Z,N,V	1
CLR	Rd	Clear Register	Rd ← Rd ⊕ Rd	Z,N,V	1
SER	Rd	Set Register	Rd ← $FF	None	1
MUL	Rd,Rr	Multiply Unsigned	R1, R0 ← Rd × Rr	C	2 (1)

4.5. Applications

Processors are used in every new vehicle system. It is very interesting to look at the correlation between fuel consumption, emissions, and... word length of specific (microprocessors) systems. Some of the research was performed by Infineon Technologies

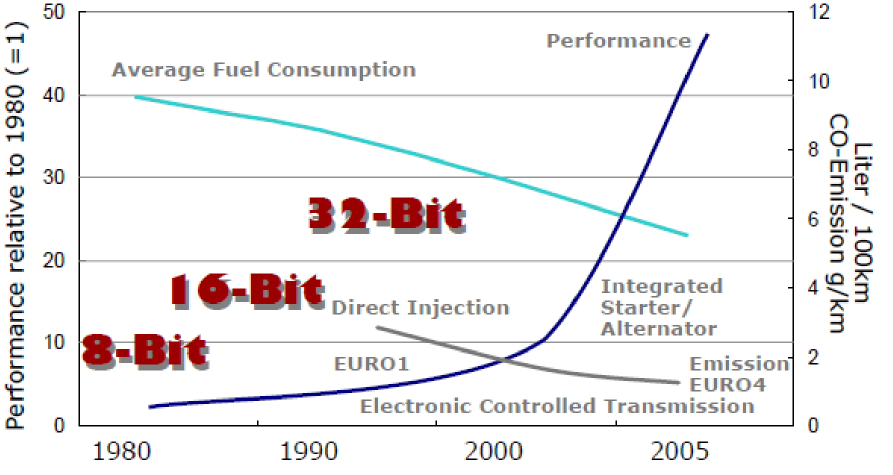


Fig. 4.22. Microprocessors word length vs. emission and fuel consumption [1]

Results from the research show that the number of functions performed is proportional to the word length of the used processor. As the number of functions and driver and passenger requirements grow, we can expect the number of processors in a vehicle to grow as well, as well as wider use of signal processors. Collision avoidance systems, road sign reading systems, and parking assistance systems are becoming a standard. However, we must remember that we are only in the **dawn of the microprocessing in automotive era.**

References

[1] Informator techniczny Bosch. *Czujniki w pojazdach samochodowych*. Wydawnictwa Komunikacji i Łączności, Warszawa 2002.

[2] Stranneby D. *Cyfrowe przetwarzanie sygnałów. Metody, algorytmy, zastosowanie*, Wydawnictwo BTC, Warszawa 2004.

[3] Doliński J.: *AVR w praktyce*, Wydawnictwo BTC, Warszawa 2003.

[4] Szabatin J. *Podstawy teorii sygnałów,* WKŁ, Warszawa 2004.

5. IP networks

Today, nobody can imagine the world without the internet – the main source of information for many people. It's very hard to describe all aspects of the network based on the internet protocol in one chapter. Therefore, one should treat this chapter as a complete propedeutics of the network, not as an encyclopedia.

The chapter is divided into three parts. The first one, titled „The architecture of the web", talks not only about the topology of the web, but mainly about its physical layout, with all its signal parameters. Those parameters can also be used in other fast networks i.e. vehicle's CAN bus.

Subchapter „Protocols" describes the necessity of having protocols. It talks about the most important ones, and also gives the example of the account of subnet address. This chapter also contains a few problems. Solving them will help YOU understand the schematic of how the network works.

The last part of the chapter, titled „Services", is the summary of most widely used services, based on protocols of different layers of the web design. The main focus was set on the services „unseen" for the average user, yet necessary in network communication.

5.1. The architecture of the web

There are a few things necessary to create a network. The physical and logical connections are the two most important ones. Physically, the connection is made possible through network adapters, connectors etc. The logical connection is based on previously established rules, known to both parts being connected. These rules, that are the language of communication, are named PROTOCOLS. The following subchapter talks more about them.

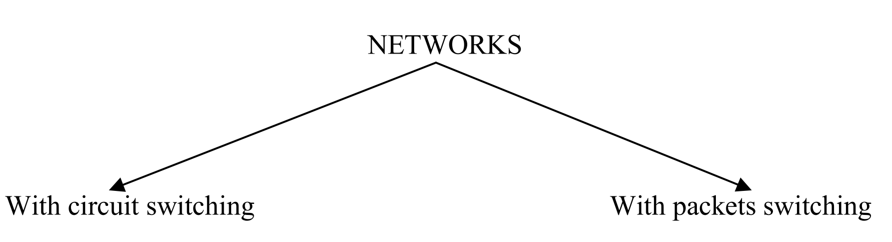
Historically, the network can be divided into two groups, based on the way the information is being supplied to the network:

1. Network with circuit switching – the connection is made while all hosts are identified. A temporary path or a circuit is created through a number of switches. If any of the connections or devices fail, the communication is stopped. To reconnect, you have to start the communication process all over again. In the end, the new circuit between the source and the destination is created,
2. Network with packet switching – the devices in this type of network are not „aware" of what's in the packets. They can only see the destination address and the following device on the path to the destination (it doesn't apply to the advanced routing protocols). There is no circuit between the source, and the destination. Every packet is being sent independently from one switching device to another. This process is called ROUTING.

The networks with circuit switching are being replaced by the networks with packet switching, and this is the main way of communication in the modern diagnostic channels.

PROBLEM

Work performed in groups of two. Name three characteristics of networks with circuit and packets switching. Show main differences between them, and determine which one is a better solution.



. . . .

PROBLEM 5.1.2

Work performed in groups of two. The network security is one of the most important and still growing subjects. Name 5 most important dangers that deal with network communication (not only internet). Compare to the information found on www.sans.org

1. . . . . 2. . . . .
3. . . . . 4. . . . .
5. . . . .

5.2. Network classification

Nowadays, one of the main ways of the network classification is based on its extension. This seems to be a mistake. The classification into local area network (LAN), and wide area network (WAN) based on their size is inadequate. A better solution would be to divide them based on the number of routers or hosts. In the diagnostic network that we see i.e. in motor vehicles, where all the information is sent to the decision unit, or straight to the controllers, the classification based on the number of detectors is unnecessary.

The other way of network classification is based on its topology. There are a few different types of topology, e.g.:

1. Star (and extended star),
2. Bus,
3. Ring (and double ring),
4. Mesh,
5. Hierarchic.

Today, the most commonly used topology is the star. We can see other topologies (i.e. bus) in the mechatronic's usage. It is possible, that the star topology can also be the hierarchic one, where central device is also the decision unit. To make the network more reliable we can create alternate excess paths. This can transform the star into the mesh topology. The excess parameter (tells us about the existence of the alternative routes), and the scalability parameter (tells us about the possibility to make the network larger) are the most appreciated.

PROBLEM 5.1.3

Independent work. Please draw the mesh and bus topology in the areas below. Write down three main parameters characterizing these topologies. Use the internet.

Star topology

1. . . . .
2. . . . .
3. . . . .

Bus topology

1. . . . .
2. . . . .
3. . . . .

5.3. Teleinformatic and telematic systems

Teleinformatic system is the key element in the existence of a network. Nowadays, the wiring is always taken into consideration while realizing projects both in the automotive, and construction industry. The manufacturers have been working for years on creating a new type of the most popular cable - the twisted-pair cable that replaced the wiring for certain systems, and is commonly used today. Despite the fact that the twisted-pair cable, especially its unshielded version is best known for its use in the computer networks, we should remember, that the shielded version of the twisted-pair cable is widely used in the vehicle's sensor and system buses.

The overwhelming popularity of the UTP cable was linked to the publication, released in 1991 by EIA/TIA - Electronic Industries Association/Telecommunication Industry Association. The publication, titled EIA/TIA-586, describes the electrical and mechanical specifications of the wiring used to build a teleinformatic network. The EIA/TIA-586 standardizes the connectors and cables with transmission characteristics up to 16MHz. The document is followed up by 4 bulletins:

1. TSB 36 – specifies cables with transmission characteristics up to 100 MHz,
2. TSB 40 – specifies connectors with transmission characteristics up to 100 MHz,
3. TSB 40A – specifies additional requirements for cross-connect jumpers and for cables with transmission characteristics up to 100 MHz,
4. SP-2840 – project that specifies cables and connectors with transmission characteristics up to 100 MHz.

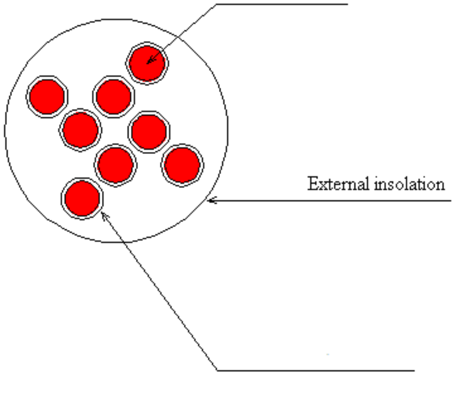
Copper transition media can be classified into unshielded and shielded ones. The placement of the wiring determines the kind of cables that should be used. The shielded cables are used in the environments, where electromagnetic interference occurs. The thickness of the shield depends on the strength of the interference, which is relatively large in motor vehicles. The other solution is to use multimode fiber optic cables. These cables are more commonly used while designing the bus system of new vehicle's and telematic systems.

PROBLEM 5.1.4

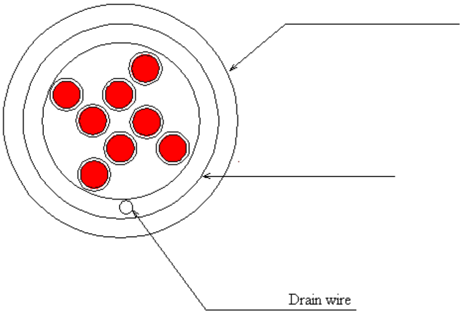
Independent work. From the definitions given below, choose the right one, and put them in the correct spaces in the illustrations.

Insulation, Aluminum foil shield, Outside insulation, Conductor, Drain wire, Copper braid, Polyester foil

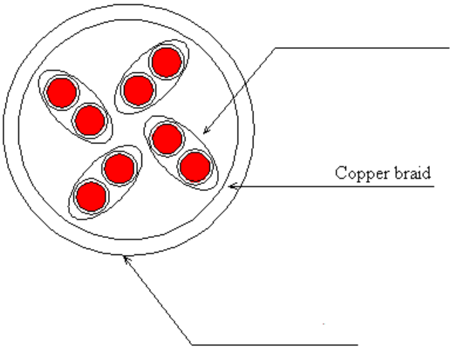
UTP cable



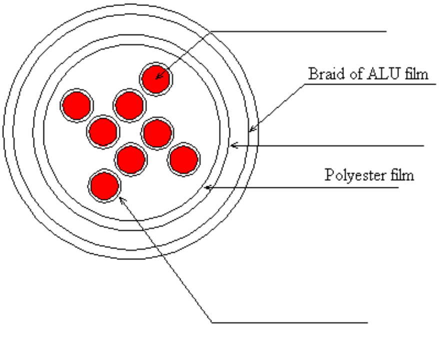
FTP cable



STP cable



SFTP cable



PROBLEM 5.1.5

Work in groups of three. Create the logical and physical wiring layout of one of the classrooms chosen by the teacher.

5.4. Protocols

Network protocols are the systematized way of communication. Systematization can be based on standardization of the entire, or some parts of communication. Some protocols are known to be popular despite the lack of their standardization. The performance of network protocols is described in the broadly available RFC documents. These documents contain technical specifications and rules that allow the use of certain protocols.

PROBLEM

Independent work. Find the RFC (791) document describing the IP protocol. Name 5 higher level protocols associated with IP. Describe their function.

In general, the functions of protocols are:

1. Establishing the format of message,
2. Establishing the structure of message,
3. Defining the way of delivering of the message,
4. Defining the time of communication,
5. Beginning and ending the communication.

Layers

Protocol stacks (i.e. TCP/IP) are part of IP networks. The definition TCP/IP can describe both: the connection of TCP and IP protocols, and the layering protocols based on TCP and IP. TCP/IP is also known as a synonym of the layering model describing DoD network.

OSI/ISO is the most popular layering model describing the performance of the network. It has 7 layers, responsible for different aspects of network communication. We need to remember that OSI/ISO, unlike the Dod network, is purely virtual and has no physical aspects in the network. It allows us to design protocols and helps us understand how the network works.

Every ISO layer has its function:

1. Physical – describes all necessary items to create a communication,
2. Data link – contains protocols necessary to communicate in the same transition media,
3. Network – responsible for physical addressing, allowing sending parts of messages through the network,
4. Transport – responsible for providing the information and flow control – putting the separated parts of transmissions together,
5. Session – begins and ends the session,
6. Presentation - responsible for switch in data representation, necessary for Application layer,
7. Application – makes the communication between a computer and person possible.

PROBLEM

Work in groups of two. Assign the definitions below to the right layer.

Connector, Web browser, IP Address, Voltage, MAC Address, Beginning of session, Skype, Data transfer, TCP, format cange, Cable, Network Interface Card.

Physical: . . . .

. . . .

. . . .

. . . .

. . . .

Data link: . . . .

Network: . . . .

Transport: . . . .

. . . .

Session: . . . .

Presentation: . . . .

Application: . . . .

. . . .

PROBLEM

Independent work. Assign the protocol's elements to the right layer.

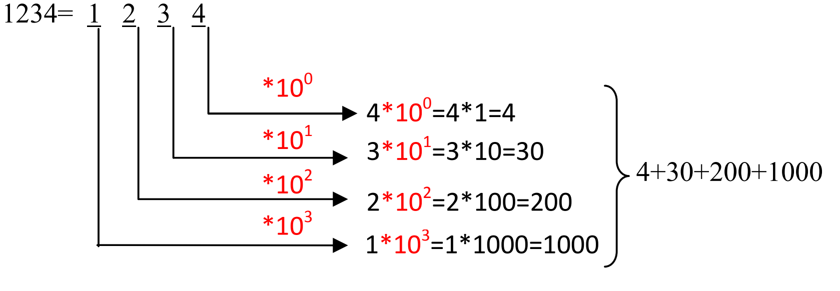
Data, Bits, Data, Data, Packets, Segments, Frames

Positional notation

Word „positional" means, that the order we put digits in, to create a number, matters i.e.

$$1234 \neq 4321$$

Position that may seem obvious for a person in its natural notation can be understood differently in the digital world. That's why expressions like „Little Indian" and „Big Indian" are so commonly used. They show the order of placing the binary digits (and most importantly – of their future interpretation). The digits can create i.e. a confident diagnosis. The order of the digit defines its value:



There are two more numeral systems used in telematic, mechatronic, and computer network: binary and hexadecimal. They are both based on positional notation.

Binary numeral system is more popular, especially when it comes to 0-1 notation. This notation is often interpreted either as a presence or lack of signal. We should remember that the "lack of signal" is not represented by "0".

One of the methods of converting decimal to binary system is by dividing the number by two with the remainder.

/2	Remainder:
1234	0
617	1
308	0
154	0
77	1
38	0
19	1
9	1
4	0
2	0
1	1
0	

$1234 = 10011010010_2 = 00000100\ 11010010_2$

The final numeral was completed by few zeros, in order to get a two complete octet result. If we missed that last step, the protocols can interpret that as an unfinished action = error.

PROBLEM

Independent work. Convert the numbers given below to binary system by dividing with the reminder. Complete the final numeral to octets.

/2	*Remainder:*
667	

/2	*Remainder:*
985	

667 = ...$_2$ *985 =* . .. .$_2$

Hexadecimal numeral system is also widely used in networks. Particularly, this notation pertains to memory cells and error codes that are read in the CAN bus systems. This is a consequence of a large number of specified values that would cause the notation to be extremely long.

Hexadecimal positional numeral system uses 16 symbols: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F. Letters represent numbers 10 to 15. The conversion from binary to hexadecimal numeral system is easy. The final octets are divided into four digit groups (quartets). Each group is converted independently.

$1234 = 10011010010_2 = 0000\ 0100\ \ 1101\ 0010_2$

$0010_2 = 2 = 2_{16}$

$1101_2 = 13 = D_{16}$

$0100_2 = 4 = 4_{16}$

$0000_2 = 0 = 0_{16}$

=> $1234 = 10011010010_2 =$ **$04D2_{16}$**

PROBLEM

Convert the numbers into binary and hexadecimal numeral systems. Complete the final numeral to octets. Work independently; discuss the results with the group.

254 = ...$_2$ = ...$_{16}$

255 = ...$_2$ = ...$_{16}$

78 = ...$_2$ = ...$_{16}$

$1024 = \dots\dots\dots\dots\dots\dots\dots\dots_{2} = \dots\dots\dots\dots_{16}$

5.5. Internet Protocol version 4 (IPv4) addressing

Every layer of the OSI model is associated with specific address. Data link layer contains physical addresses of certain devices, network layer has their logical addresses, and the transport layer contains port numbers that identify services, which in a way are addresses as well. The process of encapsulation and decapsulation consist of adding, removing (and analyzing) addresses.

IPv4 addressing consists of coding information about sub network (and sub networks), hosts, and broadcasting, in every address. Every address has 32 bits divided in four octets. The addresses are divided into classes A, B, C, D, E. Value of the first octet determines the address's class.

Tab. 5.1. Classes of IPv4 addresses

Class	1st octet in binary system	Real range of address in decimal system
A	00000000_2 - 011111111_2	1.0.0.0 – 127.255.255.255
B	10000000_2 - 10111111_2	128.0.0.0 – 191.255.255.255
C	11000000_2 - 11011111_2	192.0.0.0 – 223.255.255.255

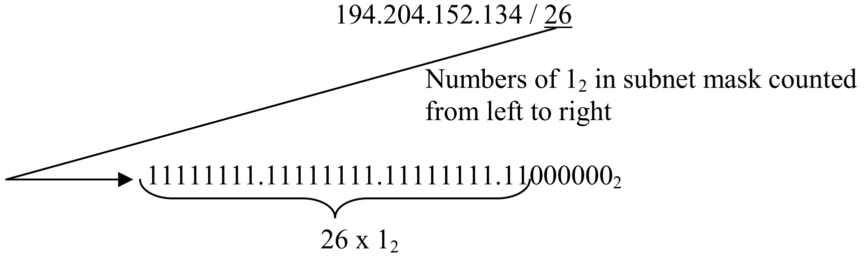
The address's classes contain other classes as well. These ones are reserved for specific processes (i.e. the process of group communication of routing protocol). Every class has a default subnet mask added to it:

1. Class A: 255.0.0.0,
2. Class B: 255.255.0.0,
3. Class C: 255.255.255.0.

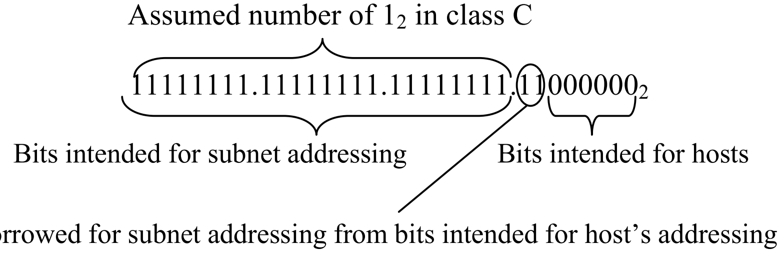
Subnet mask is the most important part of network address. It informs us about numbers of hosts in subnet, numbers of subnets, and also allows us to determine subnet's address or broadcasting address (or broadcast). Subnet mask has to be coherent, which means that we can't have 1_2 preceded by 0_2. It depresses the numbers in the octets of subnet mask.

$10000000_2 = 128$	$11111000_2 = 248$
$11000000_2 = 192$	$11111100_2 = 252$
$11100000_2 = 224$	$11111110_2 = 254$
$11110000_2 = 240$	$11111111_2 = 255$

A very popular way of writing down the subnet mask is CIDR methodology (Classless Inter-Domain Routing). CIDR notation uses the base address of the network followed by a slash i.e.



Address 194.204.152.34 belongs to class C. Default subnet mask for this class is 255.255.255.0, which means the CIDR notation is: /24. In this case, two bits assigned for host's addressing were borrowed for subnet addressing:



If two bits are borrowed for subnet addressing, it means that we can address 2^2 subnets: 00_2, 01_2, 10_2, 11_2. The last octet of address 194.204.152.134 becomes **10**000110_2 binary. Marked values borrowed for subnet addressing show that it is 3rd subnet in a row.

Logical multiplication of address and mask will result in subnet address that includes host:

$$
\begin{array}{rl}
194.204.152.134 = & 11000010.11001100.10011000.10000110_2 \\
255.255.255.192 = /26 = & 11111111.11111111.11111111.11000000_2 \\
\text{(and)} & \overline{} \\
 & 11000010.11001100.10011000.10000000_2 = \\
 & = \mathbf{194.204.152.128}
\end{array}
$$

It is not necessary to convert all address octets and mask into binary numeral system. The conversion in address value comparing to hosts address can only be seen in the mask octet that contains 0_2.

The number of 0_2 in subnet mask tells us about the number of hosts in the subnet. In this case (/26) there are 6 logical 0, which allow addressing 2^6-2 = 64-2 = 62 hosts in every subnet. The necessity to subtract two addresses is the result of designation these addresses for the subnet address (the first address in a set), and broadcast (the last address in a set). Setting the broadcasting address is done by putting in the subnet address (in the bits assigned for host's addressing) 1_2. In subnet 194.204.152.128/26 we'll have 6 binary 1 written down as followed:

Subnet address: $11000010.11001100.10011000.10\underbrace{000000}_2$

Bits assigned for host's addressing.

By rewriting the last octet and putting 1_2 in place of bits assigned for host's addressing, we'll get the following broadcasting address (broadcast):

Broadcast:

11000010.11001100.10011000.10**111111**$_2$=194.204.152.191

Summarizing, address 194.204.152.134/26 gives following information:

1. Class C,
2. Default mask class: 255.255.255.0,
3. Mask appeared from the entry: 255.255.255.192,
4. Number of bits borrowed for subnet's addressing: 2,
5. Maximum number of subnets possible for addressing: : $2^2 = 4$,
6. Remaining number of bits for host's addressing: 6,
7. Maximum number of hosts in every subnet: $2^6-2 = 62$,
8. Subnet's address: 194.204.152.128,
9. Broadcast: 194.204.152.191.

PROBLEM

Work in groups of two. Write down the following information based on given address: 192.168.60.35/25.

- *Class:* ....
- *Default mask class:* ....
- *Mask appeared from the entry:* ....
- *Number of bits borrowed for subnet's addressing:* ....
- *Maximum number of subnets possible for addressing: :* ....
- *Remaining number of bits for host's addressing:* ....
- *Maximum number of hosts in every subnet:* ....
- *Subnet's address:* ....
- *Broadcast:* ....

PROBLEM

Independent work. Find all the information as requested in PROBLEM 5.2.6 for the following address 172.16.16.84/22. Discuss the results with the group.

PROBLEM Work in groups of two. Check which sets of addresses are reserved for private network's addressing. Determine the maximum number of hosts for every set.

Class A: . . . . - . . . . ,
maximum # of hosts. . . . . . .
Class B: . . . . - . . . . ,
maximum # of hosts. . . . . . .
Class C: . . . . - . . . . ,
maximum # of hosts. . . . . . .

5.6. Services

Protocols that allow the process of communication and information exchange can also be used as industrial protocols. These protocols are standardized by various international organizations. The most popular ones are Internet Engineering Task Force (IETF) and Institute of Electrical and Electronics Engineers (IEEE).

Network frames have places signed as "port numbers". Those ports are designated for identification services connected with protocols. One of the examples is http – reserved for port 80. Some of the ports characterize the services they are assigned to. These ports are also known as "commonly known" ports from 1 -1023. Ports ranged from 1024 to 49151 are registered and used by host's processes. Higher port's numbers (49152 – 65535) are called dynamic ports and are not assigned for particular applications.

The existence of protocols allows direct identification of a service. It also allows blocking the particular service by making access list on the router showing the port's range or number.

PROBLEM

Independent work. Go to Internet Assigned Numbers Authority website. Find the following port numbers and their assigned services.

Port number	Service
20, 21	
23	
37	
80	
118	
161	
995	

The services and their protocols have to be independent from implementation or platform they work on. It means that HTTP service, started on host A has to be identified with the same service started on host B even if this computer has a different operating system.

The services can be divided into two groups:

1. Network services; mainly used by corporations. They are used to create advanced, diffused systems that have a particular function. It is hard to use them in the International Network (Internet), since they’re created for private network,
2. Internet Services; used in the International Network. At first they were a part of network services, which became very popular. A few examples are: online banking, RSS channels, Internet Relay Chat (IRC), GG, email, etc.

PROBLEM

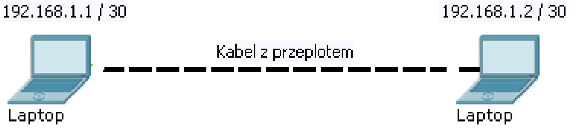
Work in groups of two. Define the following expressions and name their function.

Service	Definition	Funcion
Blog		
IRC		
Online banking		
SKYPE		
Facebook		

eMule		

PROBLEM

Work in groups of two. Configure the network according to the following picture. Allow downloading one of the files from laptop A to laptop B. Repeat the downloading from laptop B to laptop A.



References

[1] Amato V., Lewis W. *Akademia Sieci Cisco. Pierwszy Rok Nauki*. Mikom, Warszawa 2009.

[2] Dunsmore B., Skandier T. *Cisco. Technologie Telekomunikacyjne*. Mikom, Warszawa 2009.

[3] Sportack M. A. *Podstawy Adresowania IP.* Mikom, Warszawa 2008.

[4] Bradford R., *Podstawy Sieci Komputerowych*. WKiŁ, Warszawa 2009.

6. Controlled Area Network (CAN)

The first functional model of bus was introduced at the SAE (*Society of Automotive Engineers*) conference in 1986. It was a product of Robert Bosch's laboratories. In the following years CAN bus based system revolutionized the automotive industry. Despite the definite CAN bus's monopoly on the automotive market, new, revolutionary, competitive systems were introduced, i.e.: InterBus-S, Profibus, FIP, EIB, LON. Those solutions are quite innovative and are already used in different engineering fields. The explanation to why these systems are still not as popular in automotive industry as CAN is, can be narrowed to these three reasons:

1. Late patent of the standard,
2. Higher price dictated by the lack of various specialized systems,
3. Influence of big automotive groups on blocking the new system's expansion.

The primary specification of CAN was introduced by BOSCH. There are two versions of that system:

1. Standard CAN (SCAN), version 2.0A; uses 11 identification bits in frame's header,
2. Extended CAN (ECAN), version 2.0B; uses 29 identification bits in frame's header.

There are two ISO standards for CAN. They vary depending on bandwidth. ISO 11898 applies to fast applications (up to 1 Mbps). ISO 11519 applies to applications up to 125 kbps. Tables below show typical voltage levels in CAN buses.

Tab. 6.1. The voltage levels of signal in CAN ISO11898

Signal		Reverse state			Dominant state	
	min	nominally	max	min	nominally	max
CAN-High	2,0V	2,5V	3.0V	2,75V	3,5V	4,5V
CAN-Low	2,0V	2,5V	3.0V	0,5V	1,5V	2,25V

Tab. 6.2. The voltage levels of signal in CAN ISO11519

Signal		Reverse state			Dominant state	
	min	nominally	max	min	nominally	max
CAN-High	1,6V	1,75V	1,9V	3,85V	4,0V	5,0V
CAN-Low	3,1V	3,25V	3.4V	0V	1,0V	1,15V

Standard ISO 11898-2 defines fast transmission on physical level. The bus's line must be adjusted in order to avoid wave's bouncing. The adjusting is done by an electrical resistor. The resistor's parameters are standardized and ranged from 100Ω to 130Ω (nominally 120Ω). According to ISO 11519, the impedance adjustment is not necessary for slow transmitting CAN buses. It is based on the fact that the speed of transmission is too slow to cause the wave to bounce.

6.1. CAN bus structure and functions

CAN topology is a bus based on the usage of copper twisted pair cable. Bus's maximum length is 40m and the connections between the bus and sensors should not exceed 0,3m.

The exchange of information in the network based on CAN bus occurs either through broadcasting or addressing. Broadcasting, similar to global information systems, is based on OSI/ISO Model, which in turn is based on the notion of sending information to all nodes. In this case, frame containing information is sent by the sender, and it does not contain receiver's information. This type of procedure allows faster exchange of information on the network by reducing the amount of time necessary for reading and interpreting of information about the addressee in the receiving part of the network. Broadcasting, as a means for addressing, is used only in frames with high priority containing particularly important information (for example periodic pressure measurements). This is done in order to prevent an overwhelming amount of broadcasts which can occur in networks with large topology. The network independently decides whether the broadcasted information is required. Illustration 1 shows the mechanics of introduction of information with high priority.

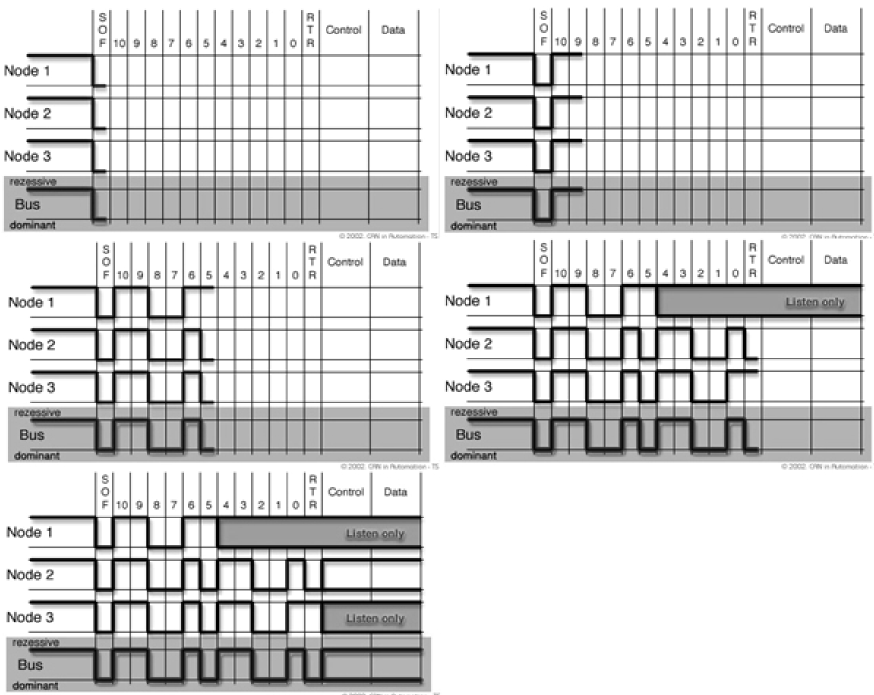


Fig. 6.1. The origin of recessive frame in CAN bus [5]

Frame based on direct addressing scheme is enlarged by the receiving node's address, which is located in the arbitration block. In this case the information is intended only for one node of the network. Inconvenience caused by the length of the frame, is compensated by the lack of broadcast to these nodes that don't need the information. The nodes simply ignore the frame, because they are not the intended recipients. If the intended node decides that the information contained in the frame is needed it sends confirmation. If the intended node did not send the confirmation, transmitting station sends the information again.
The CAN protocol contains two types of information frames. The standard information frame (version 2.0A) contains seven bit fields.

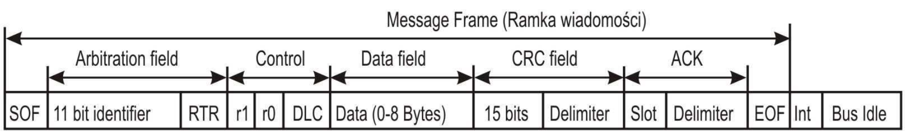


Fig. 6.2. The standard message frame for 2.0A version [5]

The frame consists of the following fields:

1. The starting field SOF (start of frame). It is the dominating bit (logical field "0"), marking the beginning of the frame,
2. Arbitration field containing eleven information bits and bit RTR (remote transmission request). The dominating bit RTR (logical level "0") marks that the information contains data. Negating value (logical value "1") marks that the information is a request from one of the nodes about data from multiple nodes in the bus,
3. Control field (CF) contains six bits; bits r0 and r1, which are reserved for future use, four bits DLC (data length code), marking the number of bytes in the field data,

4. Data field containing anywhere from zero to eight bit octets,
5. CRC field, containing fifteen bits of cyclic redundancy check.
6. Acknowledge field containing two bits. First bit is transmitted as a recessive bit. Then it is overwritten by dominating bits transmitted from other nodes that successfully received the information. The second bit is the filling bit.
7. End of frame (EOF) field contains seven bits.
8. Int (intermission field) is the last field containing three bits. After these three bits the bus becomes idle. There is no requirement for the length of the bus idle time, where all of the values are zero.

Version 2.0B transports 29 indentifying bits unlike version 2.0A that contains 11 identifying bits.

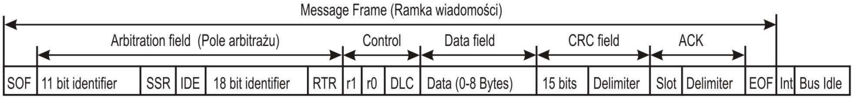


Fig. 6.3. The standard message frame for 2.0B version [5]

Differences between frames from Standard A and B:

1. Version 2.0B contains arbitration field, which in turn contains two indentifying bit fields. First (base ID) has 11 length bits to ensure compatibility with version 2.0A. Second field (extension ID) has 18 bits, thus, the whole length is 29 bits,
2. Differences between the formats were done by using identifying extension bit IDE,
3. SSR bit (substitute remote request) is located in the arbitration field. This bit determines the priority of the information.

There are two extra frames in the CAN bus; remote frame and error frame. Remote frame is used by the nodes as a request for a frame with the same priority (identical arbitration field). Error frames contain error codes which are transmitted to all nodes in the network (this will be discussed in the other parts of the chapter). In this case, all nodes generate the same error frame.

6.2. CAN coding and synchronization

CAN is coded with the Non-return to zero principle. This means that during the total bit time the generated bit level is either 'dominant' or 'recessive'. There is however a problem with NRZ encoding. When a certain amount of bits of the same value occur, there is no possibility for the receiver to synchronize with the bit stream. The receiver needs an edge to synchronize. The solution to this problem is the mechanism of bit stuffing. This simple rule inserts a bit switch after 5 consecutive bits of the same polarity. These stuff-bits are removed for interpretation of the bit-stream. They do not contain any information.

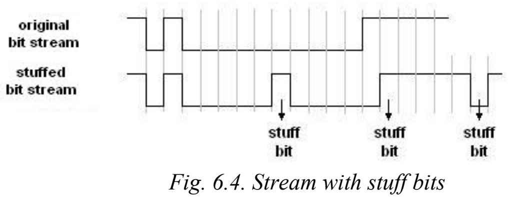


Fig. 6.4. Stream with stuff bits

Each node in the CAN-network has its own clock, usually a quartz oscillator. Timing parameters have to be set individually on each node. This has to create a common bit rate in the network, even if the nodes oscillators are different. The frequencies generated by these oscillators are not stable. These variations are caused by temperature differences, voltage differences and deteriorating components. As long as these variations remain in a specific tolerance range, the CAN node can compensate for these differences. The CAN specification defines that a bit is split up in 4 different segments: the synchronization segment, the propagation segment, the phase buffer segment 1 and the phase buffer segment 2. The sample point of the receiver is between the two phase buffer segments.

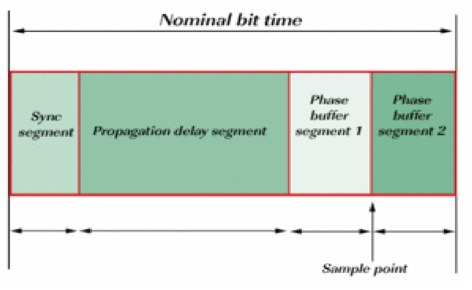


Fig. 6.5. CAN bit with sample point [4]

6.3. Errors and overflows

CAN errors are detected on the physical layer and on the data-link layer. On the physical layer 2 mechanisms are available:

1. Bit-error: Each station that transmits also observes the bus level and thus detects differences between the bit sent and the bit received. This permits reliable detection of global errors and errors local to the transmitter,
2. Stuff-errors: If more than 5 consecutive values are received, a stuff error is detected.

On the data-link layer 3 mechanisms for detecting errors are available:

1. CRC-error: If the CRC of a received frame doesn't match the calculated CRC, a CRC-error is detected,
2. Ack-error: If the acknowledge bit is not made dominant, the frame is not correctly received by at least one node,

3. Frame check: This mechanism verifies the structure of the transmitted frame by checking the bit fields against the fixed format and the frame size.

When one of these errors is detected, the transmission of the current messages is halted and an error flag is sent. A defect in a module can cause the network to break down, when this module flags all messages (even correct ones) as errors. Therefore a self- monitoring system is set up that can distinguish sporadic errors and permanent errors. This is done by assessment of station error situations. Depending on the number of faults the controller can be in 3 states:

1. Error active: An active error flag is sent when an error is detected,
2. Error passive: A passive error flag is sent when an error is detected,
3. Bus off: The controller is disconnected from the bus.

The active error frame consists of 6 dominant bits. This results with the other nodes in a stuff error. Therefore those other nodes will start to transmit their error frames (the extra error flag bits). The frame is delimited by a recessive delimiter of 8 bits. When an error frame has been detected, the data frame is automatically retransmitted.

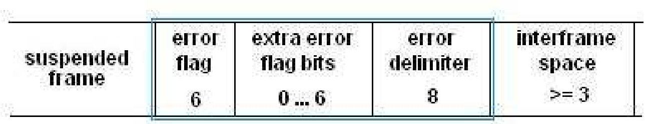


Fig. 6.6. The active error frame

The passive error frame sends 6 recessive bits, so it cannot influence the bus. It can however halt its own transmission. To avoid blocking the bus by a disturbed node sending high priority messages a transmission delay was introduced for nodes in Error Passive state. After transmission an Error

Passive node must wait 3 + 8 recessive bits before starting transmission again (additional wait state).

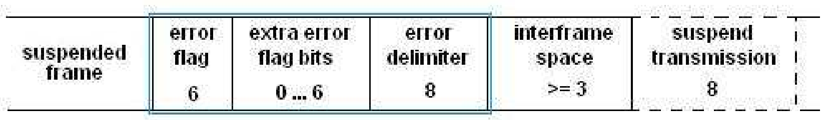


Fig. 6.7. The passive error frame

An Overload Frame can be generated by a node if due to internal conditions the node is not yet able to start reception of the next message or if during Intermission one of the first two bits is dominant.

With an Overload Frame the transmitter is requested to delay the start of the next transmission. The Overload Frame is almost identical to an Active Error Frame. The only difference is that an Overload Frame does not increase the error counters and does not cause a retransmission of a frame. Every node may transmit consecutively only 2 Overload Frames.

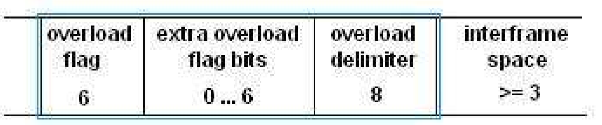


Fig. 6.8. The overload error frame

6.4. CAN and Ethernet

Conversion of CAN protocol into TCP takes place on the equipment level. There are Ethernet converters available on the market. Program conversion is not very popular because of its price and minimum effectiveness. It is possible to use Modbus/TCP protocol in the Ethernet network. The protocol puts Modbus frame in the TCP's data field frame.

The transmission control is done by standard TCP mechanisms, not by Modbus's checksum. The protocol allows the control over the driver from any computer in the Intranet or Internet, but since it is not protected, it is vulnerable to the outside attacks. It is possible to intercept sensitive data or even take over control of the entire device.

CAN converting systems in Ethernet

Most widely used converters currently available on the market are made by Philips, less popular are made by Realtec, Fujitsu and other producers. Depending on the type of interface, various microcontrollers are used. SJA 1000 system, made by Philips, is the dominating system currently available. Phillips SJA1000 is an autonomous CAN system that is capable of remotely sending and receiving frames. It performs the task of a data link. The physical layer is performed by CAN transmitter represented by 82C250 system.

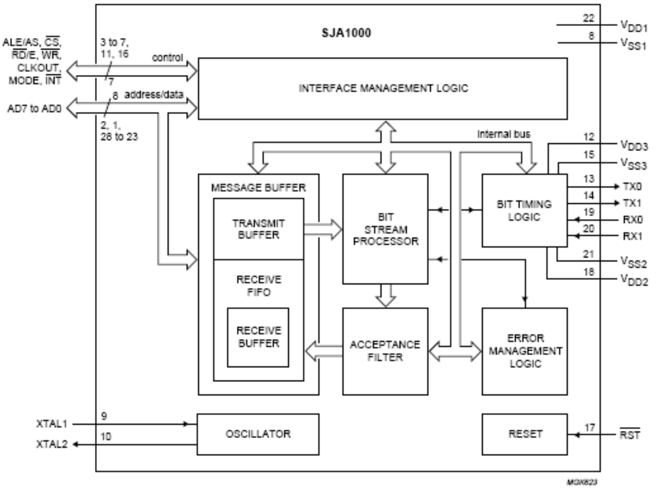


Fig. 6.9. The body of Philips SJA1000 circuit [7]

Less popular system, produced by Realtec, RTL8019AS is an integrated Ethernet controller which allows easy production of a card with full-duplex transmission capabilities. Communication between RTL8019AS and the microcomputer is performed through the help of ISA BUS. Remote control is based on writing appropriate values in the controlling registers and control's configuration. It also is based on reading of the status and data in the registers. The choice of the register to write or read is done by placing its address on the 5 line address BUS.
It must also be mentioned that control of the net is usually done with the help of popular AVR controllers produced by Amtel.

PROBLEM
List differences between CAN 2.0A and 2.0B standard.

CAN

	ISO11898	*ISO11519*
Frame length:	. . . .	. . . .
Voltages:	. . . .	. . . .
Identification bits:	. . . .	. . . .
Speed:	. . . .	. . . .

References

[1] Merkisz J., Mazurek S., *Pokładowe systemy diagnostyczne pojazdów samochodowych,* Wyd. WKŁ, Warszawa 2007.

[2] Rokosch U., *Układy oczyszczania spalin i pokładowe systemy diagnostyczne samochodów OBD,* Wyd. WKŁ, Warszawa 2007.

[3] Elektor Electronics *CAN bus*, Segment B.V., the Netherlands 1998.

[4] http://www.canopen.pl

[5] Elektronika Praktyczna *Magistrala CAN,* nr 1/2000.

[6] Milkuszka W., Trybus B., *Konwersja protokołów w rozproszonym systemie sterowani,a PSW/WWT--CAN,* Krynica 2001.

[7] Biuletyn Automatyki nr 48, (2/2006), http://www.astor.com.pl.

[8] http://www.ely.pg.gda.pl/~rkraj/can/parametr.htm#CZASY.

[9] Długosz T., Ruan S., Sun S., Zhang L., *Some Consideration on Electromagnetic Compitibility in CAN Bus Design of Automobile*, 2010 Asia-Pacific International Symposium on Electromagnetic Compatibility.

7. Local Interconnect Network (LIN)

There are few types of buses used in automatics. Their main function is to allow communication between the executing and decision units. The buses often pass the information from a sensor web. Vehicle's automatic systems are not an exception. Since 2009, every new vehicle (except special purpose vehicles) is equipped with buses based on CAN protocols. This does not mean that other bus systems are not used. There are also different types of buses commonly used in automotive, i.e.: InterBus-S, Profibus, FIP, EIB, LON, FlexRay, D2B, MOST. Frequently, vehicles have few different types of buses cooperating together.

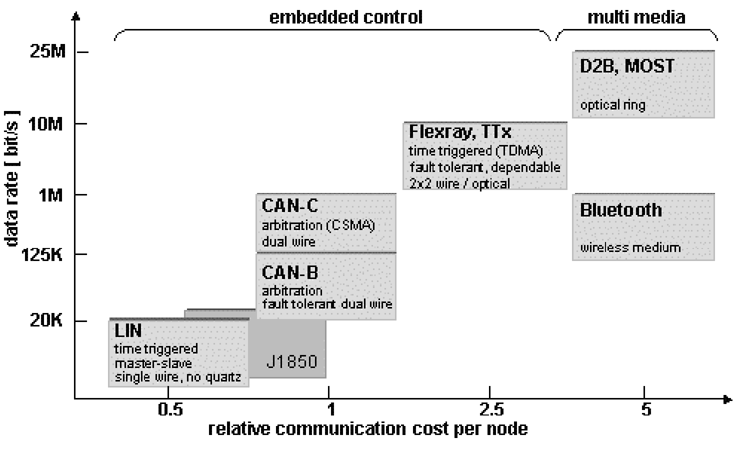


Fig. 7.1. Buses data rate vs. cost per node [1]

LIN bus is a cheaper option then CAN bus. It doesn't require twisted-pair cable, but a single copper wire. It can decrease the cost of installation by half. Maximum channel capacity is smaller in LIN buses vs. CAN buses. That's why LIN buses are used is systems which don't require quick information transfer.

Tab. 7.1. The comparison of LIN and CAN bus

	LIN	**CAN**
Automotive Use	Sub-networks	Networks
Medium Access Control	Single Master	Multiple Master
Bus Speed	2.4 – 19.6 KBaud (limited by EMI and clock synchronization considerations)	62.5 – 500KBaud
Typical Number of Nodes	2 – 10	4 – 20
Physical Layer	Single wire at battery voltage plus ground	Twisted Pair at 5V
Master Clock Generation	Crystal	Crystal
Slave Clock Generation	RC	Crystal or Resonator
Relative Cost	½	1

Most commonly used LIN buses in vehicles are based on communication with:

1. Rain sensors,
2. Light sensors,
3. Sun roof (in cabriolets),
4. Mirror position switches,
5. Window lifting and lowering switches,
6. Central door lock,
7. Windshield wipers system control,
8. Air-conditioning system,
9. Heating systems (all of the following: central, seat, windshield, and mirror heating),
10. Lighting system,
11. Seat and steering wheel positioning systems,
12. Loudspeaker and volume control systems.

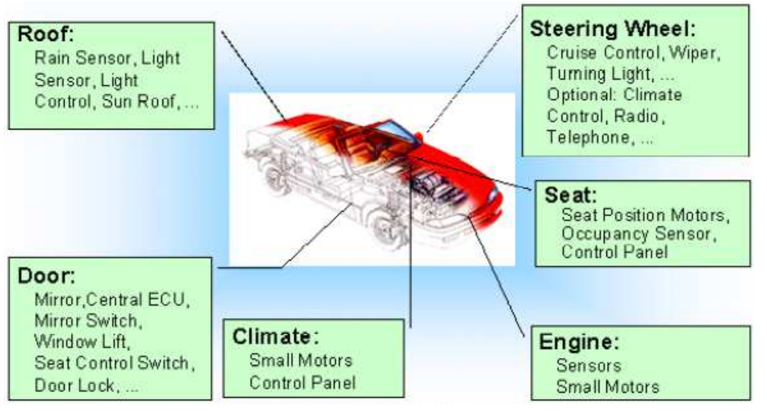


Fig. 7.2. LIN bus in vehicle [4]

7.1. LIN bus structure

LIN is based on bus topology. Unlike CAN, it is based on hierarchy node with single master node. The number of slave nodes should not exceed 16. In order to synchronize the process, the master node is equipped with an oscillator. To make the communication more reliable and improved, the master node can do both: the master and slave tasks. It is impossible for a slave node to perform master tasks.

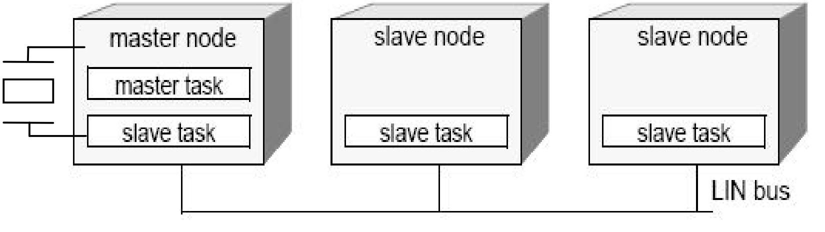


Fig. 7.3. LIN as MASTER – SLAVE bus [4]

Synchronization in LIN's master node is done by a simple RC resonator. The scheme of LIN node is simple, and consists of microcontroller, stabilizer, and sender/receiver system.

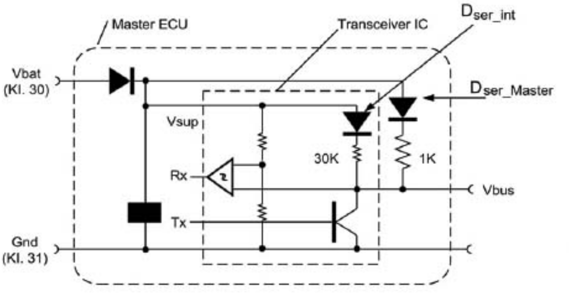


Fig. 7.4. Scheme of LIN node [4]

From an electro technical point of view the nodes and the bus are connected in parallel. The bus's resistance (not including the copper wire's resistance) is based on a formula:

$$R_b = \frac{1}{G_m + nG_s} \tag{1}$$

where: G_m – conductance output of *MASTER's* node, G_s – conductance output of *SLAVE's* node.

PROBLEM

Find the resistance of the LIN bus equipped in 10 slave nodes (30kΩ resistance) and one master node of 1kΩ resistance. The length of the copper wire equals 5m.

7.2. Frame based communication

As in the case of CAN bus, the binary digital signals can be either dominant or recessive. The recessive state is when the voltage value is close to power voltage (V_{bat} – fig.). This occurs when voltage value is $0,6V_{bat}$. The dominant state (VBUSdom) occurs when the gorund voltage value is below $0,4V_{bat}$.

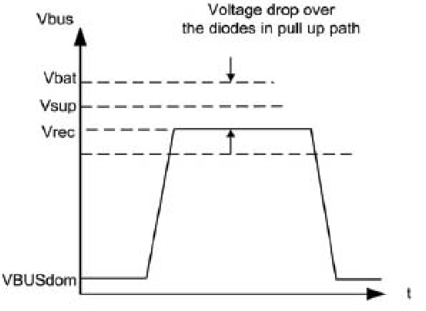


Fig. 7.5. Voltage values for dominant and recessive state

The minimum value of the bus's resistance is based on the formula:

$$R_{bmin} = \frac{Vbat - VBUSdom}{I_s} \tag{2}$$

PROBLEM

Find the minimal resistance of the LIN bus, when the power supply value equals 12V, the voltage at the dominant state is 1,2V, and the electrical current in the node is 40mA.

The master node's reply to the received frames depends on the values shown in the „*Time Schedule*" (TS) table. The TS consists of a sequence of entries which are processed one after another. Once the master task reaches the end of the active schedule table it starts from the beginning, thus establishing cyclic behavior of the signal–transmission process.

An entry comprises of two fields, the frame identifier whose transmission is controlled by the entry and the *frame slot.* The frame slot is the time allotted for the frame's transmission. It has to be larger than the maximum time it takes to transmit the frame and has to take jitter into consideration. The basic unit of time called the *time base period* depends on the configuration of the master. It typically ranges from 5 to 10 milliseconds.

Tab. 7.2. Sample schedule table, time base period 5ms

Frame Identifier	Delay in Time Base Periods	Delay in Milliseconds
2	3	15
3	2	10
4	4	20
5	1	5

Just like the CAN bus, LIN bus's frame is not addressed. The frame's destination depends on the identifier included in the header. There are five types of frames:

1. Unconditional frames. They carry standard information sent by the bus,
2. Event triggered frames. The master node sends a question to the slave nodes requesting information about the particular occurring event. The responding unconditional frame carries the identifier of the node with requested information,
3. Sporadic frames. Sent with another frame in the same time base period (shown in TS),
4. Diagnostic frames. Carry information about SLAVE node's control and configuration,
5. Reserved frames. The role of reserved frames is not clearly defined. It is reserved for the producers, and for future versions of LIN systems.

The header consists of an identifier and synchronized bits (in the MASTER node). The general structure of the frame is shown in the picture below:

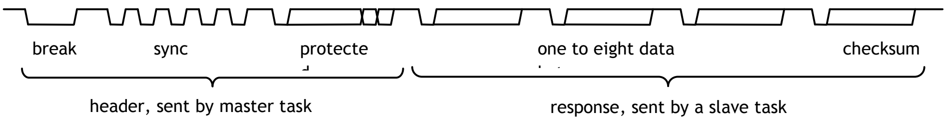


Fig. 7.6. The build of LIN frame[4]

The header of a LIN frame, sent by the master task, starts out with the break field consisting of 13 dominant bits. The sync byte (55 hexadecimal) is sent next. The protected identifier field, consisting of the frame identifier and two parity bits, concludes the header. The frame identifiers 0 to 3B hexadecimal are used for signal carrying frames, identifiers 3C and 3D are used for diagnostics and configuration data. The response part of a LIN frame is sent by the slave task. Slave task publishes the frame indicated by the frame identifier in the header. The response consists of up to eight data bytes and a checksum, which is one byte long.

LIN has 3 slave communication states. After reset, a slave enters the *initializing* state. As soon as a slave has completed its initialization it enters the *operational* state. If the slave receives a go to sleep request or the bus is inactive for 4 to 10 seconds the slave enters *sleep* state. During the sleep state the bus is driven to the recessive state, and no communication takes place. Any sleeping node can request a wakeup by driving the bus to the dominant state for 250μs to 5ms. Once the master node becomes active, it starts to process the active schedule table and transmits the first frame header.

The master node can put the whole cluster to sleep by sending a master request frame with the *go to sleep* command. Aside from managing the transmission of frames on a LIN network, the master also monitors the health of each slave by evaluating the special signal *response error* that has to be sent by each slave unconditionally.

Communication errors

There are three ways of detecting and correcting errors in LIN bus:

1. Monitoring. Consisting of checking by the broadcasting node what type of frame is being sent, checks what is being sent, and what's happening on the bus,
2. Negating checksum modulo 256 for data field with MSB bit displacement added to LSB,
3. Double detection of identifier's field parity.

The methods mentioned above allow 100% effectiveness in detecting transmitter's errors and a very high level of detection for global protocol errors.

7.3. Application ability of LIN

LIN is the newest bus used in vehicles. Its features (such as bandwidth 19200bps, ability to use a universal transmitter, deterministic communication, broadcasting, easy error detection based on checksum and pairing, and small bits of information, less than 64b) show that it will continue to be researched and implemented in new vehicles (today's version, which follows the ISO 9141 standard is named 2.1). The number of companies offering systems for this kind of bus is growing.

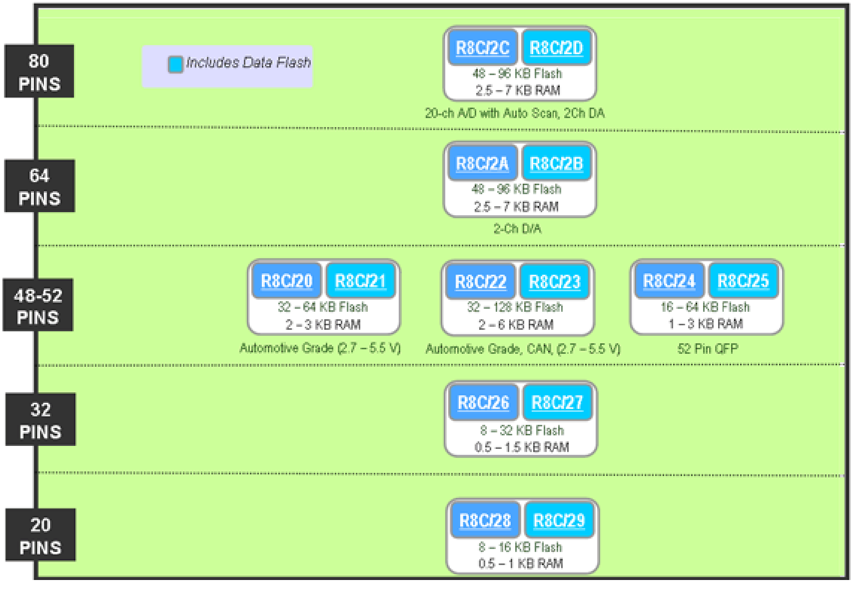


Fig. 7.7. R8C/2X family circuit by Renesas for LIN (with oscillator)[2]

The figure above shows interest of different companies in LIN Bus. It is possible to receive free samples for educational purposes. Below are the most important parameters of R2C/26 circuit:

1. 8-bit Multifunction Timer with 8-bit prescaler (Timer RA and RB): 2channels,
2. Input Capture/Output Compare Timer (Timer RC): 16-bit x 1 channel,
3. Real-Time Clock Timer with compare match function (Timer RE): 1channel,
4. UART/Clock Synchronous Serial Interface: 2 channels,
5. I^2C-bus Interface (IIC)/Chip-select Clock Synchronous Serial Interface: 1 channel,
6. LIN Module: 1 channel (Timer RA, UART0),
7. 10-bit A/D Converter: 12 channels,
8. Watchdog Timer,

9. Clock Generation Circuits: XIN Clock Generation Circuit, On-chip Oscillator(High/Low Speed), XCIN Clock Generation Circuit,
10. Oscillation Stop Detection Function,
11. Voltage Detection Circuit,
12. Power-On Reset Circuit,
13. I/O Ports: 25 (incl. LED drive ports),
14. External Interrupt Pins: 7.

References

[1] *Renesas*, www.america.renesas.com.

[2] *Dokumentacja elementów elektronicznych*, www.elenota.pl.

[3] Fryśkowiak B., Grzejszczyk E., *Systemy transmisji danych*. WKŁ, Warszawa 2010.

[4] ISO (2004). *Road Vehicles – Diagnostics on Controller Area Networks (CAN) – Part 2: Network Layer Services* (ISO 15765-2). International Standards Organization, www.iso.org/iso/home.htm.

[5] Lin Consortium. (2006). *LIN Specification Package, Revision 2.1*, www.lin-subbus.org.

8. Media Oriented System Transport (MOST)

Improvements in the automotive industry demand larger bandwidth in bus in order to process more data. First bus with larger throughput that were implemented in large scale production were in Class S Mercedez. It was the D2B (digital data BUS). After few years, D2B had new competitors; IEEE1394 (Firewire) and MOST (Media Oriented System Transport).
First version of MOST was finished in 1998 as a result of cooperation between Audi, BMW, Deimler Chrystler, Harmann-Becker and Oasis Silicon System (the last two companies are leaders in the automotive electronic industry).

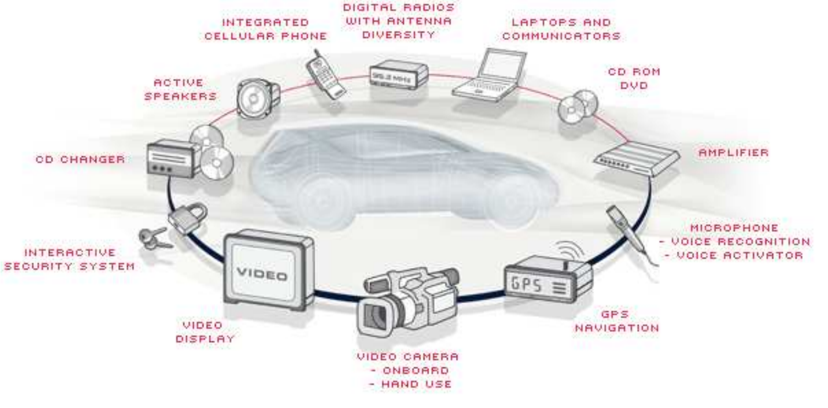


Fig. 8.1. MOST applications [4]

The MOST bus is geared towards multimedia applications. Its main features are:

1. Optical or electrical (MOST50) physical layer,
2. No ground loops,
3. Fully synchronous data transfer,
4. A control channel for controlling the participants in a network,

5. Streaming data channels for streaming synchronous audio and video data at a guaranteed data rate,
6. A packet data channel for packetized data e.g. internet protocol traffic,
7. An isochronous transfer for data that needs a guaranteed bandwidth but is variable in length e.g. MPEG data (MOST 150 only),
8. Partition of the bandwidth between the streaming data channels and the packet data channel is configurable.

8.1. Node Structure

Each piece of equipment connected to MOST bus is a node with an individual address. Comparing MOST node to its counterpart on different nets, one will notice a much greater usage, based on its ability to perform more tasks for various equipment, for example video system located in the back of the headrest. The sender / receiver system is responsible for communication in the BUS. It is integrated with the driver, which has a task of communicating with net services (and with codec's system from the above mentioned video system). The net services block is a liaison system between the driver and systems performing the function, which is characteristic for various applications – equipment (the functioning block video will contain the description of these function such as, rewind, stop the frame, open, etc). Each of these functions is a set of specific instructions; for example, for the rewind function it instructs it to increase or decrease the time by several frames. MOST bus allows sharing between functions. Cooperation allows for easier configuration of the system. Its hierarchy allows object oriented programming with objects and methods.

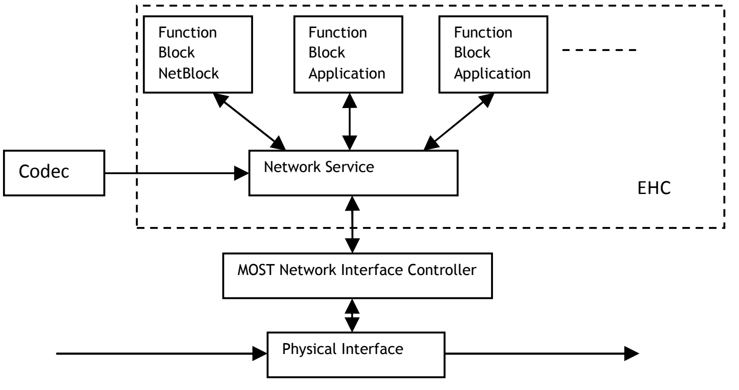


Fig. 8.2. Logical structure of MOST node [1]

One of the most important parts of the node (from the electrical point of view), other than the microprocessor, is the voltage stabilizing system. Other than powering the circuit, its task is to stop the node when it is not being used, and in emergency situations (reduce energy consumption).

The controller shown in the figure linking the node with the physical layer is able to perform tasks with the help of NIC (Network Interface Controller) and INIC (Intelligent NIC) used in faster MOST standards. They must be able to cooperate with EHC (External Host Controller) microcontroller which has a task of starting the application.

8.2. Topology and communication

MOST allows connection of 2 to 64 nodes (equipment) through the use of optical cables or twisted copper cable. It is possible to use the star topology and bus topology; however, ring topology is most widely used. The system is synchronized through Timing Maset (TM) that is connected to an oscilloscope.

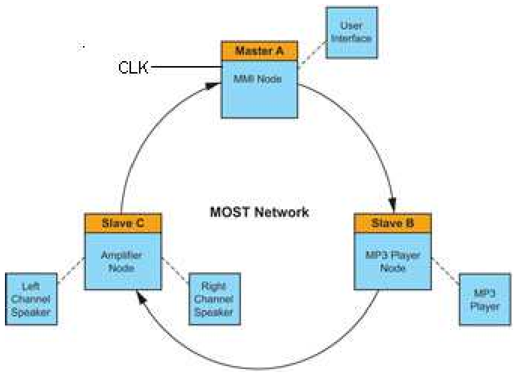


Fig. 8.3. The example of MOST topology [4]

Clock frequency of the TM node depends on equipment that make up the network. Sampling frequency 44.1kHz is commonly used which assures high quality of Hi-Fi audio signals. Bandwidth in the MOST25 standard is 24.5 Mbps and is increased to 50 Mbps in the MOST50 standard (MOST50 is still being researched)

In the case of optic cables there is no possibility of full duplex system. Thus, this problem was solved differently; each driver connected to MOST receives data, processes it, adds its own information and sends it further to the rest of subsystems according to the ring topology.

MOST uses wavelength of 650nm (ultraviolet – 1 transmission window) which is transmitted through a standard optical cable type DI shown in the figure below.

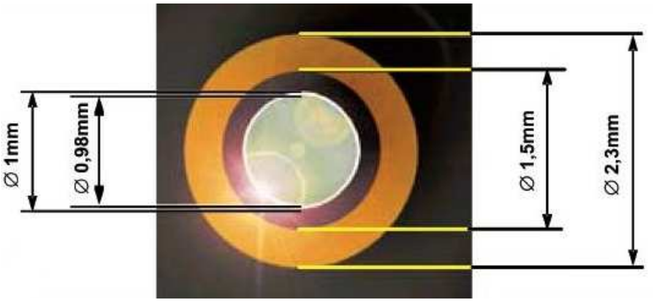
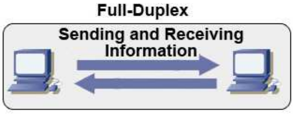


Fig. 8.4. MOST's fiber in Audi A8 [4]

PROBLEM

Work in a group. List advantages and disadvantages of full-duplex, half-duplex, and simplex systems.



Disadvantages: ....

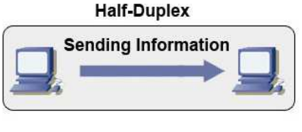
....

....

Advantages: ....

....

....



Disadvantages: ....

....

....

Advantages: ....

....

....

Disadvantages: ....

....

....

Advantages: ....

....

....

8.3. Frame Structure

MOST BUS works in three modes:

1. Sleep mode,
2. Stand-by mode,
3. Active mode (sending/receiving).

The sleep mode occurs only when:

1. There is no data transmitted by the bus,
2. All of the nodes send an information that they are ready to enter sleep mode, then go into off state,
3. There is no outside requests (driver or passenger is not using any functions that are part of the MOST bus),
4. There are no diagnostic processes in progress (i.e.: this occurs when the video is started).

Furthermore, sleep mode can be forced if the state of the battery (when the vehicle is not moving) is not allowing for normal operation or when the vehicle enters the „being transported" mode (this type of mode is possible in special purpose vehicles).
The stand-by mode is started by one of the bus ring drivers or through the use of the gateway module by different networks active in the vehicle.
In the active mode all of the functions are accessible. This mode is preceded by the stand-by mode. One of the conditions for active mode is the energy state of the battery or if the vehicle's engine is running. The active mode starts when any of the functions controlled by the BUS is activated.
The following figure shows frame structure, which is a part of the segment consisting 16 frames.

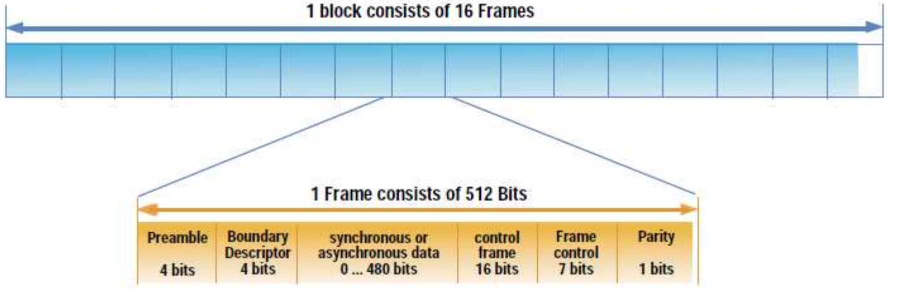


Fig. 8.5. The structure of segment and MOST frame [4]

There are 8 bits that start the transfer at the beginning of every data frame– starting field *Preamble* and *Delimiter*. Every frame must have the starting field – even in cases where several frames comprise a block containing one piece of information. The *delimiter*'s field task is to clearly separate transmitted data from the field *preamble*.
Afterwards there is data field, which contains 60 bytes of information (480 bits). The data field is dynamically divided into synchronous data fields, which can be from 24 to 60 bits (thus they can fill the entire frame) and

asynchronous data, which fill the rest of the frame. Synchronous data transfer always have priority. If asynchronous data does not fit in the frame, it is sent in the next frame. Although synchronous data has the priority, asynchronous data fills the first bits in the data field.

Synchronous data carry audio and video signals, which are transmitted in real time. This type of transfer is demanded due to clarity and fluidity of the visual and audible signals created by the CD discs. Furthermore, due to the high demands of video and audio transmissions, synchronous data fields are further divided into fragments synonymous to channels of the transmitted data, for example audio. Each channel fills one byte. Bytes are reserved by the transmission source depending on the type of the signal. This type of transmission system allows for simultaneous transmission of signals from multiple sources.

Synchronous data carry driving information, static images (pictures, maps, etc) and supporting information. Usually they are associated with the navigation system. Because this data is sent in non-regular lengths and the fill up different number of bytes, it is necessary that equipment sending this type of data store it, while waiting when it will be allowed to send it to the receiver. This means to wait for the moment when the sending equipment does not receive a frame with its address, and in which there are free bytes in the data field. Asynchronous data are stored even after being received by the receiving station, to the moment when the data comes back to the sending station. At this time, they are substituted by new data and the procedure repeats itself.

Data field is followed by a 16 bit control field. It carries information about the sender's and receiver's address and data regarding parameters of the given receiver (e.g.: change of settings in the video receiver in the headrest).

The last bits of the frame are reserved for the status field, which carries transmission parameters and one bit of parity control.
Such weak data protection is explained by the fact that optic cables are highly immune to electromagnetic disruptions. Thus, we don't have a problem of loosing or changing data. Thus, the audio and video signals are of similar quality to home entertainment systems.

PROBLEM
Independent work. List synchronous data types, and asynchronous types that are used in the MOST bus.

8.4. Application possibilities

The MOST system was first used in 3rd generation of Audi A8 (model 2002). It split components like:

1. Gateway,
2. Head unit,
3. AM / FM tuner (K-Box),
4. CD changer,
5. Digital Signal Processor (three different versions of it).

Additionally in version MMI-High (richer version) system had telephone, navigation and video tuner. When the system is started up, the main unit identifies the configuration used and initializes the corresponding components and the respective internal software (i.e.: navigation).

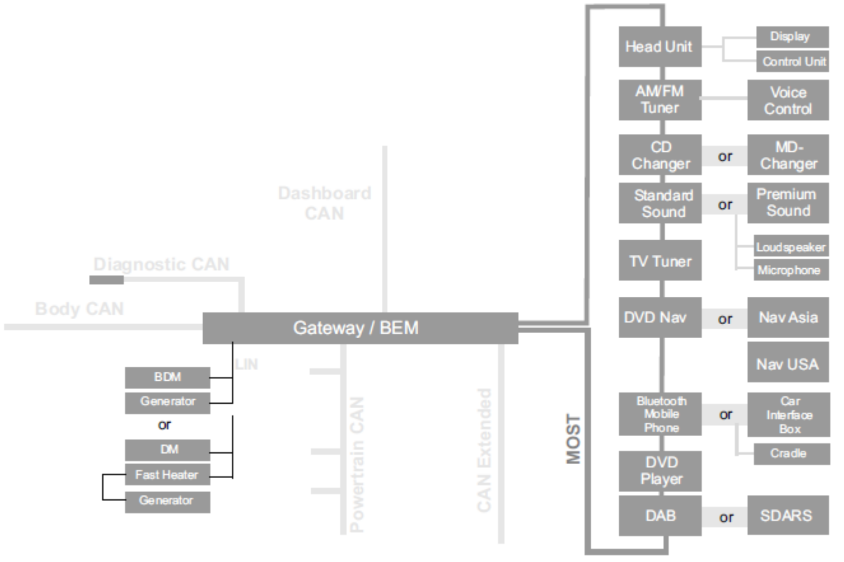


Fig. 8.6. MOST (MMI-high) in Audi A6 generation [4]

The software of a device is thus independent of the system configuration. The navigation application of the head unit is always present. Migration of system in Audi A6 model 2004 (fig.) had some extra devices like Bluetooth.

References

[1] Widerski T., *Sieć MOST, Świat Motoryzacji nr. 12/2007.*

[2]*Dokumentacja elementów elektronicznych*, www.elenota.pl.

[3] Fryśkowiak B., Grzejszczyk E., *Systemy transmisji danych*. WKŁ, Warszawa 2010.

[4] Grzemba A., *MOST – The Automotive Multimedia Network. wwwmostcooperation.com.*

[5] Atmel. *www.atmel.com*.

9. FlexRay

Word FlexRay™ comes from a combination of words *flexible* and *ray*. This name represents a consortium created in 2000 between DaimlerChrysler, BMW, Philips and Motorola. FlexRay's goal is to work on fast, reliable, and universal informatics network used in vehicles. The consortium, that in 2005 had 130 members, has a hierarchy structure. To the core of FlexRay consortium that makes all the decisions belong: BMW, Bosch, DaimlerChrysler, Freescale Semiconductor, General Motors, Philips and Volkswagen. The effect of cooperation between mentioned companies is the new version 2.1. The system has 10Mbps bandwidth in two channels, and has a collision avoidance system. FlexRay is suppose to become CAN's competition. But since it is a work in progress, it is still too expensive.

Fig. 9.1. FlexRay™ consortium logo [6]

PROBLEM

Independent work. Name 7 core members, 13 Premium members, developmental, and regular members of the consortium. Look up the information on the FlexRay website.

9.1. Node structure

FlexRay node is similar to other network's nodes e.g.: CAN, though its structure is more complicated. It is the effect of using voltage support system that can go into low energy usage state. Node's structure is based on four main blocks: host, communication controller, transceiver (most often redundant) and power supply.

The host is the part of the node where the application software is executed. The host provides the communication controller with control - and configuration information. It also provides the payload to be sent over the FlexRay network. There is also communication with the bus driver. The host can control the operational modes of the bus driver and read status information.

The communication controller is an electronic component where all aspects of the FlexRay protocol are implemented. The controller delivers the payloads received through the network to the host.

The bus driver is the transceiver of electrical signals on the bus. Each bus driver connects to one channel. So there can be two bus drivers on one communication controller.

Between the node's elements there are buses that send control, communication, and configuration signals. The node's structure is shown on the figure below.

Bus Guardian is the integral part of every node. Its main function is bus monitoring and granting permission to send signals at the appropriate timeslot. It is also possible for a single BG unit to control access to a bus for a few nodes.

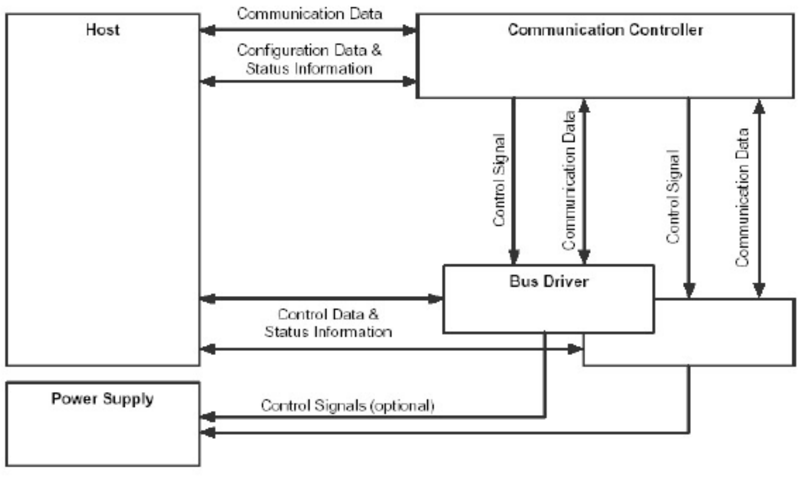


Fig. 9.2. FlexRay's node structure [6]

9.2. Topology

FlexRay is a deterministic network, allowing access to the bus only during a certain timeslot. Alike LIN, FlexRay can be a whole or a cluster network. It is possible to directly connect only two units through redundant channel. There are three main topologies:

1. Active star,
2. Passive star,
3. Bus (passive linear bus),
4. Hybrid (which aggregate above).

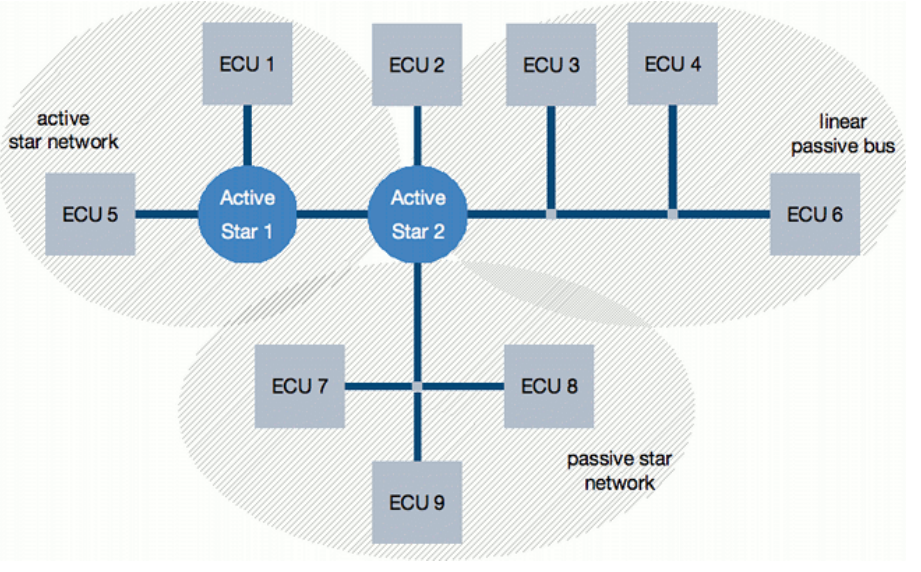


Fig. 9.3. FlexRay topologies [1]

Passive linear bus: The maximum distance between 2 nodes is 24 meters in passive linear bus. A node can connect to the bus through a split. There are maximum of 22 nodes on the bus. Practical limitations of the bus depend on the cable type and termination method.

The passive star is a special case of the linear passive bus. The passive star topology takes place when there is only one split in the network, where all the nodes connect to. The limitations of this topology are the same as the limitations of the linear passive bus.

An active star uses point to point connections between nodes and other active star. The active star to which the nodes are connected to has the function of transferring a data stream from one branch to all other branches. Since the active star device has a transmitter and receiver circuit for each branch, the branches are actually electrically decoupled from each other.

The parameters of cable and connectors are:

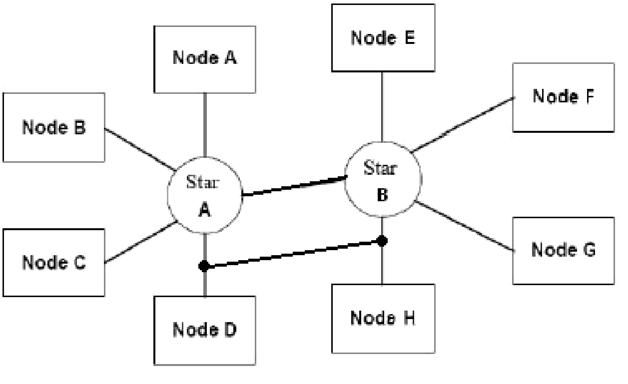
1. Cable attenuation: max. 85 db/km at 5 MHz,

2. Differential impedance: 80 - 110 Ω,
3. Line delay: max. 10 ns/m,
4. Contact resistance: max. 50,
5. Impedance of the connector: 70 - 200 Ω,
6. Length of the connector: max. 150 mm.

FlexRay topology is based on few important rules. First of all, the connection between two random active elements (nodes and active hubs) is possible in a certain cluster through device or passive topology. That means that neither passive hub, nor bus can be connected to more than one branch of active hub. Wrong connection will cause double transmission.

PROBLEM

Independent work. Find an error in the FlexRay topology shown below. Explain it and tell what kind of problems could it cause.



Connection between nodes is done by twisted pair cable or fiber optic cable. In order to avoid cross talk while using twisted pair cable, RC circuits and differential signals are being used.

9.3. Communication

Practices method of bit coding is NRZ (non-return-to-zero). After sending a bit, the signal value does not go back to neutral position. Information sending is based on assigning a certain time window to every communication node (Time Division Multiple Access). The size of a time window (also called statistic segment) is constant for the network (but it divides into macrotic and microtic). The amount of information in segment is variable, and depends on the clock used in a certain node.

Fig. 9.4. TDMA method [3]

The FTDMA (*Flexible Division Multiple Access)* communication is also possible. It is based on using two channel transmissions of frames with variable length. It is especially important in nodes that are responsible for communication with important driver (i.e.: ABS system).

Fig. 9.5. FTDMA method [3]

The static slot is a mandatory part of the communication cycle. The static segment contains a configurable number of static slots. One frame always fits into one slot. All frames within the static segment are of an equal (configurable) size. To schedule transmissions each node maintains a slot counter variable for each of the channels. These are initialized with 1 at the beginning of the communication cycle. When a static slot ends, the counter is incremented. The static segment ends when the counter reaches a predefined number of static slots. A node can only transmit a frame in an assigned static slot. FlexRay uses the *FrameID* to keep track of which slot should be used for the frame's transmission. In one static slot (per node) a special type of frame can be sent. This is the sync-frame. Sync frames have to be sent on all connected channels. We will talk more about it in the section called Clock Synchronization.

A mini-slotting based scheme is used to arbitrate the transmission of frames in the dynamic segment. These minislots are potential start times of transmissions. Frames can vary in length so no bandwidth is spilled.

In order to schedule transmissions, each node continues to maintain the 2 slot counters, each for a different communication channel. While the 2 counters increment simultaneously in the static segment, they can be incremented independently in the dynamic segment. Access on the 2 channels doesn't have to occur at the same time. A set of consecutive dynamic slots, each containing one or more minislots, are superimposed on the minislots. The duration of these dynamic slots depends on whether or not there is communication on the bus. When the slot ends, the slot counter for that channel is incremented. This is done until the channel's slot counter has reached its maximum or the dynamic segment has reached the maximum number of minislots.

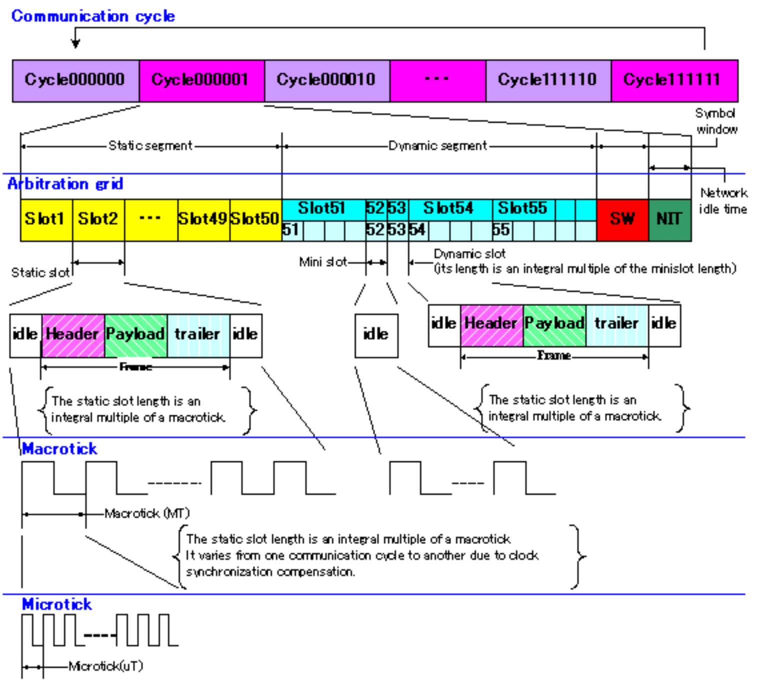


Fig. 9.6. FlexRay communication cycle [6]

9.4. Frame structure

The FlexRay's frame is shown in figure. It has the following features:

1. The header segment Header CRC is calculated and specified by the host,
2. The trailer segment CRC is calculated by the hardware,
3. The CRC also changes initial values on the connected channel to prevent incorrect connection.

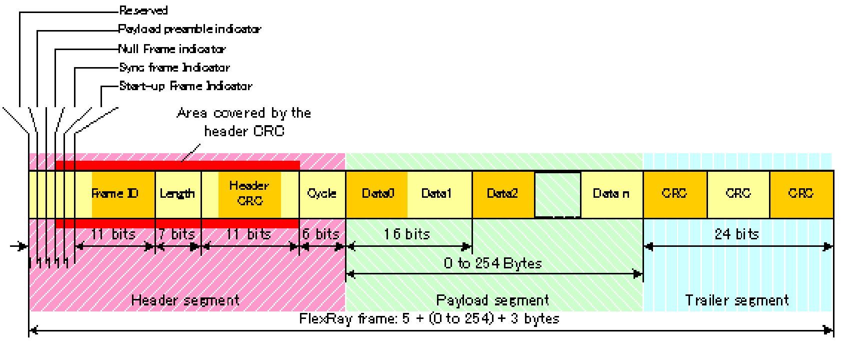


Fig. 9.7. The build of FlexRay's frame [6]

There are following segments shown in figure 3: header, payload and trailer. Payload segments consist of:

1. Reserved bit for future expansion,
2. Payload preamble indicator. This bit indicates the existence of a vector of information in the payload segment of the frame. At static frame, it indicates NW Vector, and at dynamic frame, it indicates Message ID,
3. Null frame indicator. This bit indicates whether or not the data frame in the payload segment is NULL ,
4. Sync frame indicator. This bit indicates the existence of synchronous frame,
5. Startup frame indicator. This bit indicates whether or not the node sending frame is the start-up node,
6. Frame ID. An ID is assigned to each node at system designing (valid range: 1 to 2047),
7. Length. It specifies the data length of the payload segment part,
8. Header CRC. It specifies the CRC calculation values of Sync Frame Indicator, Startup Frame Indicator, Frame ID, and Length that are calculated by the host,

9. Cycle (cycle counter). It indicates the cycle count of the node that transfers the frame during the frame transfer time,
10. Data. Valid range is from 0 to 254 bytes,
11. Message ID (optional). It uses the first two bytes of the payload segment for definition, and it can be used as the filterable data on the receiving side,
12. NW Vector (optional). The network management vector length must be from 0 to 12 bytes and common to all nodes.

Trailer segment consist of CRC. It is calculated and specified by the hardware. It changes the seed value on the connected channel to prevent incorrect connections.

PROBLEM

Work in groups of two. Name optional fields of FlexRay frame. Describe their function.

9.5. Application possibilities

Fujitsu is one of the members of FlexRay consortium. It introduced the driver system as integrated system named MB88121B. It is adjusted to work with FlexRay 2.1.

Its parameters include:

1. Providing up to a maximum of 128 message buffers,
2. 8 Kbytes of message RAM
    2.1. When the data section is 48 bytes, Maximum of 128 message buffers,
    2.2. When the data section is 254 bytes, Maximum of 30 message buffers,

3. Providing variable length message buffers
4. Each message buffer may be configured as a transmit buffer, receive buffer, or part of the receive FIFO
5. Host access to message buffers via an input buffer and an output buffer
   5.1. Input buffer: Stores a message to be transferred to message RAM,
   5.2. Output buffer: Stores a message that has been read out from message RAM,
6. Filtering by slot counter, cycle counter, and channel,
7. Each channel has a maximum bitrate of 10 Mbps,
8. Maskable interrupts,
9. 4 MHz / 5 MHz / 8 MHz / 10 MHz external oscillator circuit input,
10. Supporting external clock input,
11. CPU interface: 16-bit non-multiplexed parallel bus,
12. 16-bit multiplexed bus,
13. SPI interface,
14. Output lines for requesting DMA transfers,
15. Single 5V power supply or single 3.3V power supply.

Figure below shows block diagram of this circuit.

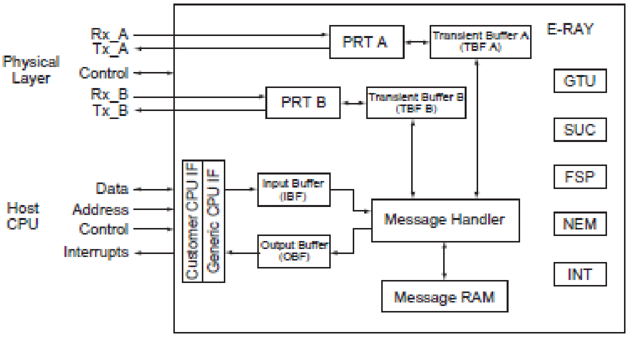


Fig. 9.8. Block diagram of circuit MB88121B [2]

This circuit can be (and is in BMW 7 series) a part of few vehicle systems such as engine control and brake control (common with CAN) and the main diagnostic bus which gathers information from MOST, LIN, CAN and others.

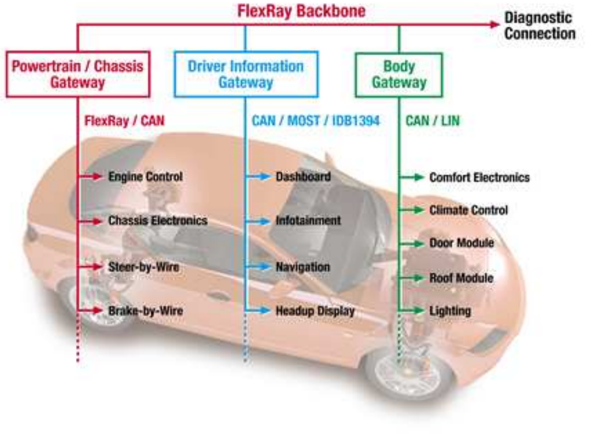


Fig. 9.9. FlexRay in a vehicle [6]

References

[1] *Fujitsu.* www.mcu.emea.fujitsu.com.

[2]*Dokumentacja elementów elektronicznych*, www.elenota.pl.

[3] Fryśkowiak B., Grzejszczyk E., *Systemy transmisji danych*. WKŁ, Warszawa 2010.

[4] ISO (2004). *Road Vehicles – Diagnostics on Controller Area Networks (CAN) – Part 2: Network Layer Services* (ISO 15765-2). International Standards Organization, www.iso.org/iso/home.htm.

[5] Lin Consortium. (2006). *LIN Specification Package, Revision 2.1*, www.lin-subbus.org.

[6] About FlexRay. http://www.flexray.com/.

10. Vehicles' sensoring

Vehicle micro processor controlled systems continually monitor the operating conditions of the car. Through sensors, computers receive vital information about a number of conditions, allowing minor adjustments to be made far more quickly and accurately than mechanical systems. Sensors convert input variables such as temperature, pressure, speed, position and other into either digital or analog electrical signals.

Fig. 10.1. Sensors in a vehicle [1]

The sensors and actuators are divided into the following areas:

1. Resistive sensors: potentiometers, thermistors, piezo resistive,
2. Voltage generating sensors: piezo electric, zirconia –dioxide, magnetic inductance,
3. Switch sensors: phototransistor, LED, speed, Air – bag impact (G – sensor),
4. Actuators: stepper motors, solenoids.

10.1. Piezoelectric effect

Piezoelectricity is the capability of certain crystals to bear a voltage when subjected to mechanical force. The word is borrowed from the word piezein (Greek), which means press or to squeeze. Piezoelectricity is reversible effect; piezoelectric crystals, subject to an extra voltage, can change shape by a small value. The effect is of the order of nanometers, but nevertheless finds useful applications such as the production and detection of vibrations, electronic frequency generation, generation of high voltages and ultrafine focusing of optical aggregates.

In a piezoelectric solid crystal, the negative and positive electrical charges are separated, but distributed symmetrically, so that the crystal overall is neutral. When a stress is appeared, this symmetry is disturbed, and the charge asymmetry generates a small amount of voltage. A 1 cm^3 of quartz with 2 kN of correctly applied force upon it, can produce 12,5 V of electricity.

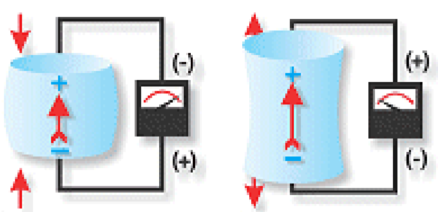


Fig. 10.2. Piezoelectric generated effect

10.2. An oxygen sensor

The Oxygen sensor is a chemical generator. It is constantly making a comparison between the Oxygen inside the exhaust manifold and air outside the engine. If this comparison shows little or no Oxygen in the exhaust manifold, a voltage is generated. The output of the sensor is usually between 0 and 1.1 V.

All spark combustion engines need the proper air to fuel ratio to operate correctly. For gasoline engines this is 14.7 parts of air to one part of fuel. When the engine has more fuel than needed, all available Oxygen is consumed in the cylinder and gasses leaving through the exhaust contain almost no Oxygen. This sends out a voltage greater than 0.45 volts. If the engine is running lean, all fuel is burned, and the extra Oxygen leaves the cylinder and flows into the exhaust. In this case, the sensor voltage goes lower than 0.45 volts. Usually the output range seen seen is 0.2 to 0.7 V.

The sensor does not begin to generate it's full output until it reaches about 600 degrees F. Prior to this time the sensor is not conductive. It is as if the circuit between the sensor and computer is not complete. The middle point is about 0.45 V. This is neither rich nor lean. A fully warm O2 sensor will not spend any time at 0.45 V. In many cars, the computer sends out a bias voltage of 0.45 through the O2 sensor wire. If the sensor is not warm, or if the circuit is not complete, the computer picks up a steady 0.45 V. Since the computer knows this is an "illegal" value, it judges the sensor to not be ready. It remains in open loop operation, and uses all sensors except the O2 to determine fuel delivery. Any time an engine is operated in open loop, it runs somewhat rich and makes more exhaust emissions. This translates into lost power, poor fuel economy and air pollution.

The O2 sensor is constantly in a state of transition between high and low voltage. Manfucturers call this crossing of the 0.45 V mark O2 cross counts. The higher the number of O2 cross counts, the better the sensor and other parts of the computer control system are working. It is important to remember that the O2 sensor is comparing the amount of Oxygen inside and outside the engine. If the outside of the sensor should become blocked, or coated with oil, sound insulation, undercoating or antifreeze, (among other things), this comparison is not possible.

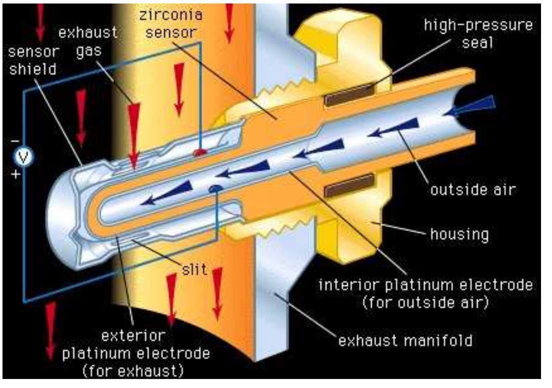


Fig. 10.3. An oxygen sensor [1]

10.3. Hall effect magnetic sensor

The Hall effect was discovered by Dr. Edwin Hall in 1879. It refers to the potential difference in value of voltage on both sides of a thin sheet of semiconducting (seldom conducting) material in the form of a van der Pauw element through which current is passing, this creates a magnetic field. The value of the current is known as the "Hall resistance", and is one of the element's material characteristics.

The Hall effect originates from the capacitance of the current flow in the conductor. The value of the current consists of many small charges (electrons) which see a force due to the magnetic field. Some of these charge elements end up forced to the sides of the conductors, where they create a pool of net charge. This is only notable in larger conductors where the separation between the two sides is large enough to create a noticeable charge.

The Hall effect differentiates between positive charges moving in one direction and negative charges moving in the opposite one. It is a very important feature. The Hall effect also presented the first real proof that electric currents in metals are carried by moving electrons, not by protons. Interestingly enough, the Hall effect also showed that in some substances (especially semiconductors), it is more appropriate to envision current as moving positive "holes" rather than negative electrons.

One can determine the strength of the magnetic field applied, by measuring the Hall voltage across the element. The so called Hall effect sensors are readily available from a number of different producers. Analog (or Linear) Hall effect sensors are the most popular types of sensors, which output a voltage that is proportional to the applied magnetic field. Another commonly used sensors are digital Hall effect sensors. Those are often used as magnetically controlled switches -- they turn on or off when the applied magnetic field reaches a predetermined level. These Hall effect switches generally consist of a Hall effect sensor, one or more logic gates and a transistor used to switch the electric current on or off.

Alternately, by applying a known magnetic field (typically from a permanent magnet) one can use the Hall voltage to measure the current through the element. This can be very useful as it allows one to measure the current in a conductor remotely through induction. This is widely used commercially in "live wire detectors", which allow you to quickly identify which wires are carrying current without directly plugging into them.

The quantum Hall effect can be observed in the presence of large magnetic field strength and low temperature. The quantum Hall effect is the quantization of the Hall resistance.

In ferromagnetic materials, the Hall resistance also shows an anomalous contribution, known as the Anomalous Hall Effect. It is proportional to the magnetization of the material. The scientists still debate over the origins of this well-recognized phenomenon.

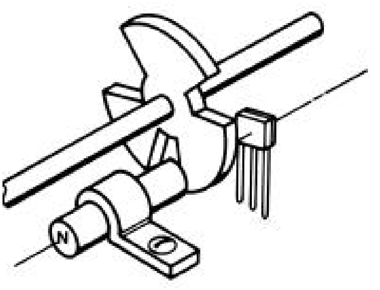


Fig. 10.4. Hall effect sensor [4]

10.4. Throttle position sensor

The TPS is a potentiometer attached to the throttle shaft. A voltage signal is supplied to the sensor, and a variable voltage is returned. The voltage increases as the throttle is opened. This signal and the MAP output determines how much air goes into the engine) so the computer can respond quickly to changes, increasing or decreasing the fuel rate as necessary.
In the figure Toyota 3L TPC is showed. The sensor basically looks at idle or closed throttle (IDL) and throttle angle opening (VTA). The TPS itself is simply a linear variable resistor that when driven by the ECU produces a linear voltage in a 0-5 volt range, 0 volts being idle and up to 5 volts representing throttle opening angle. Internally, there is also a switch that detects the idle position. Proper adjustment of the TPS is critical for engine performance, fuel economy, and emissions.

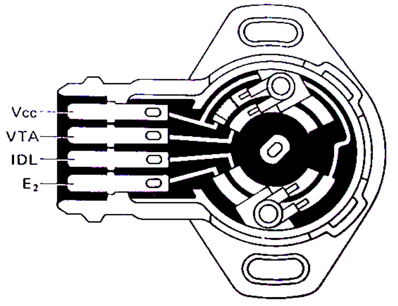


Fig. 10.5. Toyota 3L throttle position sensor [1]

10.5. Pressure sensor: Manifold Absolute Pressure (MAP) sensor

The manifold absolute pressure sensor is a variable resistor used to monitor the difference in pressure between the intake manifold at outside atmosphere. This information is used by the engine computer to monitor engine load (vacuum drops when the engine is under load or at wide open throttle). When the engine is under load, the computer may alter spark timing and the fuel mixture to improve performance and emissions.

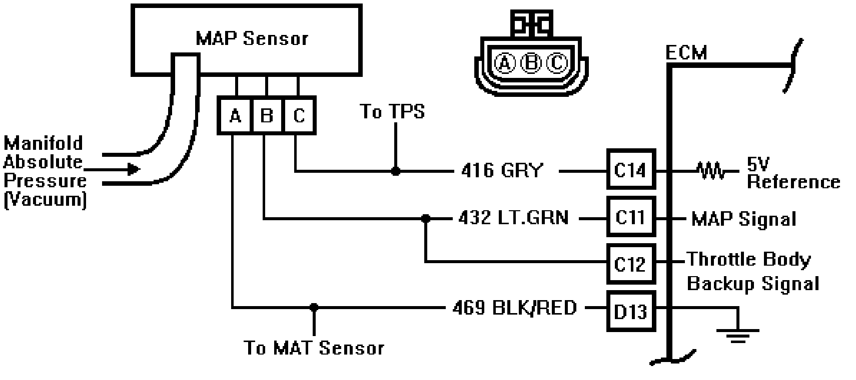


Fig. 10.6. The MAP sensor circuit [1]

10.6. Pressure sensor: knock sensor

Engine knock is characteristic of an uncontrolled combustion process and can cause engine damage. That is prevented by a knock sensor, a noise sensor near the engine.

The Knock Sensor is a Piezo Electric device that when you stress it, a voltage is produced. It senses knock and transmits information to the electronic engine management control unit. This influences process control in the engine, for example timing and fuel injection until knock is eliminated. Knock is frequently caused by fuel that does not comply with the required minimum quality.

Fig. 10.7. Engine with knock sensor

10.7. Temperature sensor: diodes (silicon band gap temperature sensor)

One of the most frequently used temperature sensors (thermometers) in electronic equipment is the silicon band gap temperature sensor. Its main advantage is that it can be incorporated in a silicon integrated circuit at very low cost. The principle of the sensor is that the forward voltage of a silicon diode is temperature-dependent, according to the following equation:

$$V_{BE} = V_{G0}(1 - T / T_0) + V_{BE0}(T / T_0) + (nKT / q)\ln(T_0 / T) + (KT / q)\ln(IC / IC_0) \quad (1)$$

where: T - temperature [K], V_{G0} – band gap voltage at absolute zero, V_{BE0} – band gap voltage at temperature T_0 and current IC_0, K - Boltzmann's constant, q - charge on an electron, n - a device-dependent constant.

By comparing the band gap voltages at two different currents, IC_1 and IC_2, many of the variables in the above equation can be eliminated, resulting in the relationship:

$$\Delta V_{BE} = (KT / q)\ln(IC_1 / IC_2) \quad (2)$$

An electronic circuit, that measures ΔV_{BE} can therefore be used to determine the temperature of the diode. The result remains valid up to from about 200 °C to 250 °C, when leakage currents become large enough to distort the measurement. Above these temperatures, more exotic materials such as silicon carbide can be used instead of silicon.

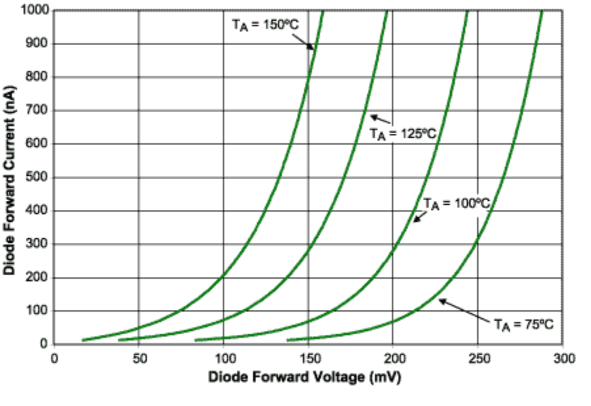


Fig. 10.8. The example of diode voltage vs. current characteristic for a few temperatures [3]

10.8. Temperature sensor: thermistor

A thermistor is a type of resistor used to measure temperature changes, which relies on the changes in its resistance with changing temperature. The relationship between resistance and temperature is linear (i.e. for a first-order approximation:

$$\Delta R = k\Delta T \qquad (3)$$

where: ΔR - change in resistance, ΔT - change in temperature, *k*-= first-order temperature coefficient of resistance.

Thermistors can be divided into two types depending on the sign of *k*. If *k* is positive, the resistance increases with increasing temperature, and the device is called a positive temperature coefficient (PTC) thermistor, or *posistor*. If *k* is negative, the resistance decreases with increasing temperature, and the device is called a negative temperature coefficient (NTC) thermistor. Resistors that are not thermistors are designed to have the smallest possible *k*, so that their resistance remains almost constant over a wide temperature range.

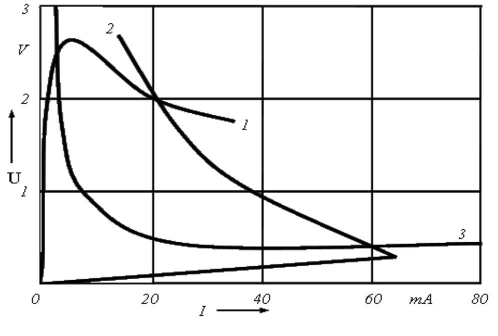


Fig. 10.9. The voltage vs. current thermistor characteristic: 1- NTC, 2 – PTC, 3 – CTR (critical temperature resistor) [5]

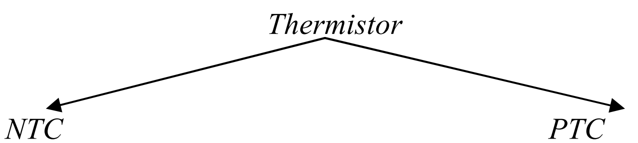
PTC thermistors can be used as current-limiting devices for circuit protection in place of fuses. Current through the device causes a small amount of resistive heating. If the current is large enough to generate more heat than the device can lose to its surroundings, the device heats up, causing its resistance to increase, and therefore causing even more heating. This creates a self-reinforcing effect that drives the resistance upwards, reducing the current and voltage available to the device.

NTC thermistors can be used as inrush-current limiting devices in power supply circuits. They have higher resistance initially which prevents large currents from flowing at turn-on, and then heat up and their resistance lowers to allow higher current flow during normal operation. These thermistors are usually much larger than measuring type thermistors, and are purposely designed for this application.

Fig. 10.10. NTC 4E7 thermistor

PROBLEM

What is the difference between NTC and PTC thermistor?



. . . .

. . . .

10.9. Temperature sensor: thermocouple

In 1822, Thomas Seebeck (an Estonian physician) accidentally discovered that the junction between two metals generates a voltage, which is a function of temperature. Thermocouples rely on this breakthrough, the so-called Seebeck effect. Although almost any two types of metals can be used to make a thermocouple, a number of standard types are used because they possess predictable output voltages and large temperature gradients.

Standard tables show the voltage produced by thermocouples at any given temperature; for example in the above diagram, the K type thermocouple at 300 °C will produce 12.2 mV. Unfortunately it is not possible to simply connect a voltmeter to the thermocouple to measure this voltage, because the connection of the voltmeter leads will make a second, undesired thermocouple junction. To make accurate measurements, this must be compensated for by using a technique known as cold junction compensation (CJC). In case you are wondering why connecting a voltmeter to a thermocouple does not make several additional thermocouple junctions (leads connecting to the thermocouple, leads to the meter, inside the meter etc), the law of intermediate metals states that a third metal, inserted between the two dissimilar metals of a thermocouple junction will have no effect provided that the two junctions are at the same temperature. This law is also important in the assembly of thermocouple junctions. It is acceptable to make a thermocouple junction by soldering the two metals together as the solder will not affect the reading. In practice, however, thermocouple junctions are made by welding the two metals together (usually by capacitive discharge) as this ensures that the performance is not limited by the melting point of a solder.

All standard thermocouple tables allow for this second thermocouple junction by assuming that it is kept at exactly zero degrees Celsius. Traditionally this was done with a carefully constructed ice bath (thus the term 'cold' junction compensation). Maintaining an ice bath is not practical for most measurement applications, so instead the actual temperature at the point of connection of the thermocouple wires to the measuring instrument is recorded.

Usually cold junction temperature is sensed by a precision thermistor in good thermal contact with the input connectors of the measuring instrument. This second temperature reading, along with the reading from the thermocouple itself is used by the measuring instrument to calculate the true temperature at the thermocouple tip. For less critical applications, the CJC is performed by a semiconductor temperature sensor. By combining the signal from this semiconductor with the signal from the thermocouple, the correct reading can be obtained without the need or expense to record two temperatures. Understanding of cold junction compensation is important; any error in the measurement of cold junction temperature will lead to the same error in the measured temperature from the thermocouple tip.

Thermocouples are available either as bare wire 'bead' thermocouples which offer low cost and fast response times, or built into probes. A wide variety of probes are available, suitable for different measuring applications (industrial, scientific, food temperature, medical research etc). One word of warning: when selecting probes make sure they have the correct type of connector. The two common types of connector are 'standard' with round pins and 'miniature' with flat pins. This causes some confusion as 'miniature' connectors are more popular than 'standard' types.

When choosing a thermocouple, consideration should be given to both the thermocouple type, insulation and probe construction. All of these will have an effect on the measurable temperature range, accuracy and reliability of the readings. Listed below is our (somewhat subjective) guide to thermocouple types:

1. Type K - Chromel (Ni-Cr alloy) / Alumel (Ni-Al alloy). Type K is the 'general purpose' thermocouple. It is low cost and, owing to its popularity, it is available in a wide variety of probes. Thermocouples are available in the −200 °C to +1200 °C range. Sensitivity is approx 41μV/°C,
2. Type E - Chromel / Constantan (Cu-Ni alloy). Type E has a high output (68 μV/°C) which makes it well suited to low temperature (cryogenic) use. Another property is that it is non-magnetic,
3. Type J - Iron / Constantan. Limited range (−40 to +750 °C) makes type J less popular than type K. The main application is with old equipment that can not accept 'modern' thermocouples. J types should not be used above 760 °C as an abrupt magnetic transformation will cause permanent decalibration,
4. Type N - Nicrosil (Ni-Cr-Si alloy) / Nisil (Ni-Si alloy). High stability and resistance to high temperature oxidation makes type N suitable for high temperature measurements without the cost of platinum (B,R,S) types. Designed to be an 'improved' type K, it is becoming more popular,
5. Type B - Platinum-Rhodium (Pt-Rh alloy). Suited for high temperature measurements up to 1800 °C. Unusually type B thermocouples (due to the shape of their temperature / voltage curve) give the same output at 0 °C and 42 °C. This makes them useless below 50 °C,

6. Type R - Platinum / Rhodium. Suited for high temperature measurements up to 1600 °C. Low sensitivity (10 μV/°C) and high cost makes them unsuitable for general purpose use,
7. Type S - Platinum / Rhodium. Suited for high temperature measurements up to 1600 °C. Low sensitivity (10 μV/°C) and high cost makes them unsuitable for general purpose use. Due to its high stability type S is used as the standard of calibration for the melting point of gold (1064.43 °C),
8. T-type thermocouple. A thermocouple best suited for measurements in the −200 to 0 °C. The positive conductor is made of copper, and the negative conductor is made of constantan.

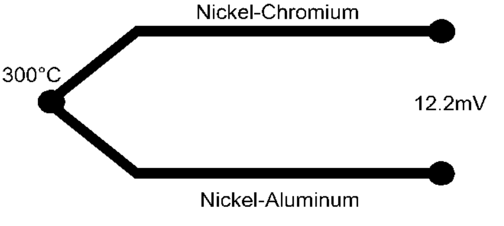


Fig. 10.11. The example of thermocouple type K [6]

PROBLEM

Work in groups. What kind of thermocouple do you know? Mention all groups and main features.

References

[1] The world of sensor. http://www.sensorsweb.com.

[2] Marciniak W., "Przyrządy półprzewodnikowe i układy scalone", WNT, Warszawa 1979.

[3] Electronic elements datasheet. http://www.elenota.pl.

[4] Hall effect. http://science.jrank.org/pages/3194/Hall-Effect.html.

[5] Thermistors. http://www.answers.com/topic/thermistor.

[6] Thermocouples. http://www.temperatures.com/tcs.html.

11. Lighting

Electromagnetic radiation travels through space as electric energy and magnetic energy. At times the energy acts like a wave and at other times it acts like a particle, called a photon. As a wave, scientists can describe the energy by its wavelength, which is the distance from the crest of one wave to the crest of the next wave. The wavelength of electromagnetic radiation can range from miles (radio waves) to inches (microwaves in a microwave oven) to millionths of an inch (the light we see) to billionths of an inch (x-rays). Visible light has wavelengths of roughly 400 nm to roughly 700 nm. This range of wavelengths is called the visible spectrum.

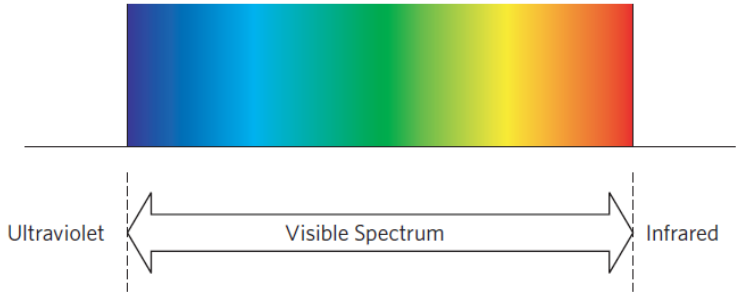


Fig. 11.1. Visible spectrum of light [4]

Electromagnetic radiation in the visible spectrum is typically generated by one of these sources:

1. Incandescent sources. The most common incandescent source is a tungsten light,
2. Non-incandescent sources such as fluorescent, metal halide, mercury vapor, neon, and HMI lights,
3. The sun.

All objects emit some electromagnetic radiation. As an object is heated, it emits relatively more of the shorter wavelengths of electromagnetic radiation and relatively less of the longer wavelengths. It is this property of light that allows a light meter to measure light's color temperature. The following figure demonstrates the visible wavelengths of the relative energy emitted at each wavelength of various color temperatures and in 5500K daylight. At 3200K there is a relatively large amount of the long wavelengths and a relatively small amount of the short wavelengths. As the color temperature increases to 5500K, 6500K, and 10000K, the relative amount of the long wavelength energy decreases and the relative amount of the short wavelength energy increases.

The 5500K daylight curve is not as smooth as the 5500K curve because daylight is a combination of the energy emitted by the sun, energy absorbed by the earth's atmosphere, and energy scattered by particles in the earth's atmosphere.

When the electrons in a molecule or a gas are excited, they rise to a higher energy level within that atom or molecule. After a period of time, the electrons return to their normal energy level and emit the diDerence in energy as electromagnetic radiation. The energy emitted is frequently in the visible spectrum.

When light strikes an object, the light can be transmitted, absorbed, or reflected. In many cases, all three occur. Transmission, absorption, or reflection can be determined by the wavelength of the light. For example, a piece of clear glass will transmit all the wavelengths of light striking the surface of the glass. If the glass is colored, some wavelengths are absorbed and some wavelengths are transmitted. If there are small particles in the glass, some of the wavelengths may be absorbed, some transmitted, and all reflected. In this case we would describe the glass as both colored and opaque. A piece of colored paper reflects some wavelengths, absorbs some wavelengths, and transmits no light.

If light strikes the surface of a transmitting object at an angle other than straight on, the light will be bent as it enters and exits the object. This property of light allows a lens to focus the light rays on a surface, such as the surface of the film used to photograph an object. Additionally, the short wavelengths are bent more than the long wavelengths. It is this property of light that produces a rainbow. As the light enters a water droplet, the light is bent. The light then reflects off the back of the water droplet. Then, as the light exits from the water droplet, the light rays bend again. Because the short wavelengths are bent more than the long wavelengths, the wavelengths of light are spread across the sky (and the rainbow is appeared).

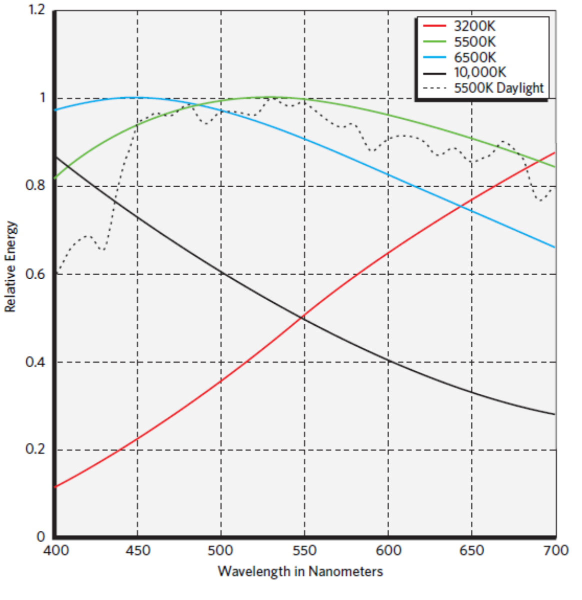


Fig. 11.2. Relative spectral energy curves for diDerent color temperatures [4]

Vision starts when light from a scene enters eye. The lens in eye focuses the light as an image onto retina. The human retina uses two types of cells to sense the light: rods and cones. These microscopic sensors are distributed across the retina, and each type serves a very different purpose. The rods and cones convert light into minute electrical impulses, which travel along nerve fibers to the brain. At the brain, they're translated into an impression of the shape and color of the observed object.

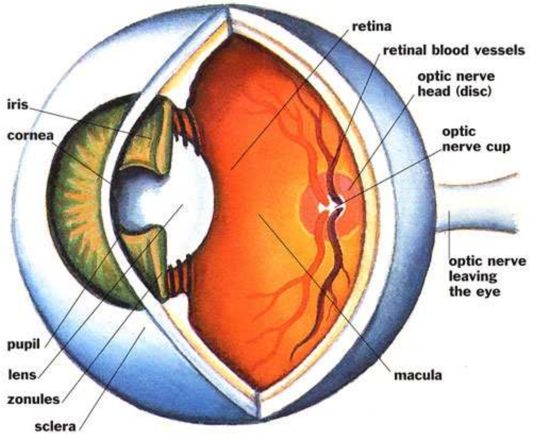


Fig. 11.3. The build of eye [4]

All rods have the same sensitivity to the wavelengths of light and, therefore, cannot see the color of an object. Rods see all objects as shades of gray. Because the rods are also very sensitive to light—much more sensitive to light than cones—they enable us to see in very low light levels, such as a night scene illuminated only by the stars or the moon. In bright scenes the rods are flooded with light and they cease to produce the signal that the brain uses for vision. In high brightness scenes only the cones provide useful information to the brain.

There are three types of cones: one has the greatest sensitivity to the long wavelengths of visible light; one has the greatest sensitivity to the middle wavelengths of visible light; and one has the greatest sensitivity to the short wavelengths of visible light.

People perceive brightness based on the total level of the signal coming from all the cones. One perceives color based on the relative signal levels coming from the three types of cones. When the cones sensitive to the long wavelengths are predominantly stimulated, one sees red; when the cones sensitive to the middle wavelengths are predominantly stimulated, one sees green; and when the cones sensitive to the short wavelengths are predominantly stimulated, one sees blue. Because there are only three types of cones, all vision is based on these three color perceptions. Therefore, most colors are described as light or dark and a combination of two colors, for example, red and blue (a reddish-blue or a bluish-red). Because of the processing of the signals from the cones in the brain, one cannot see a greenish-red or a reddish-green. The combination of red and green gives the sensation of yellow. Therefore, the object appears as greenish-yellow or yellowish-green. These sensations are the result of different amounts of signals from the red and green sensitive cones. When those signals are exactly the same, one sees yellow with no red and no green.

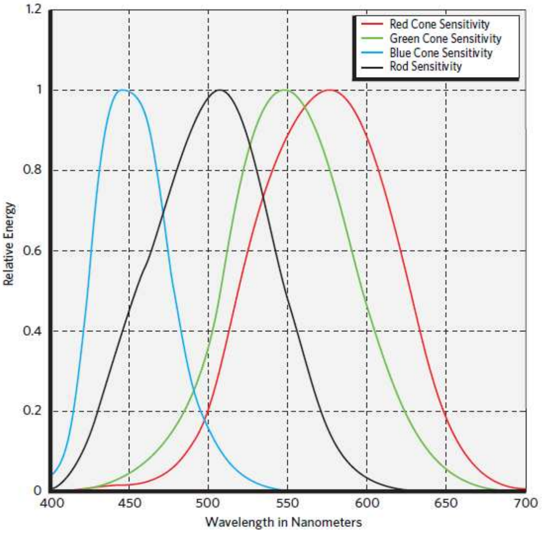


Fig. 11.4. Spectral sensitivities of the human rods and red, green, and blue sensitive cones [4]

They are 6 main phenomenon related with light ray:

1. Reflection. Defined as the bouncing back of a ray of light into the same medium, when it strikes a surface. It occurs on almost all surfaces - some reflect a major fraction of the incident light. Others reflect only a part of it, while absorb the rest,

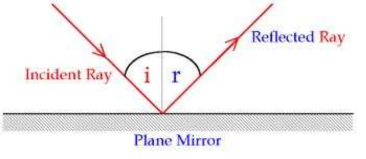


Fig. 11.5. The phenomena of reflection

2. Refraction. Defined as the bending of a light wave when it passes from one medium to another at the surface separating the two media. It basically occurs due to the speed of light being different in different media of different densities,

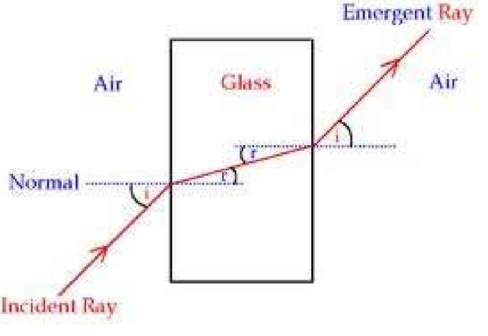


Fig. 11.6. The phenomena of refraction

3. Diffraction. A diffracted wave will "spread out". Diffraction occurs when the wavelength of a wave is of a similar size to an obstacle or a gap in a barrier,

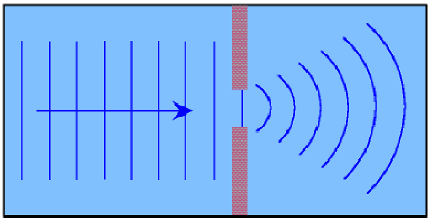


Fig. 11.7. The phenomena of diffraction

4. Polarization. Can be defined as fix in oscillation of light beam

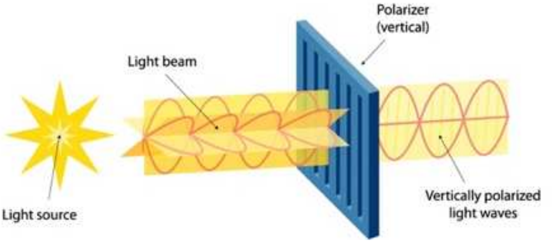


Fig. 11.8. The phenomena of polarization [5]

5. Interference. Defined as addition (superposition) of two or more waves that results in a new wave pattern

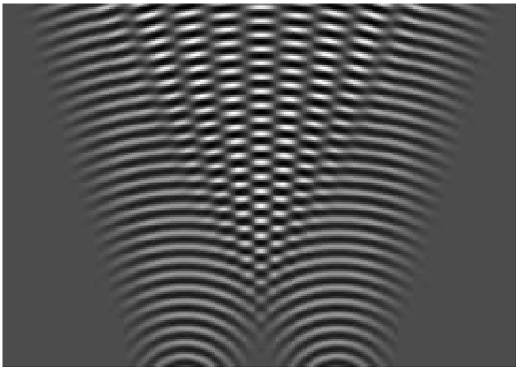


Fig. 11.9. The phenomena of interference [5]

6. Doppler effect. Based on the principle that the frequency of a wave is different when there is a relative motion between the source and the receiver, than it would have been had there been none.

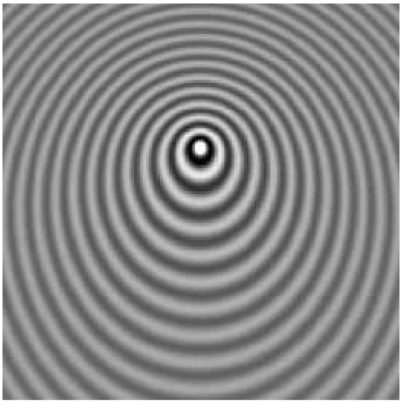


Fig. 11.10. The phenomena of Doppler effect (source of wave is moved against destination) [5]

11.1. Light sources in vehicles

Automotive, commercial, industrial, and retail facilities use several different light sources. Each lamp type has particular advantages; selecting the appropriate source depends on installation requirements, life-cycle cost, color qualities, dimming capability, and the effect wanted. Three types of lamps are commonly used:

1. Incandescent,
2. Fluorescent,
3. High Intensity Discharge (HID)
   3.1. Mercury vapor,
   3.2. Metal halide,
   3.3. High pressure sodium,

3.4. Low pressure sodium.

In cars the most popular are halogen and xenon lamp for external lighting and almost all kind for internal (dashboard and others).

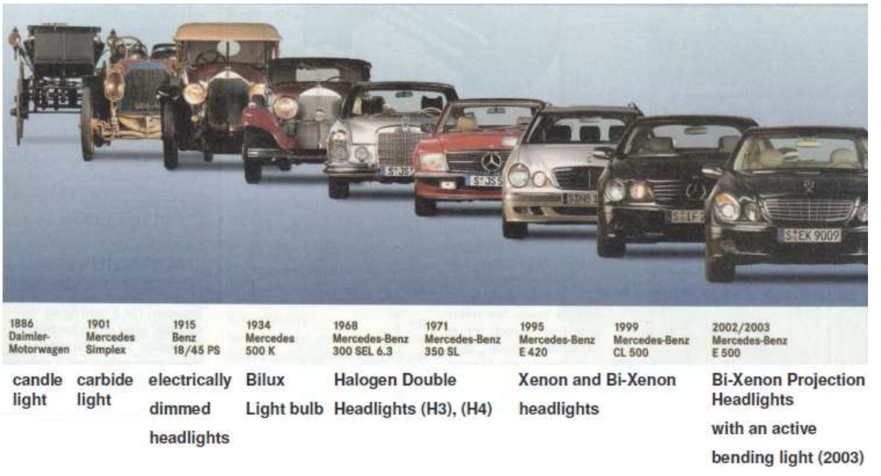


Fig. 11.11. Developing of Mercedes vehicles' light [4]

Electric light sources have many characteristics, including efficiency, color temperature, and color rendering index (CRI).

Some lamp types are more efficient in converting energy into visible light than others. The efficacy of a lamp refers to the number of lumens leaving the lamp compared to the number of watts required by the lamp (and ballast). It is expressed in lumens per watt. Sources with higher efficacy require less electrical energy to light a space.

Another characteristic of a light source is the color temperature. This is a measurement of "warmth" or "coolness" provided by the lamp. People usually prefer a warmer source in lower illuminance areas, such as dining areas and living rooms, and a cooler source in higher illuminance areas, such as grocery stores. Color temperature refers to the color of a blackbody radiator at a given absolute temperature, expressed in Kelvins [K]. A blackbody radiator changes color as its temperature increases (first to red,

then to orange, yellow, white, and finally bluish white at the highest temperature). A .warm. color light source actually has a lower color temperature.
The CRI is a relative scale (ranging from 0 - 100) indicating how perceived colors match actual colors. It measures the degree that perceived colors of objects, illuminated by a given light source conform to the colors of those same objects when they are lighted by a reference standard light source. The higher the color rendering index, the less color shift or distortion occurs. The CRI number does not indicate which colors will shift or by how much; it is rather an indication of the average shift of eight standard colors. Two different light sources may have identical CRI values, but colors may appear quite different under these two sources.

11.2. Halogen

A normal light bulb is made up of a fairly large, thin, frosted glass envelope. Inside the glass is a gas such as argon and/or nitrogen. At the center of the lamp is a tungsten filament. Electricity heats this filament up to about 2300 K . Just like any hot metal, the tungsten gets "white hot" at that heat and emits a great deal of visible light in a process called **incandescence.** The normal light bulb is not very efficient, and it only lasts about 750 to 1,000 hours in normal use. It's not very efficient because, in the process of radiating light, it also radiates a huge amount of infrared heat - far more heat than light. Since the purpose of a light bulb is to generate light, the heat is wasted energy. It doesn't last very long because the tungsten in the filament evaporates and deposits on the glass.

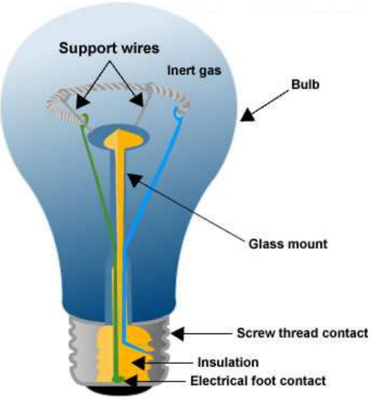


Fig. 11.12. The build of bulb [5]

A halogen lamp comes with a few modifications to eliminate this blackening problem. The bulb, made of fused quartz instead of soda lime glass, is filled with the same inert gases as incandescent lamps mixed with small amounts of a halogen gas (usually less than 1% bromine). The halogen chemically reacts with the tungsten deposit to produce tungsten halides. When the tungsten halide reaches the filament, the intense heat of the filament causes the halide to break down, releasing tungsten back to the filament. This process—known as the tungsten-halogen cycle—maintains a constant light output over the life of the lamp.

In order for the halogen cycle to work, the bulb surface must be very hot, generally over 500 K. The halogen may not adequately vaporize or fail to adequately react with condensed tungsten if the bulb is too cool. This means that the bulb needs to be smaller and made of either quartz or a high-strength, heat-resistant grade of glass known as aluminosilicate. Since the bulb is small and usually fairly strong due to its thicker walls, it can be filled with gas to a higher than usual pressure. This slows down the evaporation of the tungsten from the filament, increasing the life of the lamp.

In addition, the small size of the bulb sometimes makes it economical to use heavier premium fill gases such as krypton or xenon—which help retard the rate of tungsten evaporation—instead of the cheaper argon. The higher pressure and better fill gases can extend the life of the bulb and/or permit a higher filament temperature that results in better efficiency. Any use of premium fill gases also results in less heat being conducted from the filament by the fill gas. This results in more energy leaving the filament by radiation, slightly improving the efficiency.

Halogen bulbs thus produce light that is whiter and brighter, use less energy, and last longer than standard incandescent bulbs of the same wattage. They can last from 2000 - 4000 compared to conventional incandescent bulbs, which only operate for 750 - 1500 hours or three hours a day for about a year. However, halogen bulbs cost more.

Most halogen lamps range in power from 20 - 2000 W. Low voltage types range from 4 – 150 W. Some halogen lamps are also designed with a special infrared reflective coating on the outside of the bulb to ensure that the radiated heat, which otherwise is wasted, is reflected back to the lamp filament. The filament burns hotter so less wattage is required. These lamps can last up to 4000 hours.

Although more efficient than other large incandescent lamps, tungsten halogen lamps are inefficient relative to fluorescent and High Intensity Discharge lamp types. Halogen lamps can also pose a safety threat, as the heat generated can range from 400 – 750 K.

Fig. 11.13. Philips car halogen bulb [3]

11.3. XENON

In xenon lights, the purpose of the xenon is to greatly amplify the light that results from the high -intensity discharge electricity between two tungsten electrodes located inside the lamp tube or bulb. Xenon gas and a variety of metal salts are also located within the bulb. When a surge in voltage causes an arc of electricity to pass from one electrode to the other, the metallic salts vaporize. This vaporization gives off an extremely bright light. The intense energy of the vaporized salts excites the molecules in the xenon gas and causes the xenon to emit an intense blue glow.

Manufacturers assert that Xenon headlights will become the predominant type of headlight used in future cars. Not only are they three times brighter than traditional headlights, they provide better illumination for drivers and the high intensity discharge system they employ is lighter and less expensive for car manufacturers to install in vehicles.

Fig. 11.14. Philips car xenon bulb [3]

Xenon car bulbs have multiple advantages over standard headlight bulbs. These xenon vehicle bulbs emit much brighter and crisper quality of light as compared to ordinary halogen headlight bulbs. In fact the most up-to-date xenon bulbs emit almost twice as much light as regular halogen bulbs. The comparison for two bulbs: halogen and xenon is showed in table below.

Tab. 11.1. The comparison of xenon and halogen bulbs

Parameter	XENON	HALOGEN
Power - consumption	35W	55W
Brightness (CD)	200.000 CD	67.500 CD
Luminance (lm)	1900 ~ 3200 lm	<1000 lm
Temperature (K)	4200 ~ 12000 K	3200 K (yellow)
Life	2500 Hr sustained	200~300 Hr. sustained

The light produced by the Xenon system will give a more "even" spread and better penetration, as well as much improved contrast and reduced eyestrain. Reflectors from road traffic signs, cat-eyes and vehicles in front of vehicle that giving some "extra" seconds to react.

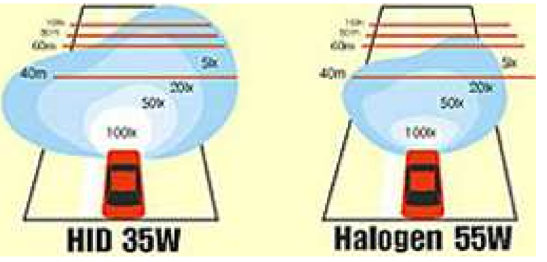


Fig. 11.15. The comparison of xenon and halogen bulbs light range [3]

11.4. LED

LED (Light Emitting Diode) is the latest in lighting technology and is considered the lightsource of the future. There are good solid reasons for this. The technology is impacting our day-to-day life all around the world. LED lighting is the most effective all lighting technologies available right now.

When a light-emitting diode is forward biased (switched on), electrons are able to recombine with holes within the device, releasing energy in the form of photons. This effect is called electroluminescence and the color of the light (corresponding to the energy of the photon) is determined by the energy gap of the semiconductor. An LED is usually small in area (less than 1 mm^2), and integrated optical components are used to shape its radiation pattern and assist in reflection.

LED technology is now being deployed in many different applications. In the automotive world, the 4x4 off-road racing fraternity were the first to embrace this technology. The incredible durability and sturdiness, combined with the superior lighting quality made it a logical choice for them.

Now, we find LED technology everywhere, in automotive, in houses, in offices, streetlighting, parking lots, industrial machinery, farm equipment, lighting in nurseries and glass houses, etc.

Each and every industry sector is using them in one form or another. The main adventages of LED are:

1. Require only a fraction of the power used by normal Halogen, or HID lights,
2. Is extremely sturdy, and can handle shock and vibration much better - for much longer,
3. Is extremely compact, allowing these units to be used in many different environments,
4. Last at least 250x longer than Halogen, and at least 20x longer than HID lights,
5. Require very little power compared to other light sources.

Fig. 11.16. LED car bulb [4]

PROBLEM

Work in pairs. What are the main advantage and disadvantage of use LED bulb in compare to XENON blub in a car?

CAR’S BULB

. . . .

. . . .

. . . .

. . . .

. . . .

11.5. HUD

Born of the military aviation world, in which systems project airspeed, compass heading, an artificial horizon, and a gunsight, vehicle HUDs use a small projector in the instrument panel that reflects the image off the base of the windshield in front of the driver. A windshield coating reflects the image so the driver can see the information, but you can still see through the windshield. The image is at the base of the windshield and appears to float just above the hood, so it isn't in the middle of one line of sight.

The focal point — how far away man eyes think it is — is somewhere between the front of the car and infinity. That means drivers who wear reading glasses (or who should wear them to see the instruments properly) don't have to refocus, as they might when glancing down to read gauges.

A dashboard knob adapts the elevation of the HUD system for drivers of different heights and also lets decide how far up into your field of vision the image projects. Only the driver sees the information, not the passengers.

There are two types of HUD. A fixed HUD requires users to look through a display element attached to the airframe or vehicle chassis. The system determines the image to be presented depending solely on the orientation of the vehicle. Most aircraft HUDs are of this type.

Helmet mounted displays (*HMD*) are technically a form of HUD, the distinction being that they feature a display element that moves with the orientation of the users' heads relative the airframe.

Fig. 11.17. The kind of HMD – binoculars [2]

HUDs are split into four generations reflecting the technology used to generate the images.

1. First Generation—Use a CRT to generate an image on a phosphor screen, having the disadvantage of the phosphor screen coating degrading over time. The majority of HUDs in operation today are of this type,
2. Second Generation—Use a solid state light source, for example LED, which is modulated by an LCD screen to display an image. These systems do not fade or require the high voltages of first generation systems. These systems are on commercial aircraft,
3. Third Generation—Use optical waveguides to produce images directly in the combiner rather than use a projection system,
4. Fourth Generation—Use a scanning laser to display images and even video imagery on a clear transparent media.

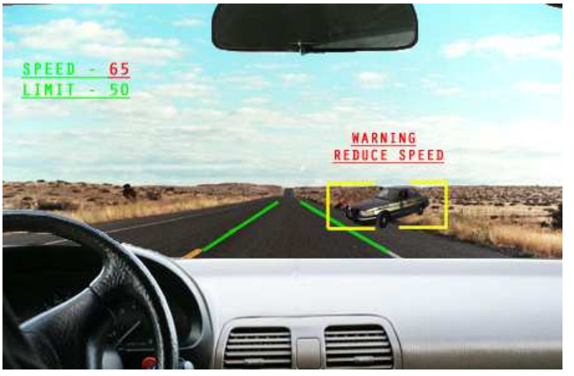


Fig. 11.18. HUD in a car [2]

References

[1] Falcon 2000 Becomes First Business Jet Certified Category IIIA by JAA and FAA; Aviation Weeks Show News Online September 7, 1998.

[2] HUD With a Velocity (Flight Path) Vector Reduces Lateral Error During Landing in Restricted Visibility; International Journal of Aviation Psychology, 2007, Vol. 17 No 1.

[3] Philips homepage*, www.philips.com.*

[4] The story of light. *http://science.howstuffworks.com*.

[5] The history of the light bulb. *http://invsee.asu.edu/Modules/lightbulb/meathist.htm*.