SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

SCHOOL OF MECHANICAL ENGINEERING

DEPARTMENT OF MECHATRONICS

UNIT – I – Batteries, AC & DC Generators – SAU1302

COURSE NAME: AUTOMOTIVE ELECTRICAL AND ELECTRONICS

COURSE CODE: SAU1302

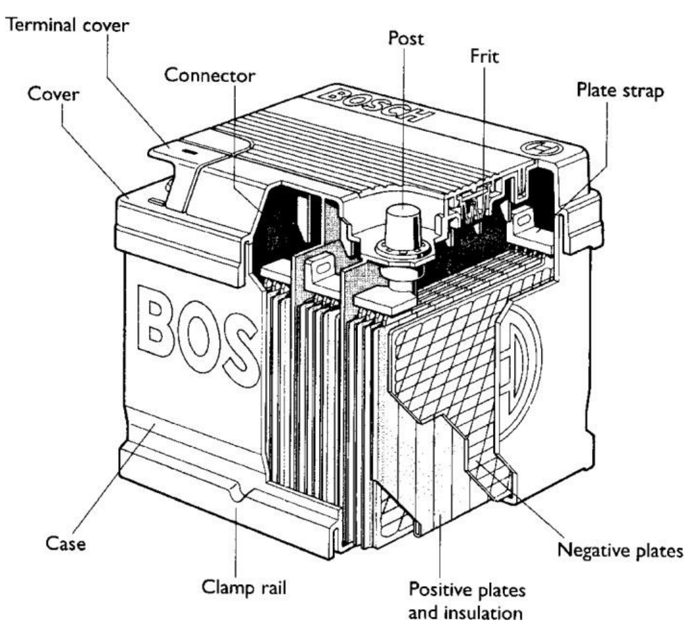
DEPARTMENT OF MECHATRONICS

UNIT I TYPES OF BATTERIES

Lead-acid batteries

Construction

Even after well over 100 years of development and much promising research into other techniques of energy storage, the lead-acid battery is still the best choice for motor vehicle use. This is particularly so when cost and energy density are taken into account. Incremental changes over the years have made the sealed and maintenance-free battery now in common use very reliable and long lasting. This may not always appear to be the case to some end-users, but note that quality is often related to the price the customer pays. Many bottom-of-the-range cheap batteries, with a 12 month guarantee, will last for 13 months! The basic construction of a nominal 12 V leadacid battery consists of six cells connected in series. Each cell, producing about 2 V, is housed in an individual compartment within a polypropylene, or similar, case. Figure shows a cut-away battery showing the main component parts.



Lead Acid Battery

The active material is held in grids or baskets to form the positive and negative plates. Separators made from a microporous plastic insulate these plates from each other. The grids, connecting strips and the battery posts are made from a lead alloy. For many years this was lead antimony (PbSb) but this has now been largely replaced by lead calcium (PbCa). The newer materials cause less gassing of the electrolyte when the battery is fully charged. This has been one of the main reasons why sealed batteries became feasible, as water loss is considerably reduced. However, even modern batteries described as sealed do still have a small vent to stop the pressure build-up due to the very small amount of gassing. A further requirement of sealed batteries is accurate control of charging voltage.

Battery rating

In simple terms, the characteristics or rating of a particular battery are determined by how much current it can produce and how long it can sustain this current. The rate at which a battery can produce current is determined by the speed of the chemical reaction.

This in turn is determined by a number of factors:

- Surface area of the plates.
- Temperature.
- Electrolyte strength.
- Current demanded.

The actual current supplied therefore determines the overall capacity of a battery. The rating of a battery has to specify the current output and the time.

Ampere hour capacity

This is now seldom used but describes how much current the battery is able to supply for either 10 or 20 hours. The 20-hour figure is the most common. For example, a battery quoted as being 44 Ah (ampere-hour) will be able, if fully charged, to supply 2.2 A for 20 hours before being completely discharged (cell voltage above 1.75 V).

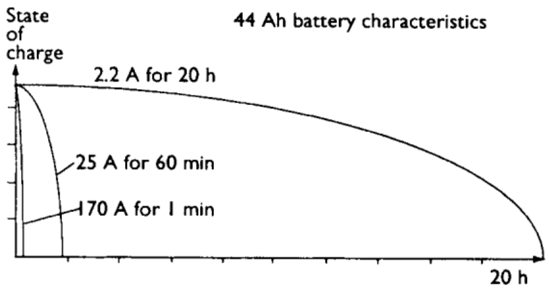
Reserve capacity

A system used now on all new batteries is reserve capacity. This is quoted as a time in minutes for which the battery will supply 25 A at 25 ° C to a final voltage of 1.75 V per cell. This is used to give an indication of how long the battery could run the car if the charging system was not working. Typically, a 44 Ah battery will have a reserve capacity of about 60 minutes.

Cold cranking amps

Batteries are given a rating to indicate performance at high current output and at low temperature. A typical value of 170 A means that the battery will supply this current for one minute at a temperature of _18 ° C, at which point the cell voltage will fall to 1.4 V (BS – British Standards).

Note that the overall output of a battery is much greater when spread over a longer time. As mentioned above, this is because the chemical reaction can only work at a certain speed. Figure 5.3 shows the above three discharge characteristics and how they can be compared.



Battery discharge characteristics compared

Maintenance and charging

Maintenance

By far the majority of batteries now available are classed as 'maintenance free'. This implies that little attention is required during the life of the battery. Earlier batteries and some heavier types do, however, still require the electrolyte level to be checked and topped up periodically.

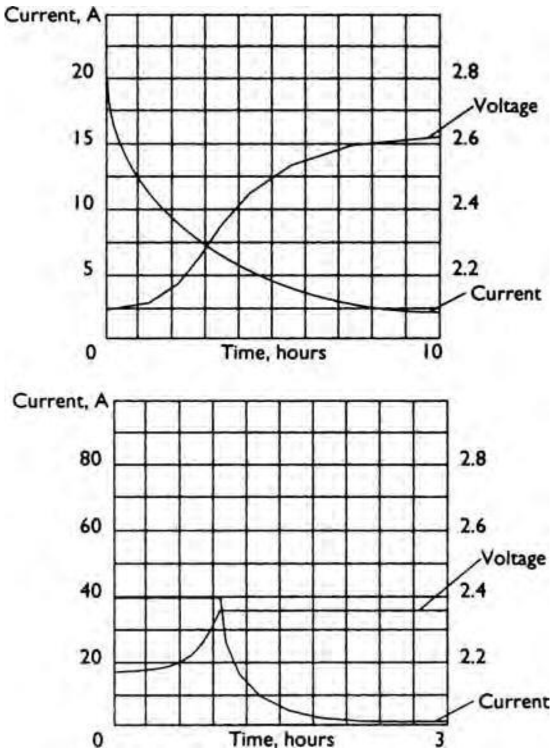
Battery posts are still a little prone to corrosion and hence the usual service of cleaning with hot water if appropriate and the application of petroleum jelly or proprietary terminal grease is still recommended. Ensuring that the battery case and, in particular, the top remains clean, will help to reduce the rate of self-discharge.

The state of charge of a battery is still very important and, in general, it is not advisable to allow the state of charge to fall below 70% for long periods as the sulphate on the plates can harden, making recharging difficult. If a battery is to be stored for a long period (more than a few weeks , then it must be recharged every so often to prevent it from becoming sulphated. Recommendations vary but a recharge every six weeks is a reasonable suggestion.

Battery Charging

The recharging recommendations of battery manufacturers vary slightly. The following methods, however, are reasonably compatible and should not cause any problems. The recharging process must 'put back' the same ampere-hour capacity as was used on discharge plus a bit more to allow for losses. It is therefore clear that the main question about charging is not how much, but at what rate.

The old recommendation was that the battery should be charged at a tenth of its ampere-hour capacity for about 10 hours or less. This is assuming that the ampere-hour capacity is quoted at the 20 hour rate, as a tenth of this figure will make allowance for the charge factor. This figure is still valid, But as ampere-hour capacity is not always used nowadays, a different method of deciding the rate is necessary. One way is to set a rate at 1/16 of the reserve capacity, again for up to 10 hours. The final suggestion is to set a charge rate at 1/40 of the cold start

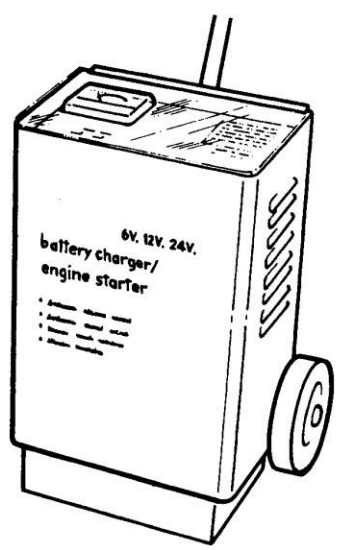


Two ways of charging a battery showing the relationship between charging voltage and charging current

Charging method	Notes
Constant voltage	Will recharge any battery in 7 hours or less without any risk of overcharging (14.4 V maximum).
Constant current	Ideal charge rate can be estimated as: 1/10 of Ah capacity, 1/16 of reserve capacity or 1/40 of cold start current (charge time of 10–12 hours or pro rata original state).

Boost charging   At no more than five times the ideal rate, a battery can be brought up to about 70% of charge in about one hour.

Performance figure, also for up to 10 hours. Clearly, if a battery is already half charged, half the time is required to recharge to full capacity. The above suggested charge rates are to be recommended as the best way to prolong battery life. They do all, however, Imply a constant current charging source. A constant voltage charging system is often the best way to charge a battery. This implies that the charger, an alternator on a car for example, is held at a constant level and the state of charge in the battery will determine how much current will flow. This is often the fastest way to recharge a flat battery. The two ways of charging are represented in Figure. This shows the relationship between charging voltage and the charging current. If a constant voltage of less than 14.4 V is used then it is not possible to cause excessive gassing and this method is particularly appropriate for sealed batteries. Boost charging is a popular technique often applied in many workshops. It is not recommended as the best method but, if correctly administered and not repeated too often, is suitable for most batteries. The key to fast or boost charging is that the battery temperature should not exceed 43° C. With sealed batteries it is particularly important not to let the battery create excessive gas in order to prevent the build-up of pressure. A rate of about five times the 'normal' charge setting will bring the battery to 78–80% of its full capacity within approximately one hour. Table summarizes the charging techniques for a lead-acid battery. Figure shows a typical battery charger.



Battery Charger

Various Tests on Battery

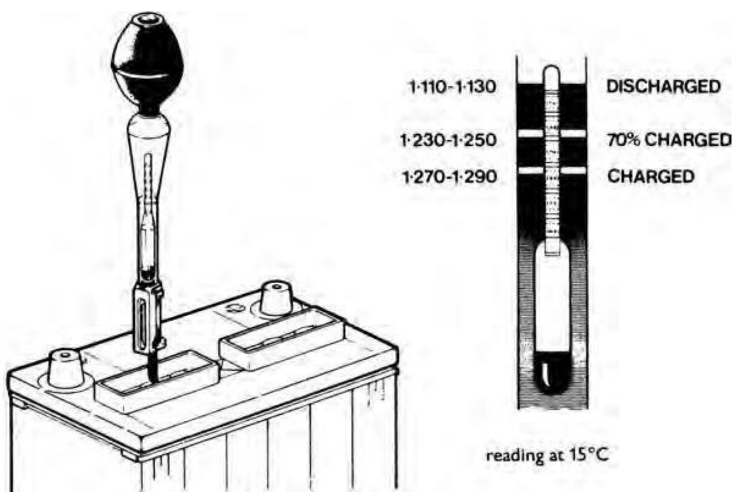
For testing the state of charge of a non-sealed type of battery, a hydrometer can be used, as shown in Figure The hydrometer comprises a syringe that draws electrolyte from a cell, and a float that will float at a particular depth in the electrolyte according to its density. The density or specific gravity is then read from the graduated scale on the float. A fully charged cell should show 1.280, 1.200 when half charged and 1.130 if discharged.

Most vehicles are now fitted with maintenance free batteries and a hydrometer cannot be used to find the state of charge. This can only be determined. From the voltage of the battery, as given in Table. An accurate voltmeter is required for this test. A heavy-duty (HD) discharge tester as shown in Figure is an instrument consisting of a low-value resistor and a voltmeter connected to a pair of heavy test prods. The test prods are firmly pressed on to the battery terminals. The voltmeter reads the voltage of the battery on heavy discharge of 200–300 A. Assuming a battery to be in a fully charged condition, a serviceable battery should read about 10V for a period of about 10 s. A sharply falling battery voltage to below 3 V indicates an unserviceable cell. Note also if any cells are gassing, as this indicates a short circuit. A zero or extremely low reading can indicate an open circuit cell. When using the HD tester, the following precautions must be observed:

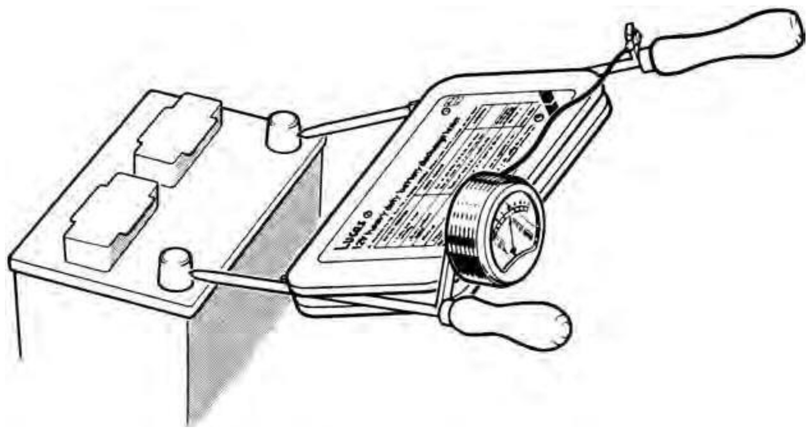
● Blow gently across the top of the battery to remove flammable gases.

● The test prods must be positively and firmly pressed into the lead terminals of the battery to minimize sparking.

● It should not be used while a battery is on charge.



Hydrometer Test of a Battery



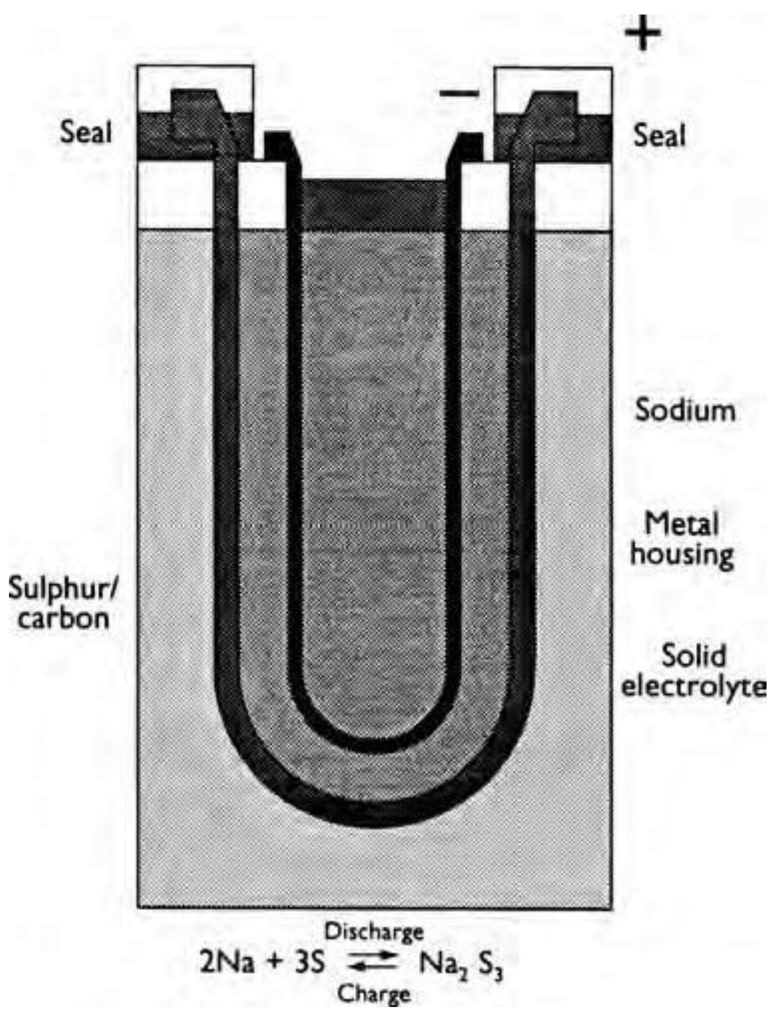
Heavy Duty Discharge Test

State of charge of a battery

Battery volts at 20 ° C	State of charge
12.0	Discharged (20% or less)
12.3	Half charged (50%)
12.7	Charged (100%)

SODIUM SULPHUR BATTERY

Much research is underway to improve on current battery technology in order to provide a greater energy density for electric vehicles. A potential major step forwards however the sodium sulphur battery, which has now reached production stage. Sodium-sulphur batteries have recently reached the production stage and, in common with the other types listed, have much potential; however, all types have specific drawbacks. For example, storing and carrying hydrogen is one problem of fuel cells.



The sodium-sulphur or NaS battery consists of a cathode of liquid sodium into which is placed a current collector. This is a solid electrode of β-alumina. A metal can that is in contact with the anode (a sulphur electrode) surrounds the whole assembly. The major problem with this system is that the running temperature needs to be 300–350 ° C. A heater rated at a few hundred watts forms part of the charging circuit. This maintains the battery temperature when the vehicle is not running. Battery temperature is maintained when in use due to I^2 R losses in the battery.

Each cell of this battery is very small, using only about 15 g of sodium. This is a safety feature because, if the cell is damaged, the sulphur on the outside will cause the potentially dangerous sodium to be converted into polysulphides – which are comparatively harmless. Small cells also have the advantage that they can be distributed around the car. The capacity of each cell is about 10 Ah. These cells fail in an open circuit condition and hence this must be taken into account, as the whole string of cells used to create the required voltage would be rendered inoperative. The output voltage of each cell is about 2V. A problem still to be overcome is the casing material, which is prone to fail due to the very corrosive nature of the sodium. At present, an expensive chromized coating is used.

This type of battery, supplying an electric motor, is becoming a competitor to the internal combustion engine. The whole service and charging infrastructure needs to develop but looks promising. It is estimated that the cost of running an electric vehicle will be as little as 15% of the petrol version, which leaves room to absorb the extra cost of production.

Alkaline batteries (Nickel – Cadmium Battery)

Lead-acid batteries traditionally required a considerable amount of servicing to keep them in good condition, although this is not now the case with the advent of sealed and maintenance-free batteries.

However, when a battery is required to withstand a high rate of charge and discharge on a regular basis, or is left in a state of disuse for long periods, the lead-acid cell is not ideal. Alkaline cells on the other hand require minimum maintenance and are far better able to withstand electrical abuse such as heavy discharge and over-charging.

The disadvantages of alkaline batteries are that they are more bulky, have lower energy efficiency and are more expensive than a lead-acid equivalent. When the lifetime of the battery and servicing requirements are considered, the extra initial cost is worth it for some applications. Bus and coach companies and some large goods-vehicle operators have used alkaline batteries.

Alkaline batteries used for vehicle applications are generally the nickel-cadmium type, as the other main variety (nickel-iron) is less suited to vehicle use. The main components of the nickel-cadmium – or Nicad – cell for vehicle use are as follows:

_ positive plate – nickel hydrate (NiOOH);

_ negative plate – cadmium (Cd);

_ electrolyte – potassium hydroxide (KOH) and water (H2O).

The process of charging involves the oxygen moving from the negative plate to the positive plate, and the reverse when discharging. When fully charged, the negative plate becomes pure cadmium and the positive plate becomes nickel hydrate. A chemical equation to represent this reaction is given next but notes that this is simplifying a more complex reaction.

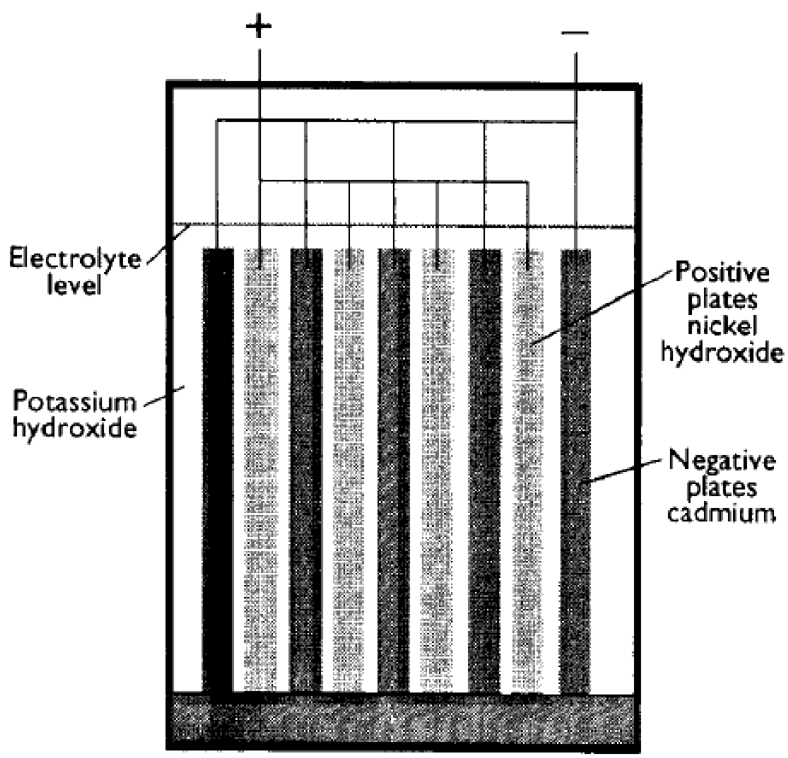


Figure 5.10 Simplified representation of a Nicad alkaline battery cell

$$2NiOOH + Cd + 2H_2O + KOH \leftrightarrow$$
$$2Ni(OH)_2 + CdO_2 + KOH$$

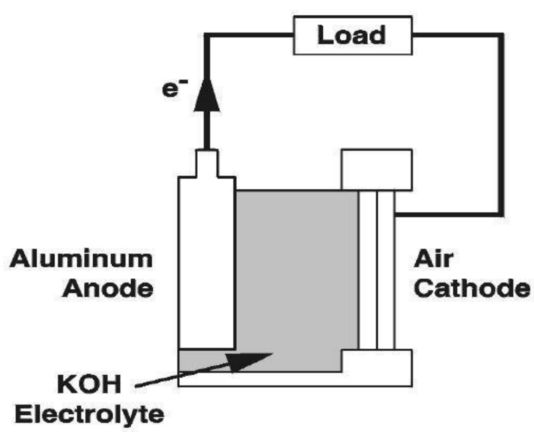
The $2H_2O$ is actually given off as hydrogen (H) and oxygen (O_2) as gassing takes place all the time during charge. It is this use of water by the cells that indicates they are operating, as will have been noted from the equation. The electrolyte does not change during the reaction. This means that a relative density reading will not indicate the state of charge.These batteries do not suffer from over-charging because once the cadmium oxide has changed to cadmium, no further reaction can take place. The cell voltage of a fully charged cell is 1.4V but this falls rapidly to 1.3 V as soon as discharge starts. The cell is discharged at a cell voltage of 1.1V.

ALUMINIUM–AIR BATTERY

Aluminium–air batteries or Al–air batteries produce electricity from the reaction of oxygen in the air with aluminium. They have one of the highest energy densities of all batteries, but they are not widely used because of problems with high anode cost and byproduct removal when using traditional electrolytes and this has restricted their use to mainly military applications. However, an electric vehicle with aluminium batteries has the potential for up to eight times the range of a lithium-ion battery with a significantly lower total weight.

Aluminium–air batteries are primary cells; i.e., non-rechargeable. Once the aluminium anode is consumed by its reaction with atmospheric oxygen at a cathode immersed in a water-based electrolyte to form hydrated aluminium oxide, the battery will no longer produce electricity. However, it is possible to mechanically recharge the battery with new aluminium anodes made from recycling the hydrated aluminium oxide. Such recycling would be essential if aluminium–air batteries are to be widely adopted.

Electrochemistry



The anode oxidation half-reaction is

$Al + 3OH^{-} \rightarrow Al(OH)_3 + 3e^{-}$ −2.31 V.

The cathode reduction half-reaction is

$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^- + 0.40\ V.$$

The total reaction is

$$4Al + 3O_2 + 6H_2O \rightarrow 4Al(OH)_3 + 2.71\ V.$$

About 1.2 volts potential difference is created by these reactions, and is achievable in practice when potassium hydroxide is used as the electrolyte. Saltwater electrolyte achieves approximately 0.7 volts per cell.

NICKEL–METAL HYDRIDE BATTERY

A nickel–metal hydride battery, abbreviated NiMH or Ni–MH, is a type of rechargeable battery. Its chemical reactions are somewhat similar to the largely obsolete nickel–cadmium cell (NiCd). NiMH use positive electrodes of nickel oxyhydroxide (NiOOH), like the NiCd, but the negative electrodes use a hydrogen-absorbing alloy instead of cadmium, being in essence a practical application of nickel–hydrogen battery chemistry. A NiMH battery can have two to three times the capacity of an equivalent size NiCd, and their energy density approaches that of a lithium-ion cell.

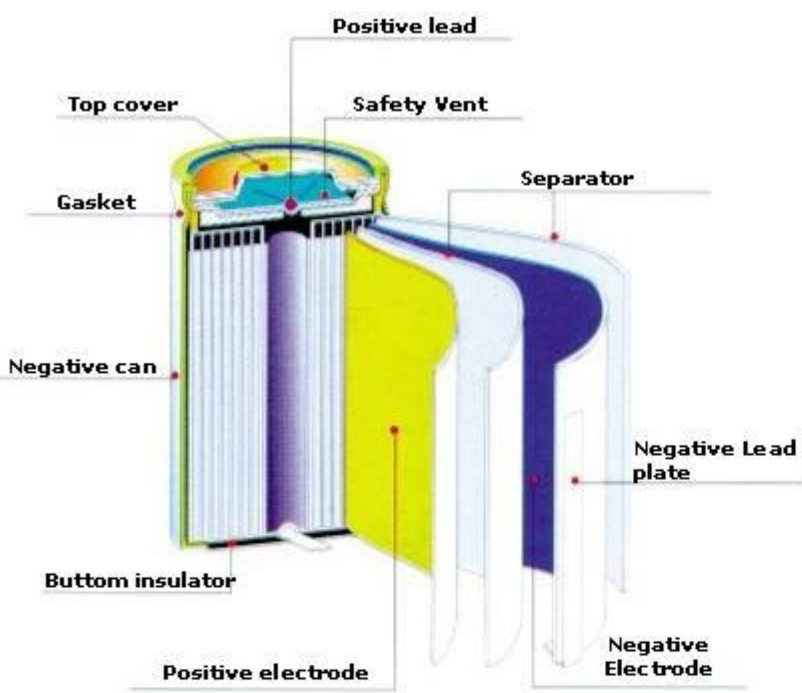
The typical specific energy for small NiMH cells is about 100 W·h/kg, and for larger NiMH cells about 75 W·h/kg (270 kJ/kg). This is significantly better than the typical 40–60 W·h/kg for NiCd, and similar to the 100–160 W·h/kg for lithium-ion batteries. NiMH has a volumetric energy density of about 300 W·h/L (1,080 MJ/m3), significantly better than NiCd at 50–150 W·h/L, and about the same as lithium-ion at 250–360 W·h/L.

NiMH batteries have replaced NiCd for many roles, notably small rechargeable batteries. NiMH batteries are very common for AA (penlight-size) batteries, which have nominal charge capacities (C) of 1.1–2.8 A·h at 1.2 V, measured at the rate that discharges the cell in five hours. Useful discharge capacity is a decreasing function of the discharge rate, but up to a rate of around 1×C (full discharge in one hour), it does not differ significantly from the nominal capacity. NiMH batteries normally operate at 1.2 V per cell, somewhat lower than conventional 1.5 V cells, but will operate most devices designed for that voltage.

Electrochemistry

The negative electrode reaction occurring in a NiMH cell is:

$$H_2O + M + e^- \rightleftharpoons OH^- + MH$$



The charge reaction is read left-to-right and the discharge reaction is read right-to-left.

On the positive electrode, nickel oxyhydroxide, NiO(OH), is formed:

$$Ni(OH)_2 + OH^- \rightleftharpoons NiO(OH) + H_2$$

The "metal" M in the negative electrode of a NiMH cell is actually an intermetallic compound. Many different compounds have been developed for this application, but those in current use fall into two classes. The most common is AB_5 , where A is a rare earth mixture of lanthanum, cerium, neodymium, praseodymium and B is nickel, cobalt, manganese, and/or aluminium. Very few cells use higher-capacity negative electrode materials based on AB_2 compounds, where A is titanium and/or vanadium and B is zirconium or nickel, modified with chromium, cobalt, iron, and/or manganese, due to the reduced life performances. Any of these compounds serve the same role, reversibly forming a mixture of metal hydride compounds.

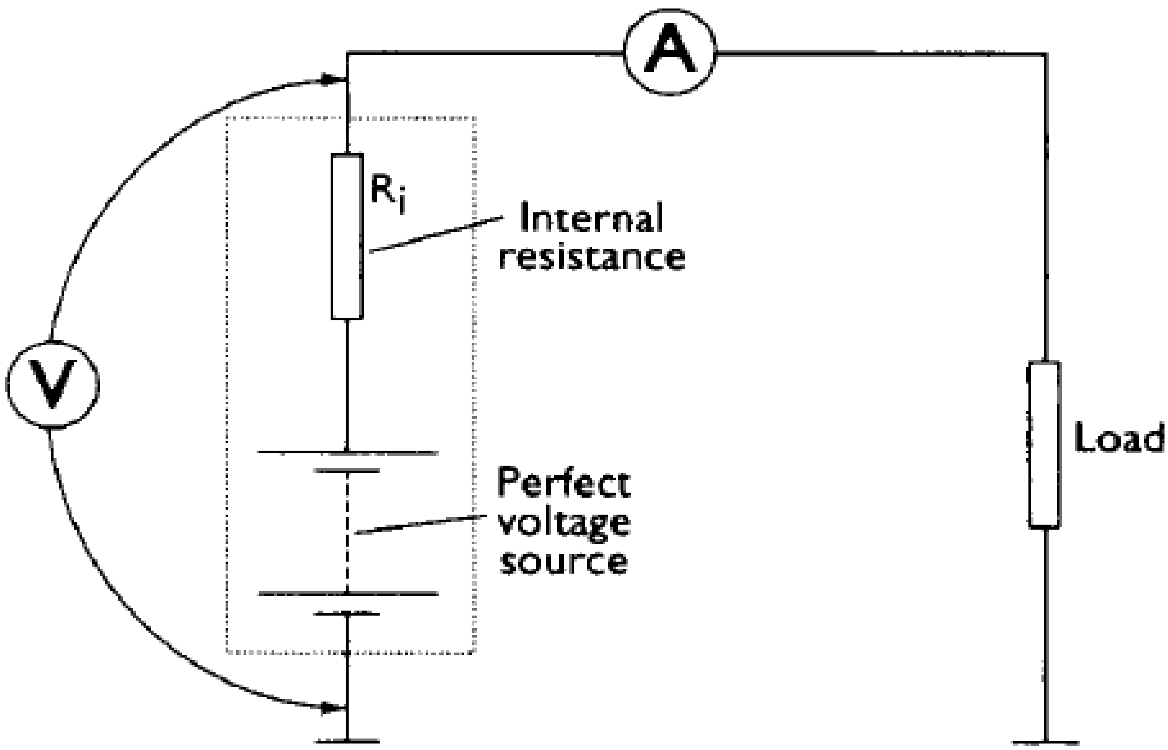
When overcharged at low rates, oxygen produced at the positive electrode passes through the separator and recombines at the surface of the negative. Hydrogen evolution is suppressed and the charging energy is converted to heat. This process allows NiMH cells to remain sealed in normal operation and to be maintenance-free.

NiMH cells have an alkaline electrolyte, usually potassium hydroxide. For separation hydrophilic polyolefin nonwovens are used.

Characteristics of Battery

Internal Resistance

- Temperature and state of charge affect the internal resistance of a battery.
- The internal resistance can also be used as an indicator of battery condition –the lower the figure, the better the condition.



Efficiency

- The efficiency of a battery can be calculated in two ways, either as the ampere-hour efficiency or the power efficiency.

Ahr (or charge) efficiency

- In an ideal world a battery would return the entire charge put into it, in which case the amp hour efficiency is 100%. However, no battery does; its charging efficiency is less than 100%.

Energy efficiency

It is defined as the ratio of electrical energy supplied by a battery to the amount of electrical energy required to return it to the state before discharge.

Self-discharge

- All batteries suffer from self-discharge, which means that even without an external circuit the state of charge is reduced.
- The rate of discharge is of the order of 0.2–1% of the Ah capacity per day. This increases with temperature and the age of the battery.
- The chemical process
- Leakage current across top of the battery

Specific energy

- Specific energy is the amount of electrical energy stored for every kilogram of battery mass. It has units of $Wh.kg^{-1}$

Energy density

- Energy density is the amount of electrical energy stored per cubic meter of battery volume. It normally has units of $Wh.m^{-3}$

Specific power

- Specific power is the amount of power obtained per kilogram of battery. It is a highly variable and rather anomalous quantity, since the power given out by the battery depends far more upon the load connected to it than the battery itself.

Battery temperature, heating and cooling needs

- Although most batteries run at ambient temperature, some run at higher temperatures and need heating to start with and then cooling when in use.
- In others, battery performance drops off at low temperatures, which is undesirable, but this problem could be overcome by heating the battery.

Battery life and number of deep cycles

- Most rechargeable batteries will only undergo a few hundred deep cycles to 20% of the battery charge.
- However, the exact number depends on the battery type, and also on the details of the battery design, and on how the battery is used.

Starter Motor types , construction and characteristics

Reducing electrical and mechanical stress at start-up, the starting current of an AC motor can vary from 3 to 7times the nominal current. This is because a large amount of energy is required to magnetise the motor enough to overcome the inertia the system has at standstill. The high current drawn from the network can cause problems such as voltage drop, high transients and, in some cases, uncontrolled shutdown. High starting current also causes great mechanical stress on the motor's rotor bars and windings, and can affect the driven equipment and the foundations. Several starting methods exist, all aiming to reduce these stresses.

The load, the motor and the supply network determine the most appropriate starting method. When selecting and dimensioning the starting equipment and any protective devices, the following factors must be taken into account:

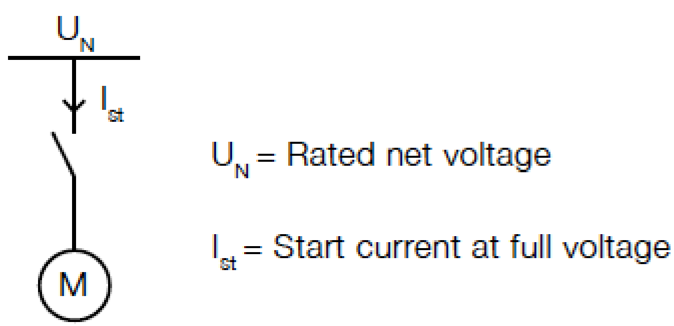
—— The voltage drop in the supply network when starting the motor

—— The required load torque during start

—— The required starting time

Direct-on-line (DOL) start

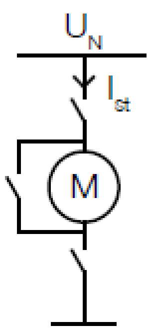
Direct on line starting is suitable for stable supplies and mechanically stiff and well-dimensioned shaft systems. It is the simplest, cheapest and most common starting method. Starting equipment for small motors that do not start and stop frequently is simple, often consisting of a hand operated motor protection circuit breaker. Larger motors and motors that start and stop frequently, or have some kind of control system, normally use a direct-on-line starter which can consist of a contactor plus overload protection, such as a thermal relay.



Star-Delta (Y/D) starting

Most low voltage motors can be connected to run at either 400 V with delta connection or at 690 V with star connection. This flexibility can also be used to start the motor with a lower voltage. Star/delta connection gives a low starting current of only about one-third of that during direct-on-line starting, although this also reduces the starting torque to about 25%.

The motor is started with Y-connection and accelerated as far as possible, then switched to D-connection. This method can only be used with induction motors delta connected for the supply voltage.



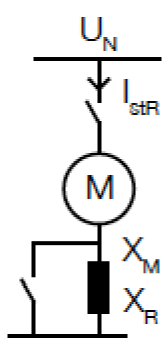
U_N = Rated net voltage

I_{st} = Start current at full voltage

Reactor start

By connecting a coil with an iron core (a reactor) in series with the motor during start, the starting current is limited in proportion with the voltage. However, this also means a substantial (quadratic) reduction in the available starting torque.

The advantage of this method is its low cost in comparison with other methods.



U_N = Rated net voltage

I_{st} = Start current at full voltage

I_{stR} = Start current at red. voltag

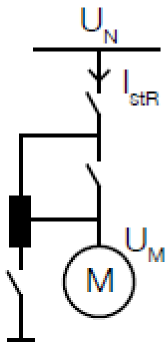
X_M = Motor reactance

X_R = Reactor reactance

$$I_{stR} = \left(\frac{U_N}{X_M + X_R}\right) \qquad T_{stR} = \left(\frac{I_{stR}}{I_{st}}\right)^2 T_{st}$$

Auto transformer start

The effect of auto transformer start is similar to that of reactor start. Using a transformer to limit the voltage reduces the starting current and the torque, but less so than the reactor start. The method is more expensive than reactor start.



U_N = Rated net voltage

I_{st} = Start current at full voltage

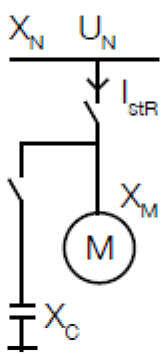
I_{stR} = Start current at red. voltage

U_M = Motor voltage

$$I_{stR} = \left(\frac{U_M}{U_N}\right)^2 I_{st} \qquad T_{stR} = \left(\frac{I_{stR}}{I_{st}}\right) T_{st}$$

Capacitor start

By storing the power required for magnetisation in capacitor banks, it is possible to start with full starting torque without disturbing the network. To avoid over-compensation, the capacitor bank must be uncoupled after start-up. The disadvantages of this method are the high cost, and the large space requirement of the capacitor banks.



U_N = Rated net voltage

I_{stR} = Start current w/capacitor

X_N = Net reactance

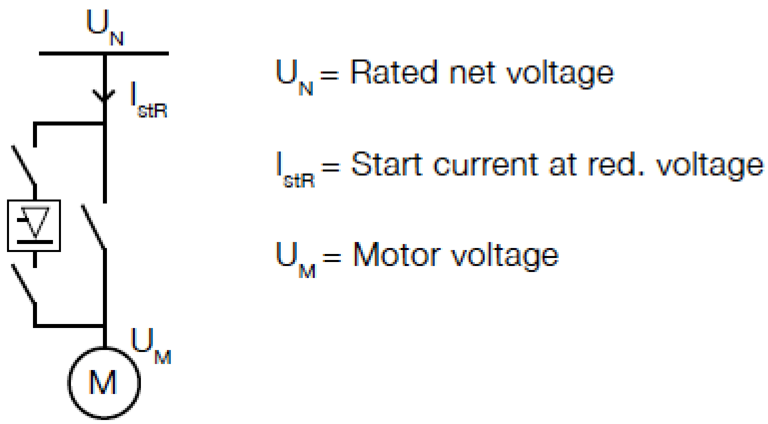
X_C = Capacitor bank reactance

X_M = Motor short circuit reactance

$$I_{stR} = \frac{U_N}{\sqrt{3}\left(X_N + \dfrac{X_C X_M}{X_C + X_M}\right)} \qquad T_{stR} = \frac{\dfrac{X_C X_M}{X_C + X_M}}{X_N + \dfrac{X_C X_M}{X_C + X_M}} T_{st}$$

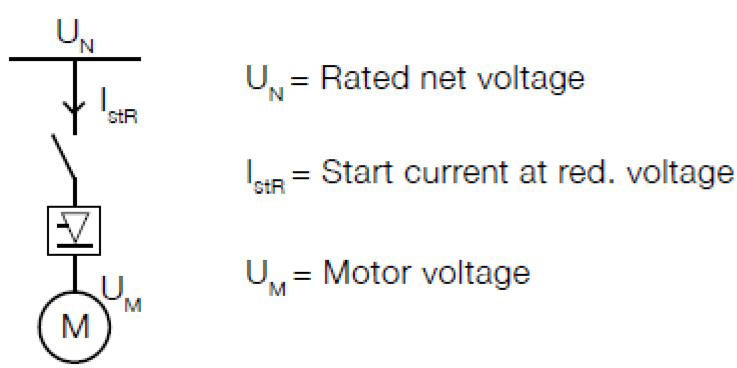
Soft starters

Soft starters are based on semiconductors, which, via a power circuit and a control circuit, initially reduces the motor voltage, resulting in lower motor torque. During the starting process, the soft starter progressively increases the motor voltage so that the motor becomes strong enough to accelerate the load to rated speed without causing torque or current peaks. Soft starters can also be used to control the stopping of a process. Soft starters are less costly than frequency converters but like frequency converters, they may inject harmonic currents into the grid, disrupting other processes.



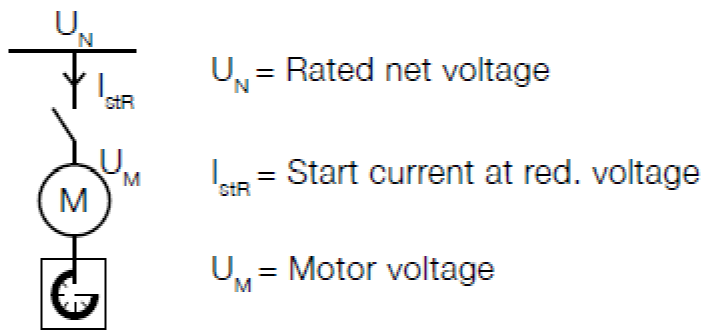
Frequency converter start

Although a frequency converter is designed for continuous feeding of motors, it can also be used for start-up only. The frequency converter enables low starting current because the motor can produce rated torque at rated current from zero to full speed. As the price of frequency converters continues to drop, they are increasingly being used in applications where soft starters would previously have been used. However in most cases they are still more expensive than soft starters, and like these, they inject harmonic currents into the network.



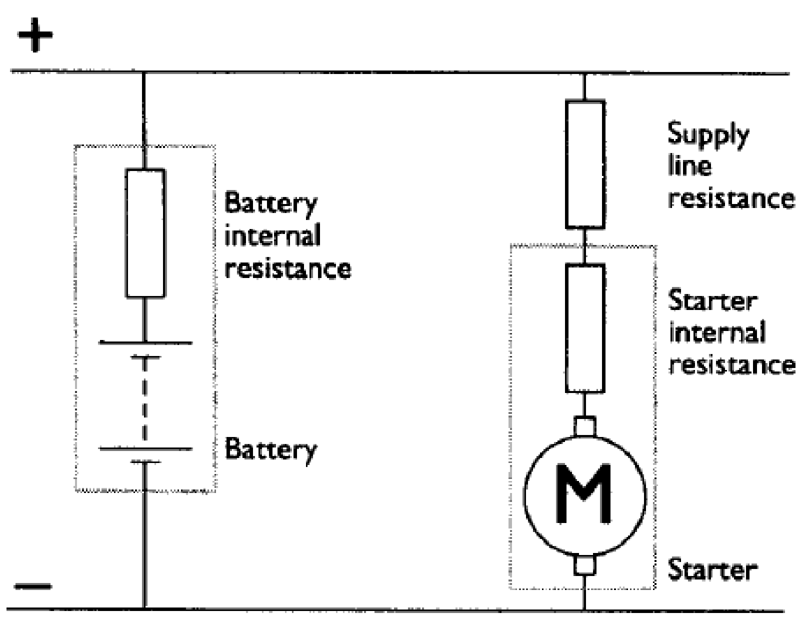
Rheostat starting

Rheostat starting can only be used with slip ring motors. On these motors, the resistance of the rotor circuits can be increased with an external resistor. This method is usually chosen when the supply net is weak and the required starting torque and moment of inertia are very high. By switching in the additional resistances in steps, normally 4 to 7 steps, the desired acceleration torque can be obtained. The normal DOL starting equipment also required.



Requirements of Starter Motor

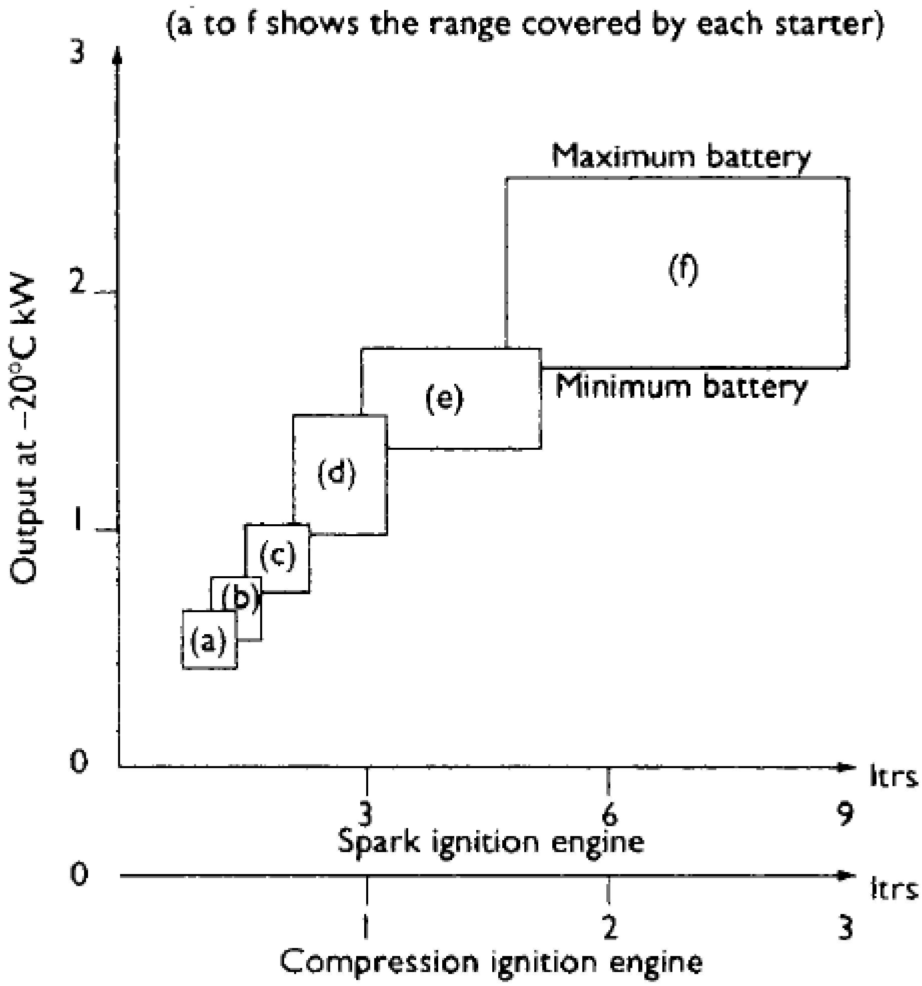
As a very general guide the stalled (locked) starter torque required per litre of engine capacity at the starting limit temperature is as shown in Table. A greater torque is required for engines with a lower number of cylinders due to the greater piston displacement per cylinder. This will determine the peak torque values. The other main factor is compression ratio.



Equivalent circuit for a starter system

Torque required for various engine sizes

Engine cylinders	Torque per litre [Nm]
2	12.5
4	8.0
6	6.5
8	6.0
12	5.5



To illustrate the link between torque and power, we can assume that, under the worst conditions (_20 ° C), a four-cylinder 2-litre engine requires 480 Nm to overcome static friction and 160 Nm to maintain the minimum cranking speed of 100 rev/min. With a starter pinion-to-ring gear ratio of 10 : 1,the motor must therefore, be able to produce a maximum stalled torque of 48 Nm and a driving torque of 16 Nm. This is working on the assumption that stalled torque is generally three to four times the cranking torque.

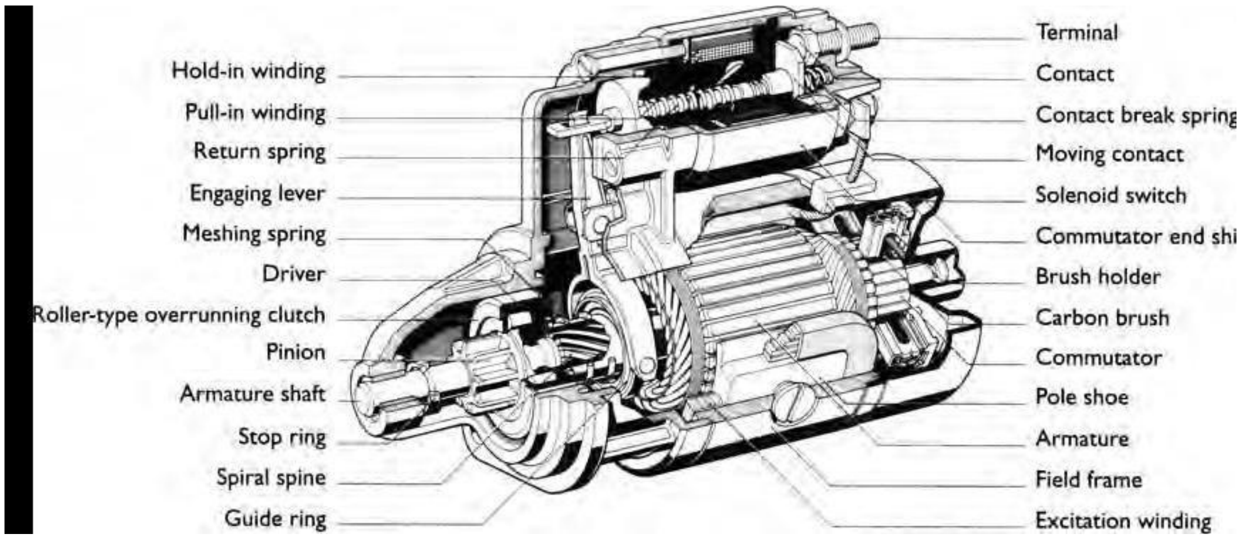
Starter drive mechanisms

Inertia starters

In all standard motor vehicle applications it is necessary to connect the starter to the engine ring gear only during the starting phase. If the connection remained permanent, the excessive speed at which the starter would be driven by the engine would destroy the motor almost immediately. The inertia type of starter motor has been the technique used for over 80 years, but is now becoming redundant. The starter shown in Figure shows the Lucas M35J type. It is a four-pole, four-brush machine and was used on small to medium-sized petrol engined vehicles. It is capable of producing 9.6 Nm with a current draw of 350 A. The M35J uses a face-type commutator and axially aligned brush gear. The fields are wave wound and are earthed to the starter yoke. The starter engages with the flywheel ring gear by means of a small pinion. The toothed pinion and a sleeve splined on to the armature shaft are threaded such that when the starter is operated, via a remote relay, the armature will cause the sleeve to rotate inside the pinion. The pinion remains still due to its inertia and, because of the screwed sleeve rotating inside it, the pinion is moved to mesh with the ring gear.

When the engine fires and runs under its own power, the pinion is driven faster than the armature shaft. This causes the pinion to be screwed back along the sleeve and out of engagement with the flywheel. The main spring acts as a buffer when the pinion first takes up the driving torque and also acts as a buffer when the engine throws the pinion back out of mesh.

One of the main problems with this type of starter was the aggressive nature of the engagement. This tended to cause the pinion and ring gear to wear prematurely. In some applications the pinion tended to fall out of mesh when cranking due to the engine almost, but not quite, running. The pinion was also prone to seizure often due to contamination by dust from the clutch. This was often compounded by application of oil to the pinion mechanism, which tended to attract even more dust and thus prevent engagement. The pre-engaged starter motor has largely overcome these problems.



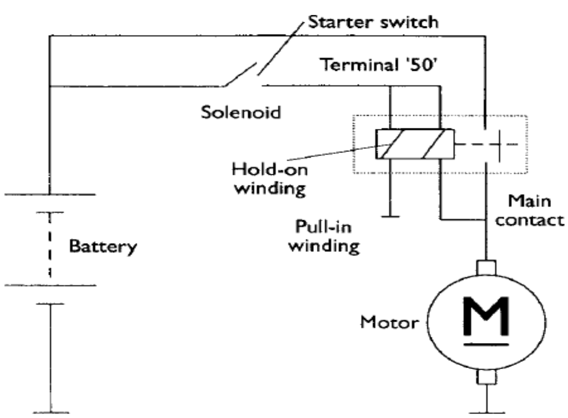
Pre-engaged starters

Pre-engaged starters are fitted to the majority of vehicles in use today. They provide a positive engagement with the ring gear, as full power is not applied until the pinion is fully in mesh. They prevent premature ejection as the pinion is held into mesh by the action of a solenoid. A one-way clutch is incorporated into the pinion to prevent the starter motor being driven by the engine. One example of a pre-engaged starter in common use is shown in Figure.

The basic operation of the pre-engaged starter is as follows. When the key switch is operated, a supply is made to terminal 50 on the solenoid. This causes two windings to be energized, the hold-on winding and the pull-in winding. Note that the pull-in winding is of very low resistance and hence a high current flows. This winding is connected in series with the motor circuit and the current flowing will allow the motor to rotate slowly to facilitate engagement. At the same time, the magnetism created in the solenoid attracts the plunger and, via an operating lever, pushes the pinion into mesh with the flywheel ring gear. When the pinion is fully in mesh the plunger, at the end of its travel, causes a heavy-duty set of copper contacts to close. These contacts now supply full battery power to the main circuit of the starter motor. When the main contacts are closed, the pull-in winding is effectively switched off due to equal voltage supply on both ends. The hold-on winding holds the plunger in position as long as the solenoid is supplied from the key switch. When the engine starts and the key is released, the main supply is removed and the plunger and pinion return to their rest positions under spring tension. A lost motion spring located on the plunger ensures that the main contacts open before the pinion is retracted from mesh.

During engagement, if the teeth of the pinion hit the teeth of the flywheel (tooth to tooth abutment), the main contacts are allowed to close due to the engagement spring being compressed. This allows the motor to rotate under power and the pinion will slip into mesh.

Figure shows a sectioned view of a one-way clutch assembly. The torque developed by the starter is passed through the clutch to the ring gear. The purpose of this free-wheeling device is to prevent the starter being driven at an excessively high speed if the pinion is held in mesh after the engine has started. The clutch consists of a driving and driven member with several rollers between the two. The rollers are spring loaded and either wedge-lock the two members together by being compressed against the springs, or free-wheel in the opposite direction. Many variations of the pre-engaged starter are in common use, but all work on similar lines to the above description. The wound field type of motor has now largely been replaced by the permanent magnet version.



Starter circuit

Permanent magnet starters

Permanent magnet starters began to appear on production vehicles in the late 1980s. The two main advantages of these motors, compared with conventional types, are less weight and smaller size. This makes the permanent magnet starter a popular choice by vehicle manufacturers as, due to the lower lines of today's cars, less space is now available for engine electrical systems. The reduction in weight provides a contribution towards reducing fuel consumption.

The standard permanent magnet starters currently available are suitable for use on spark ignition engines up to about 2 litre capacity. They are rated in the region of 1kW. The principle of operation is similar in most respects to the conventional pre-engaged starter motor. The main difference being the replacement of field windings and pole shoes with high quality permanent magnets. The reduction in weight is in the region of 15% and the diameter of the yoke can be reduced by a similar factor.

Permanent magnets provide constant excitation and it would be reasonable to expect the speed and torque characteristic to be constant. However, due to the fall in battery voltage under load and the low resistance of the armature windings, the characteristic is comparable to series wound motors. In some cases, flux concentrating pieces or interpoles are used between the main magnets. Due to the warping effect of the magnetic field, this tends to make the characteristic curve very similar to that of the series motor.

Development by some manufacturers has also taken place in the construction of the brushes. A copper and graphite mix is used but the brushes are made in two parts allowing a higher copper content in the power zone and a higher graphite content in the commutation zone. This results in increased service life and a reduction in voltage drop, giving improved starter power.

For applications with a higher power requirement, permanent magnet motors with intermediate transmission have been developed. These allow the armature to rotate at a higher and more efficient speed whilst still providing the torque, due to the gear reduction. Permanent magnet starters with intermediate transmission are available with power outputs of about 1.7 kW and are suitable for spark ignition engines up to about 3 litres, or compression ignition engines up to about 1.6 litres. This form of permanent magnet motor can give a weight saving of up to 40%. The principle of operation is again similar to the conventional pre-engaged starter. The intermediate transmission, is of the epicyclic type. The sun gear is on the armature shaft and the planet carrier drives the pinion. The ring gear or annulus remains stationary and also acts as an intermediate bearing. This arrangement of gears gives a reduction ratio of about 5 : 1. This can be calculated by the formula:

Ratio=AS/S

where A =number of teeth on the annulus, and

S =number of teeth on the sun gear.

The annulus gear in some types is constructed from a high grade polyamide compound with mineral additives to improve strength and wear resistance.

The sun and planet gears are conventional steel. This combination of materials gives a quieter and more efficient operation.

Heavy vehicle starters

The subject area of this book is primarily the electrical equipment on cars. This short section is included for interest, hence further reference should be made to other sources for greater detail about heavy vehicle starters.

The types of starter that are available for heavy duty applications are as many and varied as the applications they serve. In general, higher voltages are used, which may be up to 110 V in specialist cases, and two starters may even be running in parallel for very high power and torque requirements.

Large road vehicles are normally 24 V and employ a wide range of starters. In some cases the design is simply a large and heavy duty version of the pre-engaged type discussed earlier. This starter may also be fitted with a thermal cut-out to prevent overheating damage due to excessive cranking. Rated at 8.5kW, it is capable of producing over 80 Nm torque at 1000 rev/min. Other methods of engaging the pinion include sliding the whole armature or pushing the pinion with a rod through a hollow armature. This type uses a solenoid to push the pinion into mesh via a rod through the centre of the armature. Sliding-armature-type starters work by positioning the field windings forwards from the main armature body, such that the armature is attracted forwards when power is applied. A trip lever mechanism will then only allow full power when the armature has caused the pinion to mesh.

Integrated starters

A device called a 'dynastart' was used on a number of vehicles from the 1930s through to the 1960s.This device was a combination of the starter and a dynamo. The device, directly mounted on the crankshaft, was a compromise and hence not very efficient. The method is now known as an Integrated Starter Alternator Damper (ISAD). It consists of an electric motor, which functions as a control element between the engine and the transmission, and can also be used to start the engine and deliver electrical power to the batteries and the rest of the vehicle systems. The electric motor replaces the mass of the flywheel. The motor transfers the drive from the engine and is also able to act as a damper/vibration absorber unit. The damping effect is achieved by a rotation capacitor. A change in relative speed between the rotor and the engine due to the vibration, causes one pole of the capacitor to be charged. The effect of this is to take the energy from the vibration. Using ISAD to start the engine is virtually noiseless, and cranking speeds of 700 rev/min are possible. Even at _25 ° C it is still possible to crank at about 400 rev/min. A good feature of this is that a stop/start function is possible as an economy and emissions improvement technique. Because of the high speed cranking, the engine will fire up in about 0.1–0.5 seconds. The motor can also be used to aid with acceleration of the vehicle. This feature could be used to allow a smaller engine to be used or to enhance the performance of a standard engine.

When used in alternator mode, the ISAD can produce up to 2 kW at idle speed. It can supply power at different voltages as both AC and DC. Through the application of intelligent control electronics, the ISAD can be up to 80% efficient. Citroën have used the ISAD system in a Xsara model prototype. The car can produce 150 Nm for up to 30 seconds, which is significantly more than the 135 Nm peak torque of the 1580 cc, 65 kW fuel injected version. Citroën call the system ‘Dynalto’. A 220 V outlet is even provided inside the car to power domestic electrical appliances!

Starter switches

An Ignition (or starter) switch is a switch in the control system of an internal combustion engined motor vehicle that activates the main electrical systems for the vehicle. Besides providing power to the starter solenoid and the ignition system components (including the engine control unit and ignition coil) it also usually switches on power to many "accessories" (radio, power windows, etc.). The ignition switch usually requires a key be inserted that works a lock built into the switch mechanism. It is frequently combined with the starter switch which activates the starter motor. The ignition locking system may be bypassed by disconnecting the wiring to the switch and manipulating it directly; this is known as hotwiring.

Regulation of output voltage

To prevent the vehicle battery from being overcharged the regulated system voltage should be kept below the gassing voltage of the lead-acid battery. A figure of 14.2 +/- 0.2 V is used for all 12 V charging systems. Accurate voltage control is vital with the ever-increasing use of electronic systems. It has also enabled the wider use of sealed batteries, as the possibility of over-charging is minimal. Figure 6.15 shows two common voltage regulators. Voltage regulation is a difficult task on a vehicle alternator because of the constantly changing engine speed and loads on the alternator. The output of an alternator without regulation would rise linearly in proportion with engine speed. Alternator output is also proportional to magnetic field strength and this, in turn, is proportional to the field current. It is the task of the regulator to control this field current in response to alternator output voltage. Figure 6.16 shows a flow chart which

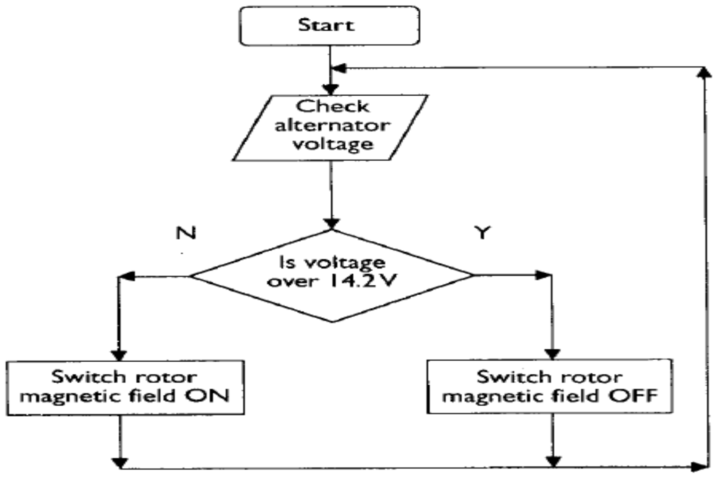


Figure 6.16 Action of the voltage regulator

Figure 6.19 Electronic voltage regulator

represents the action of the regulator, showing how the field current is switched off as output voltage increases and then back on again as output voltage falls. The abrupt switching of the field current does not cause abrupt changes in output voltage due to the very high inductance of the field (rotor) windings. In addition, the whole

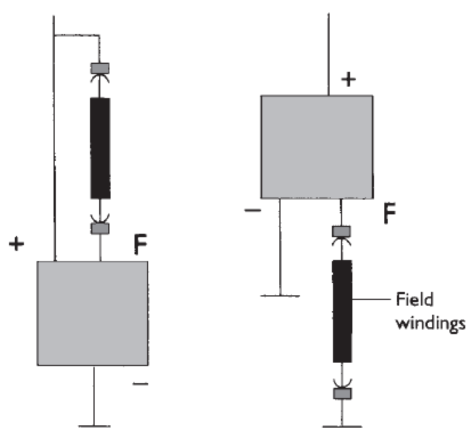


Figure 6.17 How the voltage regulator is incorporated in the field circuit

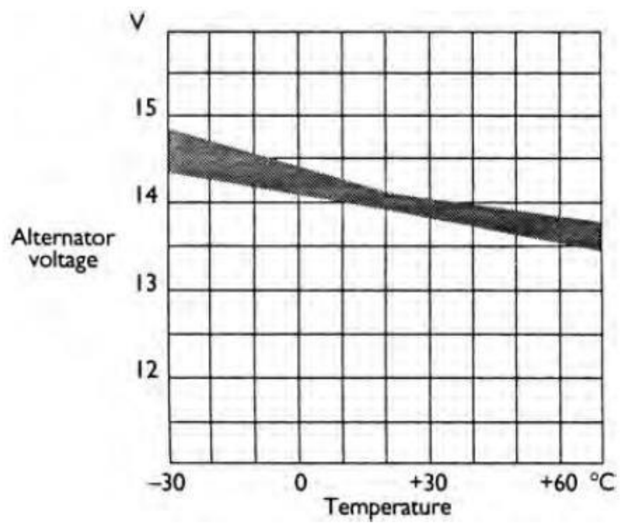


Figure 6.21 How the regulator response changes with temperature

switching process only takes a few milliseconds. Many regulators also incorporate some temperature compensation to allow

a higher charge rate in colder conditions and to reduce the rate in hot conditions.

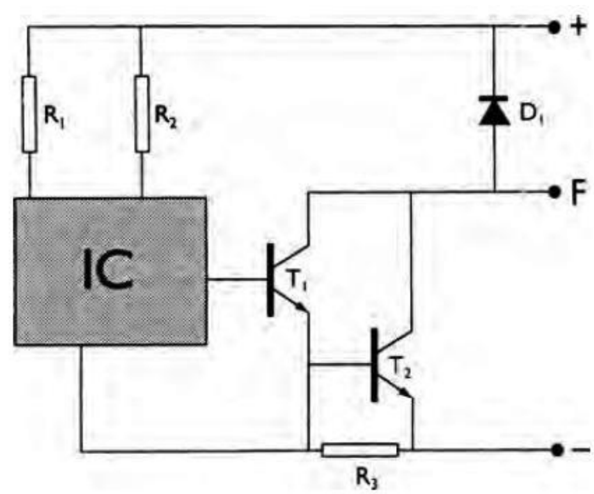


Figure 6.20 Hybrid IC regulator circuit

When working with regulator circuits, care must be taken to note ‘where’ the field circuit is interrupted. For example, some alternator circuits supply a constant feed to the field windings from the excitation diodes and the regulator switches the earth side. In other systems, one side of the field windings is will switch off, allowing T2 to switch back on and so the cycle will continue. The conventional diode, D1, absorbs the back EMF from the field windings and so prevents damage to the other components. Electronic regulators can be made to sense either the battery voltage, the machine voltage (alternator), or a combination of the two. Most systems in use at present tend to be machine sensed as this offers some protection against over-voltage in the event of the alternator being driven with the battery disconnected. Figure 6.20 shows the circuit of a hybrid integrated circuit (IC) voltage regulator. The hybrid system involves the connection of discrete components on a ceramic plate using film techniques. The main part of the regulator is an integrated circuit containing the sensing elements and temperature compensation components. The IC controls an output stage such as a Darlington pair. This technique produces a very compact device and, because of the low number of components and connections, is very reliable. Figure 6.21 is a graph showing how the IC regulator response changes with temperature. This change is important to ensure correct charging under ‘summer’ and ‘winter’ conditions. When a battery is cold, the electrolyte resistance increases. This means a higher voltage is necessary to cause the correct recharging current. Over-voltage protection is required in some applications in order to prevent damage to electronic components. When an alternator is connected to a vehicle battery system, the voltage, even in the event of regulator failure, will not often exceed about 20V due to the low resistance and swamping effect of the battery. If an alternator is run with the battery disconnected (which is not recommended), a heavy duty Zener diode connected across the output of the WL/field diodes will offer some protection as, if the system voltage exceeds its breakdown figure, it will conduct and cause the system voltage to be kept within reasonable limits.

Charging circuits

For many applications, the charging circuit is one of the simplest on the vehicle. The main output is connected to the battery via a suitably sized cable (or in some cases two cables to increase reliability and flexibility), and the warning light is connected to an ignition supply on one side and to the alternator terminal at the other. A wire may also be connected to the phase terminal if it is utilized. Figure 6.22 shows two typical wiring circuits. Note that the output of the alternator is often connected to the starter main supply simply for convenience of wiring. If the wires are kept as short as possible this will reduce voltage drop in the circuit. The voltage drop across the main supply wire when the alternator is producing full output current, should be less than 0.5V.

Some systems have an extra wire from the alternator to 'sense' battery voltage directly. An ignition feed may also be found and this is often used to ensure instant excitement of the field windings. A number of vehicles link a wire from the engine management ECU to the alternator. This is used to send a signal to increase engine idle speed if the battery is low on charge.

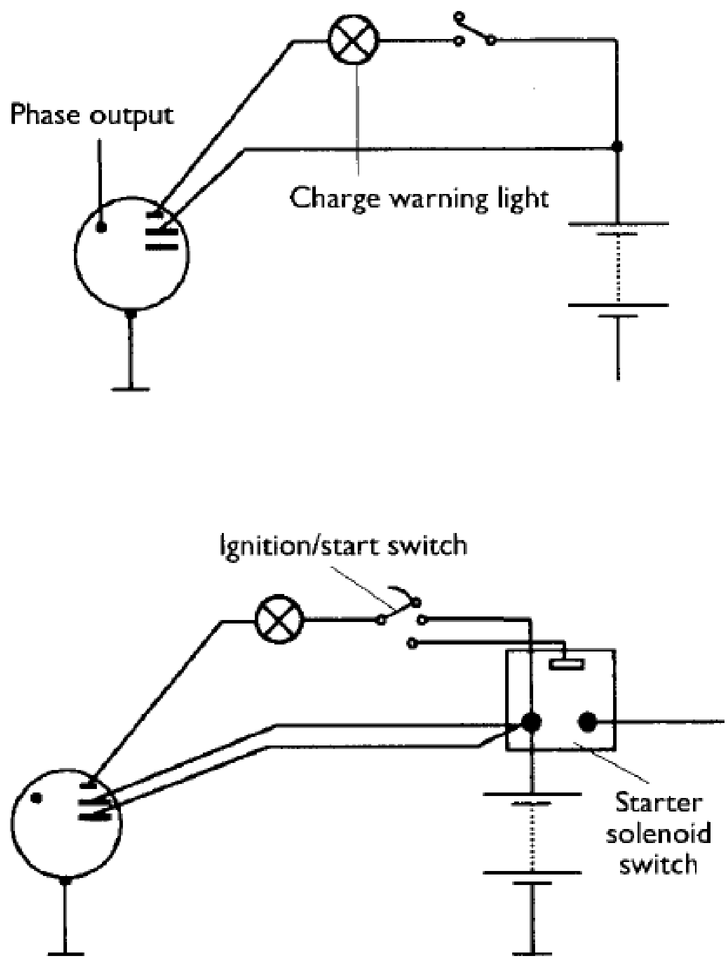


Figure 6.22 Example charging circuits

CUTOUT RELAY

The cutout relay (Figs. 2 & 3) has two windings, a series winding of a few turns of heavy wire (shown in solid red) and a shunt winding of many turns of fine wire (shown in dashed red). The shunt winding is connected across the generator so that generator voltage is impressed upon it at all times.

The series winding is connected in series with the charging circuit so that all generator output passes through it. The relay core and windings are assembled into a frame. A flat steel armature is attached to the frame by a flexible hinge so that it is centered just above the end of the core. The armature contact points are located just above the stationary contact points. When the generator is not operating, the armature contact points are held a wax from the stationary points by the tension of a flat spring riveted on the side of the armature.

CUTOUT RELAY ACTION-When the generator voltage builds up a value great enough to charge the battery, the magnetism induced by the relay windings is sufficient to pull the armature toward the core so that the contact points close. This com- pletes the circuit between the generator and battery. The current which flows from the generator to the battery passes through the series winding in a direction to add to the magnetism holding the armature down and the contact points closed.

When the' generator slows down or stops, current begins to flow from the battery to the generator. This reverse flow of current through the series winding causes a reversal of the series winding magnetic field. The magnetic field of the shunt winding does not reverse. Therefore, instead of helping each other, the two windings now magnetically oppose so that the resultant magnetic field becomes insufficient to hold the armature down. The flat spring pulls the armature away from the core so that the points separate; this opens the circuit between the generator and battery.

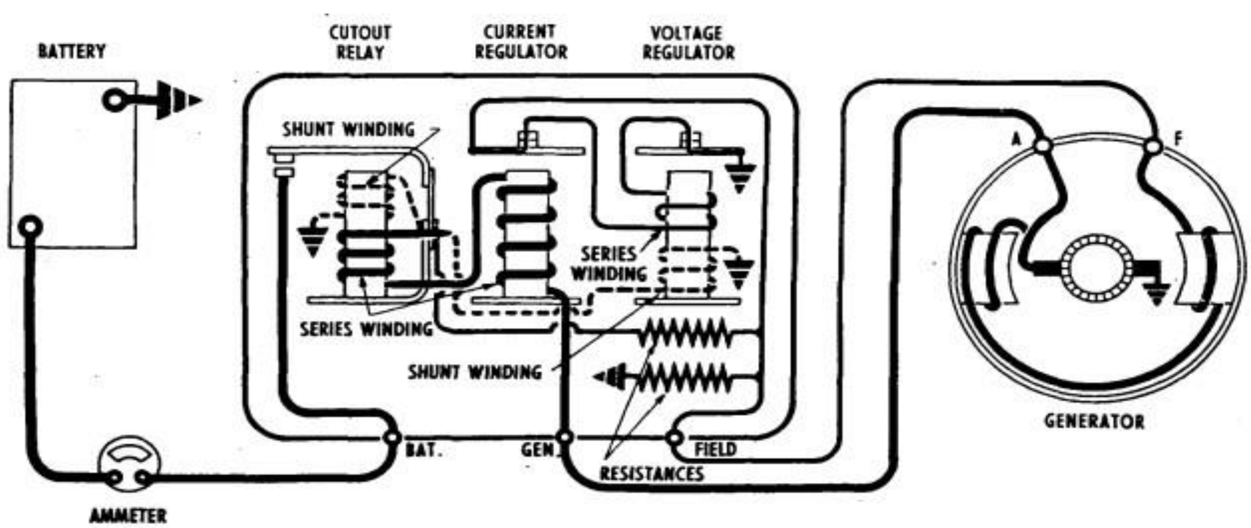


Figure 2—Wiring circuit of Delco-Remy grounded type, three-unit regulator shown in Figure 1. The shunt windings in the cut-out relay and voltage regulator are shown in dashed red. The series windings in the cutout relay and current regulator are shown in solid red. The series winding 1n the voltage regulator is shown in blue.

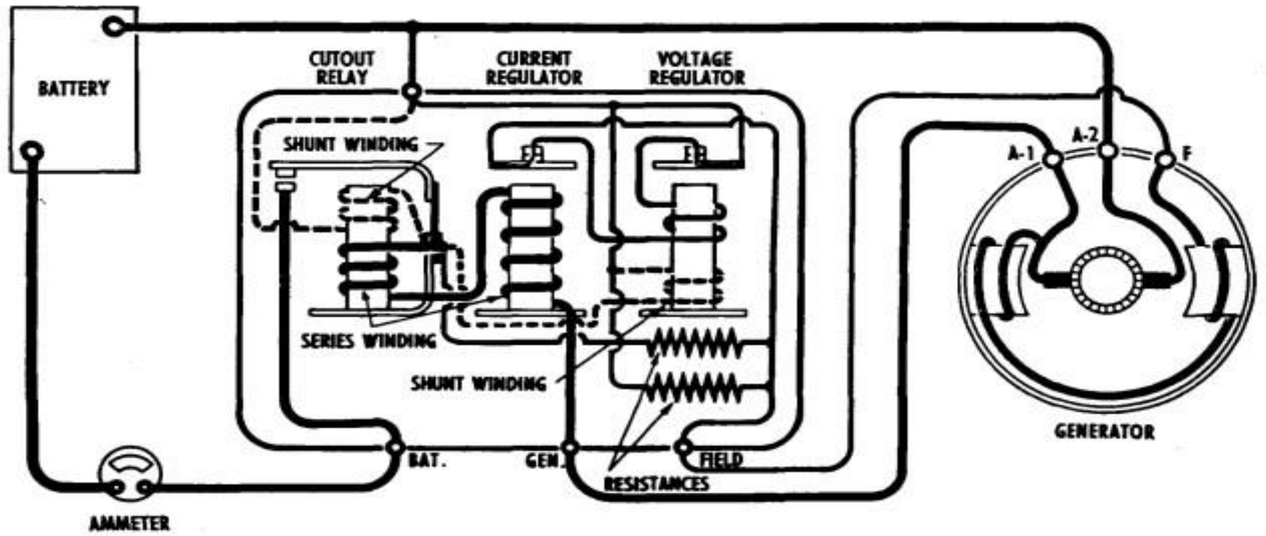


Figure 3—Wiring Circuit of Delco-Remy insulated type, three-unit regulator similar to that shown in Figure 1. The shunt windings in the cutout relay and voltage regulator are shown in dashed red. The series windings in the cutout relay and current regulator are shown in solid red. The series winding in the voltage regulator is shown in blue.

VOLTAGE REGULATOR

The voltage regulator (Figs. 2 & 3) has two windings assembled on a single core, a shunt winding consisting of many turns of fine wire (shown in dashed red) which is shunted across the generator, and a series winding of a few turns of relatively heavy wire (shown in solid blue) which is connected in series with the generator field circuit when the regulator contact points are closed.

The windings and core are assembled into a frame. A flat steel armature is attached to the frame by a flexible hinge so that it is just above the end of the core. The armature contains a contact point which is just beneath a stationary contact point. When the voltage regulator is not operating, the tension of a spiral spring holds the armature away from the core so that the points are in contact and the generator field circuit is completed to ground through them.

VOLTAGE REGULATOR ACTION-When the generator voltage reaches the value for which the voltage regulator is adjusted, the magnetic field produced by the two windings (shunt and series) overcomes the armature spring tension and pulls the armature down so that the contact points separate. This inserts resistance into the generator field circuit so that the generator field current and voltage are reduced. Reduction of the generator voltage reduces the magnetic field of the regulator shunt winding. Also, opening the regulator points opens the regulator series winding circuit so that its magnetic field collapses completely. The consequence is that the magnetic field is reduced sufficiently to allow the spiral spring to pull the armature away from the core so that the contact points again close. This directly grounds the generator field circuit so that generator voltage and output increase. The above cycle of action again takes place and the cycle continues at a rate of 50 to 200 times a second, regulating the voltage to a predetermined value. With the voltage thus limited the generator supplies varying amounts of current to meet the varying states of battery charge and electrical load.

Solenoid

A solenoid is simply a specially designed electromagnet. A solenoid usually consists of a cylindrical coil wound with one or more layers of insulated wire and a movable iron core called the *armature*. When current flows through a wire, a magnetic field is set up around the wire. If we make a coil of many turns of wire, this magnetic field becomes many times stronger, flowing around the coil and through its center in a doughnut shape. The length of the solenoid is much larger in comparison with its diameter. When the coil of the solenoid is energized with current, the core moves to increase the flux linkage by closing the air gap between the cores. The movable core is usually spring-loaded to allow the core to retract when the current is switched off. The force generated is approximately proportional to the square of the current and inversely proportional to the square of the length of the air gap.

Solenoids are inexpensive, and their use is primarily limited to on-off applications such as latching, locking, and triggering. They are frequently used in home appliances (e.g. washing machine valves), office equipment (e.g. copy machines), automobiles (e.g. door latches and the starter solenoid), pinball machines (e.g., plungers and bumpers), and factory automation.

Flux Distribution of a Solenoid

The direction of the field inside the solenoid may be found by applying the right hand rule for solenoids. This rule states that if a solenoid is grasped with the right hand such that the fingers point in the direction of the current in the wire, then the thumb will point in the direction of magnetic flux. The application of the right hand rule for solenoids is illustrated in figure 1.1. The flux distribution due to solenoid carrying current is given in figure 1.2. The magnetic field produced by a solenoid is similar to that of a bar magnet. One end of the solenoid becomes a north pole (N) where the flux leaves the solenoid. The other end becomes the South Pole (S) where the flux enters it.

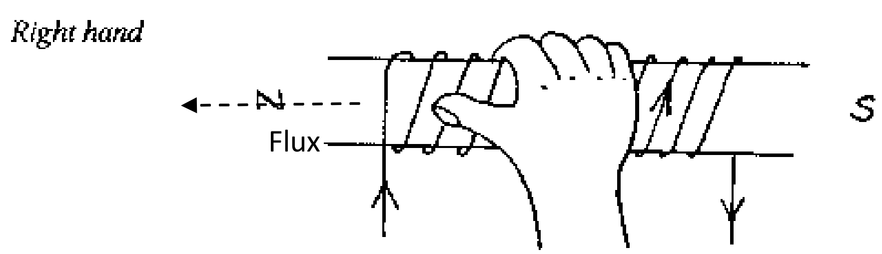


Figure 1.1 Right hand rule for solenoid

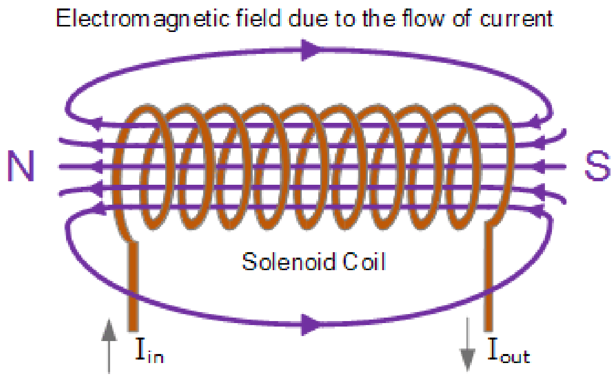


Figure 1.2 Magnetic field due to a solenoid carrying current

Application of Solenoid

An electromechanical *relay* is a solenoid used to make or break mechanical contact between electrical leads. A small voltage input to the solenoid controls a potentially large current through the relay contacts. Applications include power switches and electromechanical control elements. A relay performs a function similar to a power transistor but has the capability to switch extremely large currents if necessary. However, transistors have a much shorter switching time than relays.

As illustrated in figure 2, a *voice coil* consists of a coil that moves in a magnetic field produced by a permanent magnet and intensified by an iron core. The force on the coil is directly proportional to the current in the coil. The coil is usually attached to a movable load such as the diaphragm of an audio speaker, the spool of a hydraulic proportional valve, or the read-write head of a computer disk drive. The linear response and bidirectional capability make voice coils more attractive than solenoids for control applications.

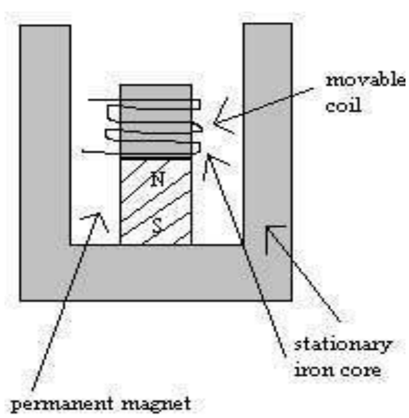


Fig 2. Voice coil

Charging System Components

The Charging system is an important part of the electrical system. The charging system has two essential functions:

- Generate electrical power to run the vehicle's electrical systems
- Generate current to recharge the vehicle's battery

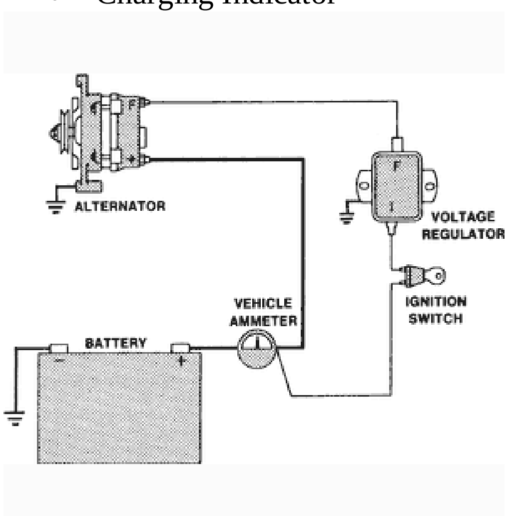
Electrical Power: At low engine speeds, the battery may supply some of the power the vehicle needs. At high engine speeds, the charging system handles all of the vehicle's electrical requirements.

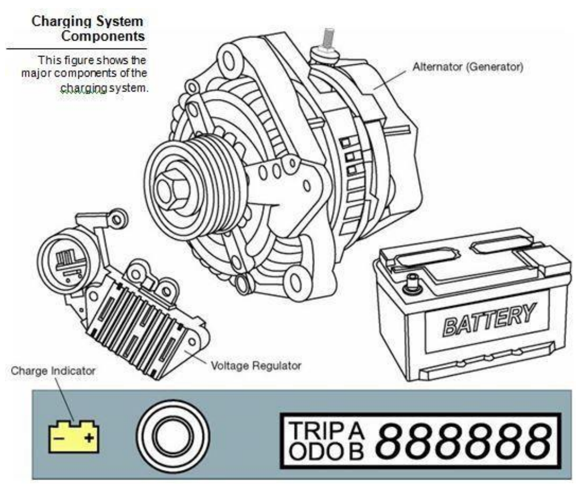
Charging: Alternator (generator) output is higher than battery voltage to recharge the battery.

The charging system components:

These components make up the charging system:

- Alternator
- Voltage regulator
- Battery
- Charging Indicator





I. Alternator

The alternator generates electrical power to run accessories and to recharge the batteries. It is normally driven by a belt located off the crankshaft. Mechanical energy from the crankshaft is converted by the alternator into electrical energy for the batteries and accessories.
The alternator contains three main components:

- Stator (attached to alternator housing, remains stationary),

- Rotor (spins inside the stator),
- Rectifier,

Slip ring and brushes make an electrical connection to the spinning rotor.

The alternator generates electricity through these steps:

- Engine power drives the alternator rotor through a pulley and drive belt.
- The alternator rotor spins inside the windings of the stator.
- The stator windings generate an alternating current.
- Rectifier diodes change the alternating current (AC) into direct current (DC).

II. Voltage regulator

The voltage regulator acts as an electrical traffic cop to control alternator output. It senses when the batteries need recharging, or when the vehicles electrical needs increase, and adjusts the alternators output accordingly. ie., it controls the alternator's output current to prevent over-charging and under charging of the battery. It does this by regulating the current flowing from the battery to the rotor's field coil.

The voltage regulator can be mounted inside or outside of the alternator housing. If the regulator is mounted outside there will be a wiring harness connecting it to the alternator. Today's IC voltage regulator is a fully electronic device, using resistors and diodes.

III. Battery

The batteries are a reservoir of chemical electrical power. Their primary purpose is to crank the engine. They also supply power to vehicle accessories when the electrical load is too great for the alternator alone. The battery also acts as a voltage stabilizer. The battery must always remain attached to the electrical system while the engine is running.

IV. Charging Indicator

The charging indicator is usually an ON /OFF warning lamp. When the system is running, the light should be off. The lamp lights when the charging system is not providing sufficient charge.

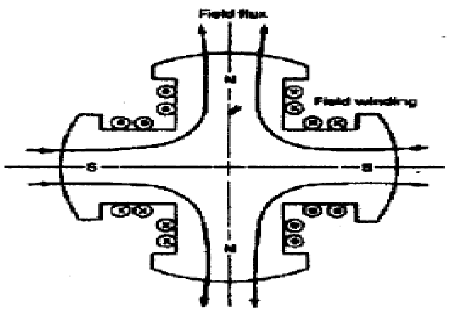
Synchronous generator (Alternators)

AC generators are usually called alternators. They are also called synchronous generators. A synchronous generator is a machine for converting mechanical power from a prime mover to ac electric power at a specific voltage and frequency.

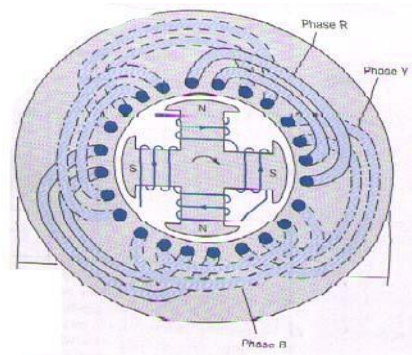
Principle of Operation: The operation of a synchronous generator is based on Faraday's law of electromagnetic induction, and in an ac synchronous generator the generation of emf's is by relative motion of conductors and magnetic flux. The rotating magnetic field induces an AC voltage in the stator windings. Since the currents in the stator windings vary in step with the position of the rotor, an alternator is a synchronous generator.

In constructing a synchronous machine a point to note is that the stator is fixed and the poles rotate.

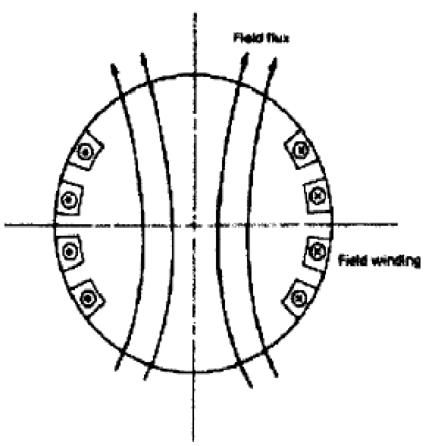
There are two categories of Synchronous machines:
(a) those with salient or projecting poles
(b) those with cylindrical rotors



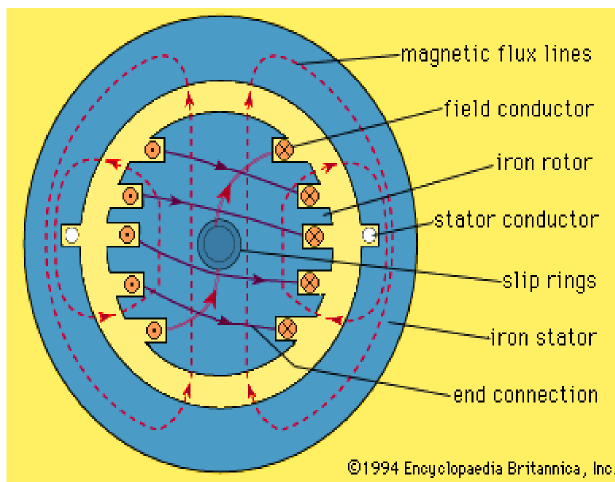
4-Pole Salient Rotor



A Salient Pole Rotor



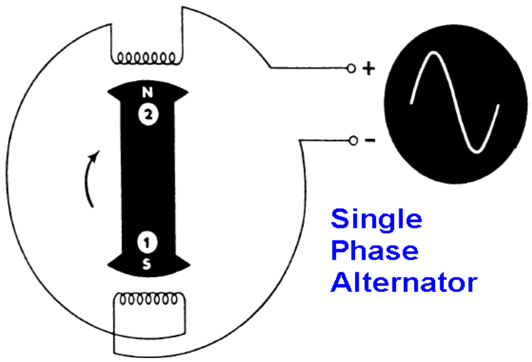
2-pole Cylindrical Rotor



A Cylindrical Rotor

<u>Single Phase Alternator</u>

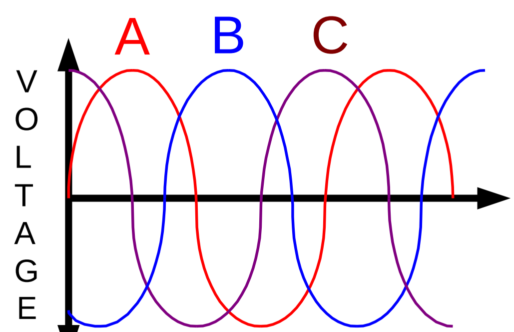
- Single-phase alternator is an alternating current electrical generator that produces a single, continuously alternating voltage. Single-phase generators can be used to generate power in single-phase electric power systems.
- A single-phase alternator has all the armature conductors connected in series
- The stator is two poles. The winding is wound in two distinct pole groups, both poles being wound in the same direction around the stator frame.
- The rotor also consists of two pole groups, adjacent poles being of opposite polarity.

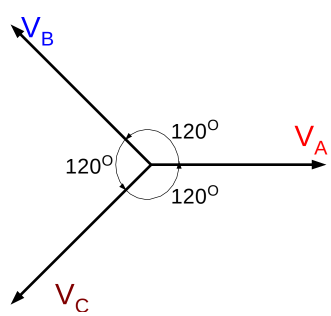


- The two poles of the stator winding are connected to each other so that the AC voltages are in phase, so they add.
- As the rotor (field) turns, its poles will induce AC voltages in the stator (armature) windings. Since one rotor pole is in the same position relative to a stator pole as any other rotor pole, both the stator poles are cut by equal amounts of magnetic lines of force at any time. As a result, the voltages induced in the two poles of the stator winding have the same amplitude or value at any given instant.

Three Phase Alternator

- The three-phase alternator has three single-phase windings spaced so that the voltage induced in any one is phase-displaced by 120 degrees from the other two.
- The voltage waveforms generated across each phase are drawn on a graph phase-displaced 120 degrees from each other.

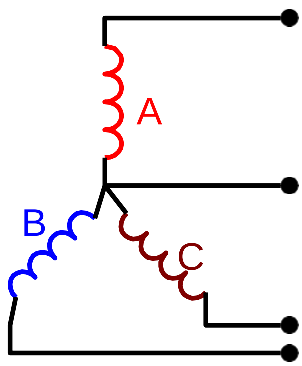


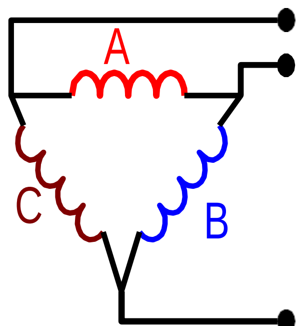


- The three phases are independent of each other.
- One point from each winding can be connected to form a neutral and thus make a wye (star) connection.
- The voltage from this point to any one of the line leads will be the phase voltage. The line voltage across any two line leads is the vector sum of the individual phase voltages. The line voltage is 1.73, ($\sqrt{3}$), times the phase voltage.
- Since the windings form only one path for current flow between phases, the line and phase currents are equal.

- A three-phase stator can also be connected so that the phases form a “delta” connection.
- In the delta connection the line voltages are equal to the phase voltages, but the line currents will be equal to the vector sum of the phase currents.
- Since the phases are 120 degrees out of phase, the line current will be 1.73, ($\sqrt{3}$), times the phase current. Both "star" and the "delta" connections are used in alternators.

Three Phase Stator Connection



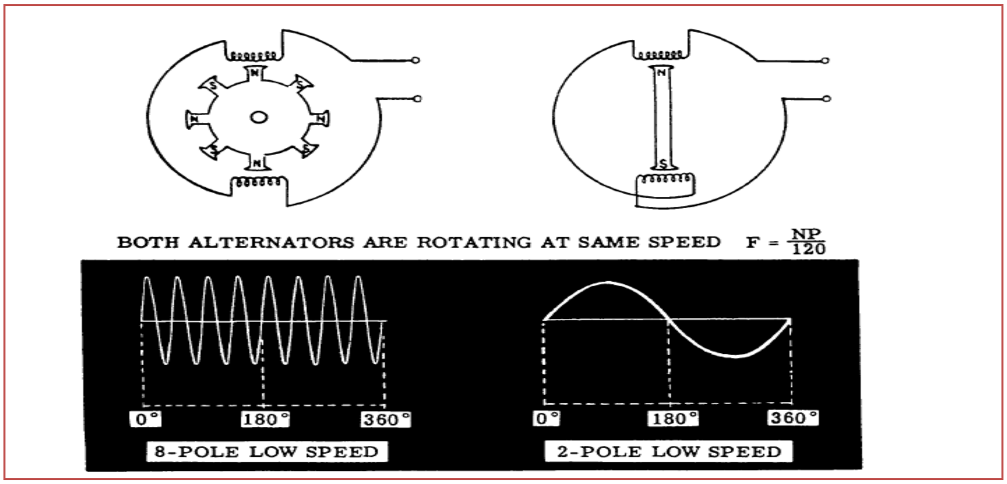


- The frequency of the AC generated by an alternator depends upon the number of poles and the speed of the rotor
- When a rotor has rotated through an angle so that two adjacent rotor poles (a north and a south) have passed one winding, the voltage induced in that one winding will have varied through a complete cycle of 360 electrical degrees.
- A two pole machine must rotate at twice the speed of a four-pole machine to generate the same frequency.
- The magnitude of the voltage generated by an alternator can be varied by adjusting the current on the rotor which changes the strength of the magnetic field.
- A two pole alternator produces one electrical cycle for each complete mechanical rotation.
- A four pole alternator will produce two electrical cycles for each mechanical rotation because two north and two south poles move by each winding on the stator for one complete revolution of the rotor.
- $f = (N)(P/2)/60 = (NP)/120$

  where N is the speed of the rotor in revolutions per minute, P is the number of poles

f is the electrical line frequency produced by the alternator.

- The speed of the rotor must be divided by 60 to change from revolutions per minute to revolutions per second.



- In an alternator the output voltage varies with the load.
- There are two voltage drops.{ IR & IX_L }
- The IX_L drop is due to the inductive reactance of the armature windings.
- Both the IR drop and the IXL drop decrease the output voltage as the load increases.
- The change in voltage from no-load to full-load is called the "voltage regulation" of an alternator.
- A constant voltage output from an alternator is maintained by varying the field strength as required by changes in load.

EMF Equation of an alternator

Let,
P= No. of poles
Z= No. of Conductors or Coil sides in series/phase i.e. Z= 2T
where T is the number of coils or turns per phase
f = frequency of induced e.m.f in Hz
ϕ = Flux per pole (Weber)
N = rotor speed (RPM)

If induced e.m.f is assumed sinusoidal then, K_f= Form factor = 1.11

In one revolution of the rotor i.e. in 60/N seconds, each conductor is cut by a flux of Pϕ Webers.

dϕ= ϕP and also dt= seconds60/N

then induced e.m.f per conductor (average) = dϕ/ dt= Pϕ/(60/N) =P N ϕ/60 …..(a)

But We know that f = PN/120 or N= 120f/P

Putting the value of N in Equation (a)...

We get the average value of e.m.f per conductor is

Eav = Pϕ/60 x 120 f/P **= 2f ϕ Volts.** {N= 120f/P}

If there are Z conductors in series per phase,

then average e.m.f per phase = 2fϕZ Volts***= 4fϕT Volts....*** {Z=2T}

Also we know that Form factor= RMS Value/Average Value...

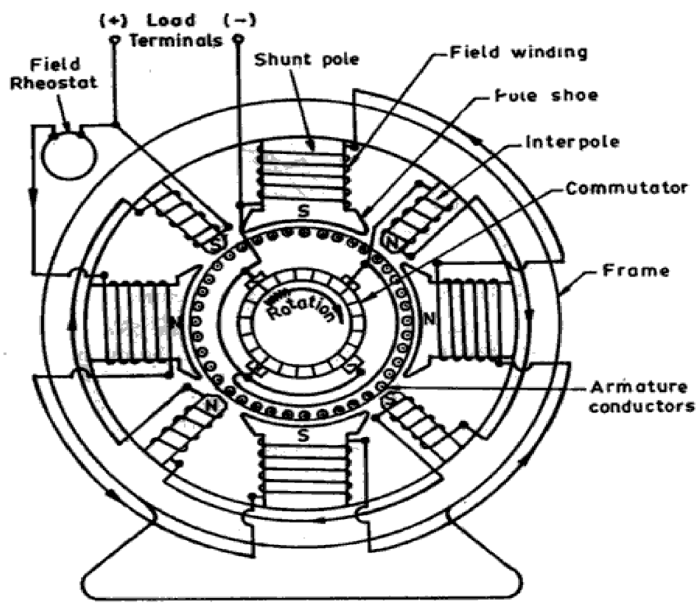
= RMS value= Form factor x Average Value,

= 1.11 x 4fϕT = 4.44fϕT Volts.

GENERATORS

Construction of DC Generator

In construction dc machine consists of four parts mainly, 1. Field magnets 2. Armature 3. Commutator 4. Brush and brush gear. DC machine (can be a generator or motor) with four poles is shown in figure 1.1



1. Field System:

The object of the field system is to create a uniform magnetic field within with the armature rotates. Electromagnets are preferred on the account of their magnetic effects and field strength regulation which can be achieved by controlling the magnetizing current. Field magnets consist of the following parts:

(i) Yoke or Frame (ii) Pole cores (iii) Pole shoes (iv) Magnetizing coils

Cylindrical yoke is usually used which acts as a frame of the machine and carries the magnetic flux produced by the poles. Since the field is stationary there is no need to use laminated yoke for normal machine. In small machines, cast iron yokes are used because of its cheapness but yoke of large machines are made of fabricated steel due to its high permeability.

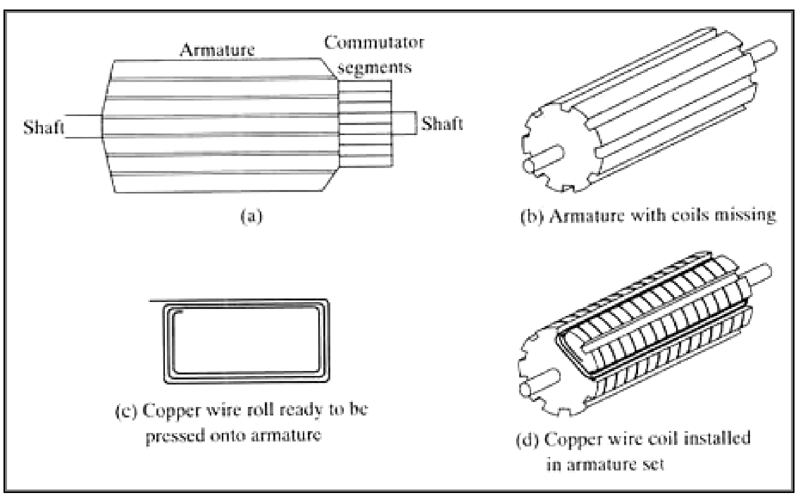
Pole core is usually of circular section and is used to carry the coils of insulated wires carrying the exciting current. Pole cores are usually not laminated and made of cast steel.

Each pole core has a pole shoe serves having a curved surface. The pole shoe serves two purposes:

(i) It supports the field coils
(ii) It increases the cross-sectional area of the magnetic circuit and reduces its reluctance.

Each pole core has one or more field coils or magnetizing coils placed over it to produce a magnetic field. The field coils are connected in series with one another such that when the current flows through the coils, alternate north and south poles are produced in the direction of rotation.

2. Armature



It is a rotating part of a DC machine and is built up in a cylindrical or drum shape. The purpose of armature is to rotate the conductors in the uniform magnetic field. It consists of coils of insulated wires wound around an iron and so arranged that electric currents are induced in these wires when the armature is rotated in a magnetic field. Its most important function is to provide a path of low reluctance to the magnetic flux. The armature core is made of high permeability silicon-steel stampings.

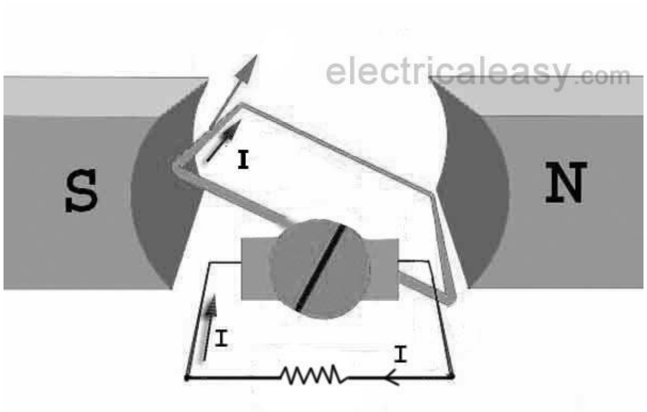
3. Commutator:

It is a form of rotating switch. They are placed between armature and external circuit. The commutator will reverse the connections to the external circuit at the instant aech reversal of circuit in the armature coil.

4. Brushes & Bearings

Brushes are made of carbon or graphite. It collects current from the commutator and convey it to external load resistance. It is rectangular in shape. Brushes are housed in brush holders and mounted over brush holder studs. Ball bearings are used as they are reliable for light machines.For heavy machines roller bearings are used.

Working Principle:



(a)

(b) According to Faraday's law of electromagnetic induction, when a conductor moves in a magnetic field (thereby cutting the magnetic flux lines), a dynamically induced emf is produced in the conductor. The magnitude of generated emf can be given by emf equation of DC generator. If a closed path is provided to the moving conductor then generated emf causes a current to flow in the circuit.

(c) Thus in DC generators, as we have studied earlier, when armature is rotated with the help of a prime mover and field windings are excited (there may be permanent field magnets also), emf is induced in armature conductors. This induced emf is taken out via commutator-brush arrangement.

E.M.F Equation of a DC Generator

Let

Φ = flux/pole in weber
Z = total number of armture conductors
= No.of slots x No.of conductors/slot
P = No.of generator poles
A = No.of parallel paths in armature
N = armature rotation in revolutions per minute (r.p.m)
E = e.m.f induced in any parallel path in armature

Generated e.m.f Eg = e.m.f generated in any one of the parallel paths i.e E.
Average e.m.f geneated /conductor = **$d\Phi/dt$** volt (n=1)
Now, flux cut/conductor in one revolution **$d\Phi = \Phi P$** Wb
No.of revolutions/second = N/60
Time for one revolution, **dt = 60/N** second
Hence, according to Faraday's Laws of Electroagnetic Induction,
E.M.F generated/conductor is

$$\frac{d\varnothing}{dt} = \frac{\varnothing PN}{60}$$

For a simplex wave-wound generator
No.of parallel paths = 2
No.of conductors (in series) in one path = Z/2
E.M.F. generated/path is

$$\frac{\varnothing PN}{60} \text{ x } \frac{Z}{2} = \frac{\varnothing ZPN}{120} \text{ volt}$$

For a simplex lap-wound generator
No.of parallel paths = P
No.of conductors (in series) in one path = Z/P
E.M.F.generated/path

$$\frac{\varnothing PN}{60} \text{ x } \frac{Z}{P} = \frac{\varnothing ZN}{60} \text{ volt}$$

In general generated e.m.f

$$E_g = \frac{\varnothing ZN}{60} \text{ x } \left(\frac{P}{A}\right) \text{ volt}$$

where A = 2 for simplex wave-winding
and A= P for simplex lap-winding

Types of Generators

Generators are usually classified according to the way in which their fields are excited.The field windings provide the excitation necessary to set up the magnetic fields in the machine. There are various types of field windings that can be used in the generator or motor circuit. In addition to the following field winding types, permanent magnet fields are used on some smaller DC products.

Generators may be divided in to

(a) Separately-excited generators

(b) Self-excited generators.

Separately-excited generators are those whoe field magnets are energised from an independent external source of DC current.

Self-excited generators are those whose field magnets are energused by the current produced by the generators themselves.Due to residual magnetism, there is always present someflux in the poles.When the armature is rotated, some e.m.f and hence some induced current is produced which is partly or fully passed through the field coils thereby strengthening the residual pole flux.

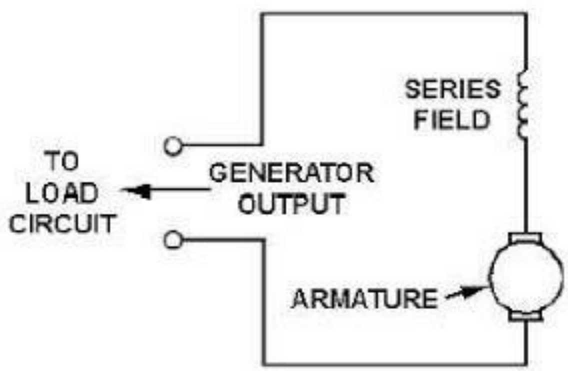
Self-excited generators are classed according to the type of field connection they use. There are three general types of field connections —

SERIES-WOUND, SHUNT-WOUND (parallel), and COMPOUND-WOUND.

Compound-wound generators are further classified as cumulative-compound and differential-compound.

Series-wound generator

In the series-wound generator, shown in figure, the field windings are connected in series with the armature. Current that flows in the armature flows through the external circuit and through the field windings. The external circuit connected to the generator is called the load circuit



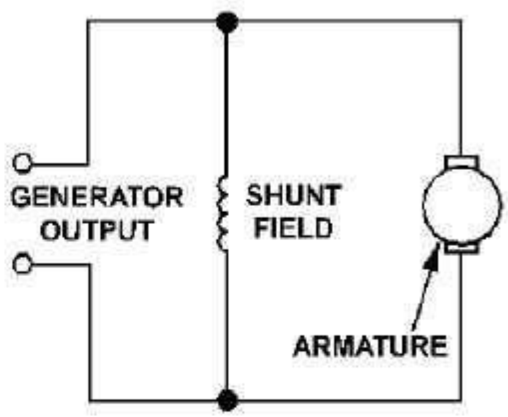
A series-wound generator uses very low resistance field coils, which consist of a few turns of large diameter wire.

The voltage output increases as the load circuit starts drawing more current. Under low-load current conditions, the current that flows in the load and through the generator is small. Since small current means that a small magnetic field is set up by the field poles, only a small voltage is induced in the armature. If the resistance of the load decreases, the load current increases.

Under this condition, more current flows through the field. This increases the magnetic field and increases the output voltage. A series-wound dc generator has the characteristic that the output voltage varies with load current. This is undesirable in most applications. For this reason, this type of generator is rarely used in everyday practice.

Shunt wound

In this field winding is connected in parallel with the armature conductors and have the full voltage of the generator applied across them.The field coils consist of many turns of small wire. They are connected in parallel with the load. In other words, they are connected across the output voltage of the armature.



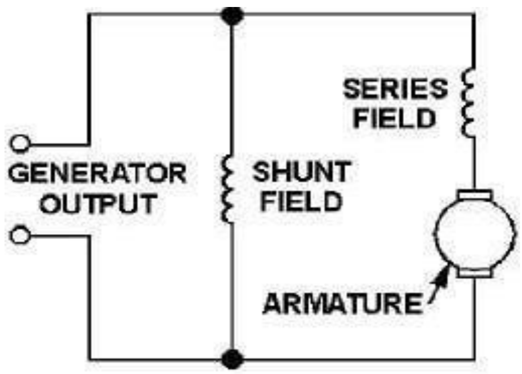
Current in the field windings of a shunt-wound generator is independent of the load current (currents in parallel branches are independent of each other). Since field current, and therefore field strength, is not affected by load current, the output voltage remains more nearly constant than does the output voltage of the series-wound generator.

In actual use, the output voltage in a dc shunt-wound generator varies inversely as load current varies. The output voltage decreases as load current increases because the voltage drop across the armature resistance increases (E = IR).

In a series-wound generator, output voltage varies directly with load current. In the shunt-wound generator, output voltage varies inversely with load current. A combination of the two types can overcome the disadvantages of both. This combination of windings is called the compound-wound dc generator.

Compound-wound generator :

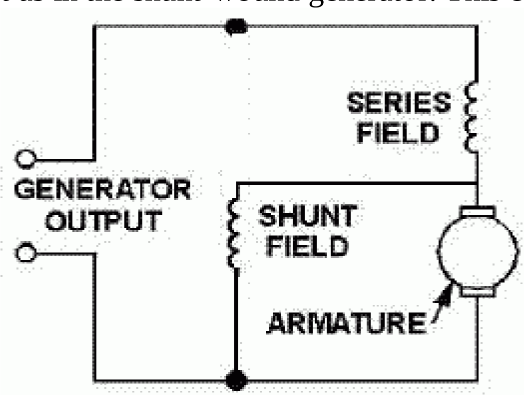
Compound-wound generators have a series-field winding in addition to a shunt-field winding, as shown in figure.



Compound wound long shunt

The shunt and series windings are wound on the same pole pieces. They can be either short-shunt or long-shunt as shown in figures. In a comound generator, the shunt field is stronger than the series field.When series field aids the shunt field, generator is said to be commutatively-compounded.On the other hand if series field opposes the shunt field,the generator is said to be differentially compounded.

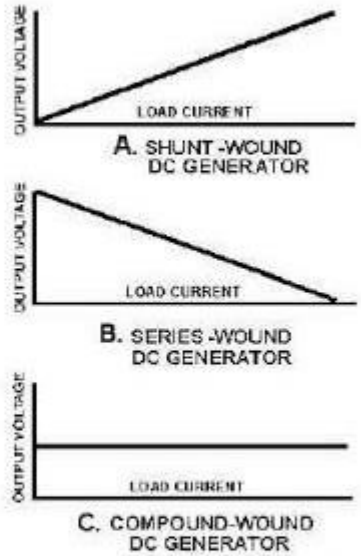
In the compound-wound generator when load current increases, the armature voltage decreases just as in the shunt-wound generator. This causes the voltage applied to the shunt-field winding



Compound wound short shunt

to decrease, which results in a decrease in the magnetic field. This same increase in load current, since it flows through the series winding, causes an increase in the magnetic field produced by that winding.

By proportioning the two fields so that the decrease in the shunt field is just compensated by the increase in the series field, the output voltage remains constant. This is shown in figure, which shows the voltage characteristics of the series-, shunt-, and compound-wound generators. As you can see, by proportioning the effects of the two fields (series and shunt), a compound-wound generator provides a constant output voltage under varying load conditions. Actual curves are



Voltage output characteristics

seldom, if ever, as perfect as shown.

Characteristics of DC generators

The speed of a d.c. machine operated as a generator is fixed by the prime mover. For general-purpose operation, the prime mover is equipped with a speed governor so that the speed of the generator is practically constant. Under such condition, the generator performance deals primarily with the relation between excitation, terminal voltage and load. These relations can be best exhibited graphically by means of curves known as generator characteristics. These characteristics show at a glance the behaviour of the generator under different load conditions.

D.C. Generator Characteristics The following are the three most important characteristics of a d.c. generator:

1. Open Circuit Characteristic (O.C.C.):

This curve shows the relation between the generated e.m.f. at no-load (E0) and the field current (If) at constant speed. It is also known as magnetic characteristic or no-load saturation curve. Its shape is practically the same for all generators whether separately or self-excited. The data for O.C.C. curve are obtained experimentally by operating the generator at no load and constant speed and recording the change in terminal voltage as the field current is varied.

2. Internal or Total characteristic (E/Ia)

This curve shows the relation between the generated e.m.f. on load (E) and the armature current (Ia). The e.m.f. E is less than E0 due to the demagnetizing effect of armature reaction. Therefore, this curve will lie below the open circuit characteristic (O.C.C.). The internal characteristic is of interest chiefly to the designer. It cannot be obtained directly by experiment. It is because a voltmeter cannot read the e.m.f. generated on load due to the voltage drop in armature resistance.

The internal characteristic can be obtained from external characteristic if winding resistances are known because armature reaction effect is included in both characteristics

3. External characteristic (V/IL):

This curve shows the relation between the terminal voltage (V) and load current (IL). The terminal voltage V will be less than E due to voltage drop in the armature circuit. Therefore, this curve will lie below the internal characteristic. This characteristic is very important in determining the suitability of a generator for a given purpose. It can be obtained by making simultaneous measurements of terminal voltage and load current (with voltmeter and ammeter) of a loaded generator.

Unit-II

Wiring, Lighting and Ignition systems

AUTOMOTIVE WIRING

Electrical power and control signals must be delivered to electrical devices reliably and safely so that the electrical system functions are not impaired or converted to hazards. To fulfill power distribution military vehicles, use one-and two-wire circuits, wiring harnesses, and terminal connections.

Among your many duties will be the job of maintaining and repairing automotive electrical systems. All vehicles are not wired in exactly the same manner; however, once you understand the circuit of one vehicle, you should be able to trace an electrical circuit of any vehicle using wiring diagrams and color codes.

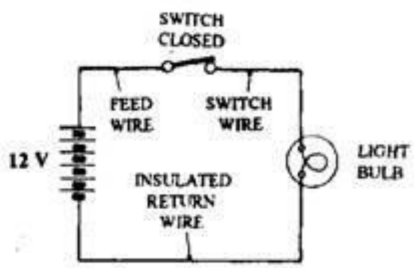
ONE-AND TWO-WIRE CIRCUITS

Tracing wiring circuits, particularly those connecting lights or warning and signal devices, is no simple task. The branch circuits making up the individual systems have one wire to conduct electricity from the battery to the unit requiring it and ground connections at the battery and the unit to complete the circuit. These are called ONE-WIRE CIRCUITS or branches of a GROUND RETURN SYSTEM. In automotive electrical systems with branch circuits that lead to all parts of the equipment, the ground return system saves installation time and eliminates the need for an additional wiring to complete the circuit. The all-metal construction of the automotive equipment makes it possible to use this system.

The TWO-WIRE CIRCUIT requires two wires to complete the electrical circuit- one wire from the source of electrical energy to the unit it will operate, and another wire to complete the circuit from the unit back to the source of the electrical power. Two-wire circuits provide positive connection for light and electrical brakes on some trailers. The coupling between the trailer and the equipment, although made of metal and a conductor of electricity, has to be jointed to move freely. The rather loose joint or coupling does not provide the positive and continuous connection required to use a ground return system between two vehicles. The two-wire circuit is commonly used on equipment subject to frequent or heavy vibrations. Tracked equipment, off-road vehicles (tactical), and many types of construction equipment are wired in this manner.

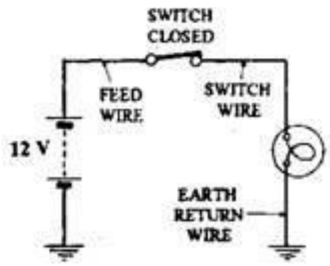
Insulated Return

Some vehicle application requires a separate insulated-cable system for both the feed and the return conductors. It is also safer because with separate feed and return cables, it is practically impossible for the cable conductors to short even if chafed and touching any of the metal bodywork, as the body is not live since it is not a part of the electrical circuits. From the safety reasons, an insulated return is essential for vehicles transporting highly flammable liquids and gases, where a spark could very easily set off an explosion or a fire. The vehicles, such as coaches and double-decker buses use large quantity of plastic panelling. For these vehicles an insulated return is more reliable and safer. The insulated return off course uses extra cable that makes the overall wiring harness heavier, less flexible, and bulky, consequently increases the cost to some extent.



Earth Return.

All electrical circuits incorporate both a feed and a return conductor between the battery and the component requiring supply of electrical energy. The vehicle with a metal structure can be used as one of the two conducting paths. This is called as the earth return (Fig. 13.51). A live feed wire cable forms the other conductor. To complete the earth-return path, one end of a short thick cable is bolted to the chassis structure while the other end is attached to one of the battery terminal posts. The electrical component is also required to be earthed in a similar way. Only one battery-to-chassis conductor is necessary for a complete vehicle's wiring system and similarly any number of separate earth-return circuits can be wired. An earth-return system, therefore, reduces and simplifies the amount of wiring so that it is easy to trace electrical faults.



Positive and Negative earthing

In the beginning, it was the general practice of earthing the negative terminal of the battery, whereas the positive current was supplied to the electrical units. The negative earthing system is still used in the cars of American make.

In some countries, the negative earth system has been replaced by the positive earth system. This is because the positive earth system possesses certain advantages over the negative earth system. These advantages concern the temperature of the central spark plug electrode and the corrosion of some parts, it is well known fact that the positive terminal of the leadacid battery is attacked by the liberated gases. If this is the live terminal and the negative terminal earthed, the exposed part of the positive will become corroded.

Further it is also a well known fact that the positive point of the spark plug wears away more quickly than the negative point. In view of this fact, the central electrode of the plug will wear away quickly if made electrode of the plug will wear away quickly if made positive when compared with the metal electrode of the shell. Alternatively, the central electrode of the plug will have much longer life if made negative by earthing the positive terminal of the battery.

Another factor which plays an important role in the voltage requirements of a spark plug is the temperature of the negative electrode. The hotter this electrode is, the lower will be the voltage required for producing the spark, It has also been observed that more uniform voltages at the sparking points have been obtained with the central electrode being negative. Further, the metal rotor arm of the distributor, if made negative, will wear at a slower rate than if it were made positive.

There is an additional advantage of the positive earth method in the ignition coil elements the primary circuit voltage is added to the secondary circuit voltage, making it more economical.

Recently, with the adoption of alternators in place of generators, it has been observed that employing negative earth method is advantageous along with an ac current rectifier having transistors and diodes. This has meant shifting back to the negative earth method. However it is worth mentioning that the important advantages of the positive earth for the ignition system still hold good.

LIGHTING

The lighting circuit includes the battery, vehicle frame, all the lights, and various switches that control their use. The lighting circuit is known as a single-wire system since it uses the vehicle frame for the return.

The complete lighting circuit of a vehicle can be broken down into individual circuits, each having one or more lights and switches. In each separate circuit, the lights are connected in parallel, and the controlling switch is in series between the group of lights and the battery.

The marker lights, for example, are connected in parallel and are controlled by a single switch. In some installations, one switch controls the connections to the battery, while a selector switch determines which of two circuits is energized. The headlights, with their high and low beams, are an example of this type of circuit.

In some instances, such as the courtesy lights, several switches may be connected in parallel so that any switch may be used to turn on the light. When a wiring diagram is being studied, all light circuits can be traced from the battery through the ammeter to the switch (or switches) to the individual light.

LAMPS

Small gas-filled incandescent lamps with tungsten filaments are used on automotive and construction equipment. The filaments supply the light when sufficient current is flowing through them. They are designed to operate on a low voltage current of 12 or 24 volts, depending upon the voltage of the the vehicle will be of the single-or double-contact small one-half-candlepower bulbs to large 50- candlepower bulbs. The greater the candlepower of the lamp, the more current it requires when lighted. Lamps are identified by a number on the base. When you replace a lamp in a vehicle, be sure the new lamp is of the proper rating. The lamps within Lamps are rated as to size by the candlepower (luminous intensity) they produce. They range from types with nibs to fit bayonet sockets, as shown in lamp is also whiter than a conventional lamp, which increases lighting ability.

HEADLIGHTS

The headlights are sealed beam lamps that illuminate the road during nighttime operation. Headlights consist of a lens, one or two elements, and a integral reflector. When current flows through the element, the element gets white hot and glows. The reflector and lens direct the light forward.

Many modern passenger vehicles use halogen headlights. A halogen headlight contains a small, inner halogen lamp surrounded by a conventional sealed housing. A halogen headlamp increases light output by 25 percent with no increase in current. The halogen The headlight switch is an ON/ OFF switch and rheostat (variable resistor) in the dash panel) or on the steering column. The headlight switch controls current flow to the lamps of the headlight system. The rheostat is for adjusting the brightness of the instrument panel lights.

Military vehicles that are used in tactical situations are equipped with a headlight switch that is integrated with the blackout lighting switch. An important feature of this switch is that it reduces the possibility of accidentally turning on the lights in a blackout.

With no lights on, the main switch can be turned to the left without operating the mechanical switch to get blackout marker lights (including blackout taillights and stoplights) and blackout driving lights. But for stoplights for daylight driving or headlights for ordinary night driving, you must first lift the mechanical switch lever and then turn the main switch to the right. The auxiliary switch gives panel lights when the main switch is in any of its ON positions. But it will give parking lights only when the main switch is in service drive (to the extreme right). When the main switch is off, the auxiliary switch should not be moved from the OFF position.

DIMMER SWITCH

The dimmer switch controls the high and low headlamp beam function and is normally mounted on the floorboard or steering column. When the operator activates the dimmer switch, it changes the electrical connection to the headlights.

In one position, the high beams are turned on, and, in the other position, the dimmer changes them to low beam.

Aiming Headlights

The headlights can be aimed using a mechanical aimer or a wall screen. Either method assures that the headlight beams point in the direction specified by the vehicle manufacturer. Headlights that are aimed too high can blind oncoming vehicles. Headlights that are aimed too low or to one side will reduce the operator's visibility.

TURN-SIGNAL SYSTEMS

Vehicles that operate on any public road must be equipped with turn signals. These signals indicate a left or right turn by providing a flashing light signal at the rear and front of the vehicle.

The turn-signal switch is located on the steering column. It is designed to shut off automatically after the turn is completed by the action of the canceling cam. A common design for a turn-signal system is to use the same rear light for both the stop and turn signals. This somewhat complicates the design of the switch in that the stoplight circuit must pass through the turn-signal switch. When the turn-signal switch is turned off, it must pass stoplight current to the rear lights. As a left or right turn signal is selected, the stoplight circuit is open and the turn-signal circuit is closed to the respective rear light.

The turn signal flasher unit creates the flashing of the turn signal lights. It consists basically of a bimetallic (two dissimilar metals bonded together) strip wrapped in a wire coil. The bimetallic strip serves as one of the contact points.

When the turn signals are actuated, current flows into the flasher- first through the heating coil to the bimetallic strip, then through the contact points, then out of the flasher, where the circuit is completed through the turn-signal light. This sequence of events will repeat a few times a second, causing a steady flashing of the turn signals.

Electrical and Electronic Fuel Lift Pumps

In many modern cars the fuel pump is usually electric and located inside the fuel tank. The pump creates positive pressure in the fuel lines, pushing the gasoline to the engine. The higher gasoline pressure raises the boiling point. Placing the pump in the tank puts the component least likely to handle gasoline vapor well (the pump itself) farthest from the engine, submersed in cool liquid. Another benefit to placing the pump inside the tank is that it is less likely to start a fire. Though electrical components (such as a fuel pump) can spark and ignite fuel vapors, liquid fuel will not explode (see flammability limit) and therefore submerging the pump in the tank is one of the safest places to put it. In most cars, the fuel pump delivers a constant flow of gasoline to the engine; fuel not used is returned to the tank. This further reduces the chance of the fuel boiling, since it is never kept close to the hot engine for too long.

The ignition switch does not carry the power to the fuel pump; instead, it activates a relay which will handle the higher current load. It is common for the fuel pump relay to become oxidized and cease functioning; this is much more common than the actual fuel pump failing. Modern engines utilize solid-state control which allows the fuel pressure to be controlled via pulse-width modulation of the pump voltage. This increases the life of the pump, allows a smaller and lighter device to be used, and reduces electrical load.

Cars with electronic fuel injection have an electronic control unit (ECU) and this may be programmed with safety logic that will shut the electric fuel pump off, even if the engine is running. In the event of a collision this will prevent fuel leaking from any ruptured fuel line. Additionally, cars may have an inertia switch (usually located underneath the front passenger seat) that is "tripped" in the event of an impact, or a roll-over valve that will shut off the fuel pump in case the car rolls over.

Some ECUs may also be programmed to shut off the fuel pump if they detect low or zero oil pressure, for instance if the engine has suffered a terminal failure (with the subsequent risk of fire in the engine compartment).

The fuel sending unit assembly may be a combination of the electric fuel pump, the filter, the strainer, and the electronic device used to measure the amount of fuel in the tank via a float attached to a sensor which sends data to the dash-mounted fuel gauge. The fuel pump by itself is a relatively inexpensive part. But a mechanic at a garage might have a preference to install the entire unit assembly.

Fuel Level Indicator

Most fuel gauges are operated electrically and are composed of two units- the gauge, mounted on the instrument panel; and the sending unit, mounted in the fuel tank. The ignition switch is included in the fuel gauge circuit, so the gauge operates only when the ignition switch is in the ON position. Operation of the electrical gauge depends on either coil action or thermostatic action. The four types of fuel gauges are as follows:

The THERMOSTATIC FUEL GAUGE, SELF-REGULATING contains an electrically heated bimetallic strip that is linked to a pointer. A bimetallic strip consists of two dissimilar metals that, when heated, expand at different rates, causing it to deflect or bend. In the case of this gauge, the deflection of the bimetallic strip results in the movement of the pointer, causing the gauge to give a reading. The sending unit consists of a hinged arm with a float on the end. The movement of the arm controls a grounded point that makes contact with another point which is attached to an electrically heated bimetallic strip. The heating coils in the tank and the gauge are connected to each other in series.

The THERMOSTATIC FUEL GAUGE, EXTERNALLY REGULATED differs from a self-regulating system in the use of a variable resistance fuel tank sending unit and an external voltage-limiting device. The sending unit controls the gauge through the use of a rheostat (wire wound resistance unit whose value varies with its effective length). The effective length of the rheostat is controlled in the sending unit by a sliding brush that is operated by the float arm. The power supply to the gauge is kept constant through the use of a voltage limiter. The voltage limiter consists of a set of contact points that are controlled by an electrically heated bimetallic arm.

The THERMOSTATIC FUEL GAUGE, DIFFERENTIAL TYPE is similar to the other type of thermostatic fuel gauges, except that it uses two electrically heated bimetallic strips that share equally in operating and supporting the gauge pointer. The pointer position is obtained by dividing the available voltage between the two strips (differential). The tank unit is a rheostat type similar to that already described; however, it contains a wire-wound resistor that is connected between external terminals of one of the gauges of the bimetallic strip. The float arm moves a grounded brush that raises resistance progressively to one terminal, while lowering resistance to the other. This action causes the voltage division and resulting heat differential to the gauge strips formulating the gauge reading.

The MAGNETIC FUEL GAUGE consists of a pointer mounted on an armature. Depending upon the design, the armature may contain one or two poles. The gauge is motivated by a magnetic field that is created by two separate magnetic coils that are contained in the gauge. One of these coils is connected directly to the battery, producing a constant magnetic field. The other coil produces a variable field, whose strength is determined by a rheostat-type tank unit. The coils are placed 90 degrees apart.

Pressure Gauge

A pressure gauge is used widely in automotive and construction applications to keep track of such things as oil pressure, fuel line pressure, air brake system pressure, and the pressure in the hydraulic systems. Depending on the equipment, a mechanical gauge, an electrical gauge, or an indicator lamp may be used.

TEMPERATURE GAUGE

The temperature gauge is a very important indicator in construction and automotive equipment. The most common uses are to indicate engine coolant, transmission, differential oil, and hydraulic system temperature. Depending on the type of equipment, the gauge may be mechanical, electric, or a warning light.

The ELECTRIC GAUGE may be the thermostatic or magnetic type, as described previously. The sending unit (fig. 2-83) that is used varies, depending upon application.

1. The sending unit that is used with the thermostatic gauge consists of two bimetallic strips, each having a contact point. One bimetallic strip is heated electrically. The other strip bends to increase the tension of the contact points. The different positions of the bimetallic strip create the gauge readings.

2. The sending unit that is used with the magnetic gauge contains a device called a thermistor. A thermistor is an electronic device whose resistance decreases proportionally with an increase in temperature.

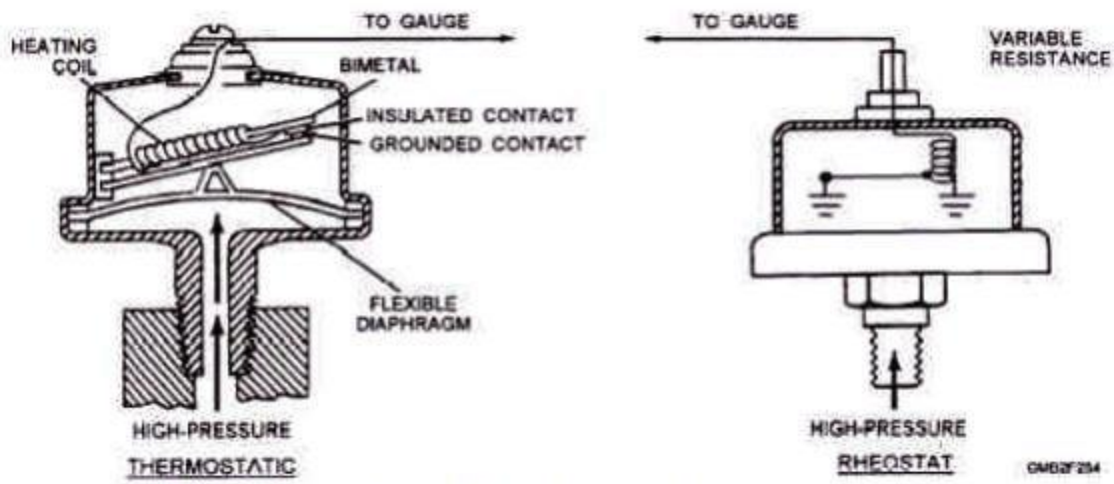


Figure 2-82.- Types of sending units for pressure gauges.

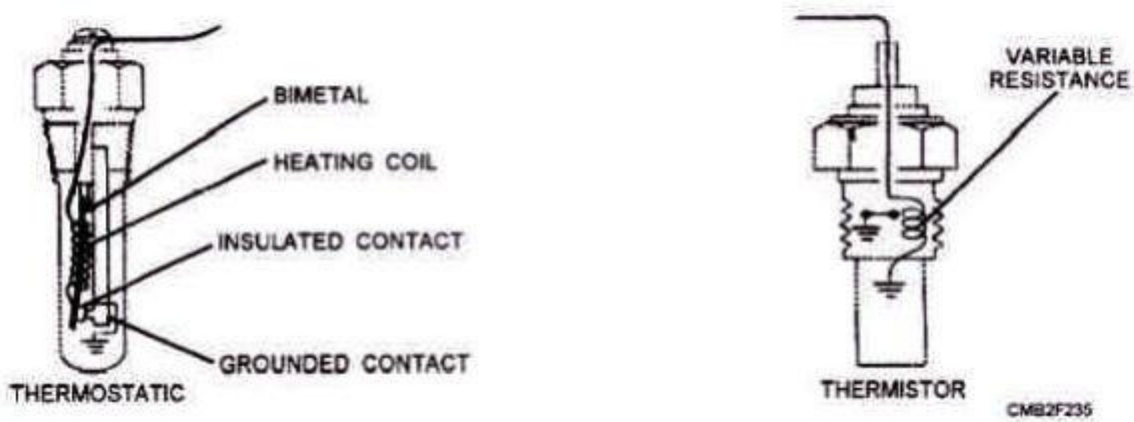


Figure 2-83.- Types of temperature gauge sending units.

SPEEDOMETER AND ODOMETERS

Mechanical Speedometers And Odometers

Both the mechanical speedometer and the tachometer consist of a permanent magnet that is rotated by a flexible shaft. Surrounding the rotating magnet is a metal cup that is attached to the indicating needle. The revolving magnetic field exerts a pull on the cup that forces it to rotate. The rotation of the cup is countered by a calibrated hairspring.

The influence of the hairspring and the rotating magnetic field on the cup produces accurate readings by the attached needle. The flexible shaft consists of a flexible outer casing that is made of either steel or plastic and an inner drive core that is made of wire-wound spring steel. Both ends of the core are molded square, so they can fit into the driving member at one end and the driven member at the other end and can transmit torque.

Gears on the transmission output shaft turn the flexible shaft that drives the speedometer. This shaft is referred to as the speedometer cable. A gear on the ignition distributor shaft turns the flexible shaft that drives the tachometer. This shaft is referred to as the tachometer cable.

The odometer of the mechanical speedometer is driven by a series of gears that originate at a spiral gear on the input shaft. The odometer consists of a series of drums with digits printed on the outer circumference that range from zero to nine. The drums are geared to each other so that each time the one furthest to the right makes one revolution, it will cause the one to its immediate left to advance one digit. The second to the right then will advance the drum to its immediate left one digit for every revolution it makes. This sequence continues to the left through the entire series of drums. The odometer usually contains six digits to record 99,999.9 miles or kilometers. However, models with trip odometers do not record tenths, thereby contain only five digits. When the odometer reaches its highest value, it will automatically reset to zero. Newer vehicles incorporate a small dye pad in the odometer to color the drum of its highest digit to indicate the total mileage is in excess of the capability of the odometer.

Electric Speedometers and Odometers

The electric speedometer and tachometer use a mechanically driven permanent magnet generator to supply power to a small electric motor. The electric motor then is used to rotate the input shaft of the speedometer or tachometer. The voltage from the generator will increase proportionally with speed, and speed will likewise increase proportionally with voltage enabling the gauges to indicate speed.

The signal generator for the speedometer is usually driven by the transmission output shaft through gears. The signal generator for the tachometer usually is driven by the distributor through a power takeoff on gasoline engines. When the tachometer is used with a diesel engine, a special power takeoff provision is made, usually on the camshaft drive.

Electronic Speedometers and Odometers

Electronic speedometers and Odometers are self-contained units that use an electric signal from the engine or transmission. They differ from the electric unit in that they use a generated signal as the driving force. The gauge is transistorized and will supply information through either a magnetic analog (dial) or light-emitting diode (LED) digital gauge display. The gauge unit derives its input signal in the following ways:

An electronic tachometer obtains a pulse signal from the ignition distributor, as it switches the coil on and off. The pulse speed at this point will change proportionally with engine speed. This is the most popular signal source for a tachometer that is used on a gasoline engine.

A tachometer that is used with a diesel engine uses the alternating current generated by the stator terminal of the alternator as a signal. The frequency of the ac current will change proportionally with engine speed.

An electronic speedometer derives its signal from a magnetic pickup coil that has its field interrupted by a rotating pole piece. The signal units operation is the same as the operation of the reluctor and pickup coil described earlier in this TRAMAN. The pickup coil is located strategically in the transmission case to interact with the reluctor teeth on the input shaft.

HORN

The horn currently used on automotive vehicles is the electric vibrating type. The electric vibrating horn system typically consists of a fuse, horn button switch, relay, horn assembly, and related wiring. When the operator presses the horn button, it closes the horn switch and activates the horn relay. This completes the circuit, and current is allowed through the relay circuit and to the horn.

Most horns have a diaphragm that vibrates by means of an electromagnetic. When the horn is energized, the electromagnet pulls on the horn diaphragm. This movement opens a set of contact points inside the horn. This action allows the diaphragm to flex back towards its normal position. This cycle is repeated rapidly. The vibrations of the diaphragm within the air column produce the note of the horn.

Tone and volume adjustments are made by loosening the adjusting locknut and turning the adjusting nut. This very sensitive adjustment controls the current consumed by the horn.

Increasing the current increases the volume. However, too much current will make the horn sputter and may lock the diaphragm.

When an electric horn will not produce sound, check the fuse, the connections, and test for voltage at the horn terminal. If the horn sounds continuously, a faulty horn switch is the most probable cause. A faulty horn relay is another cause of horn problems. The contacts inside the relay may be burned or stuck together.

WINDSHIELD WIPERS

The windshield wiper system is one of the most important safety factors on any piece of equipment. A typical electric windshield wiper system consists of a switch, motor assembly, wiper linkage and arms, and wiper blades. The description of the components is as follows:

The WINDSHIELD WIPER SWITCH is a multi position switch, which may contain a rheostat. Each switch position provides for different wiping speeds. The rheostat, if provided, operates the delay mode for a slow wiping action. This permits the operator to select a delayed wipe from every 3 to 20 seconds. A relay is frequently used to complete the circuit between the battery voltage and the wiper motor.

The WIPER MOTOR ASSEMBLY operates on one, two, or three speeds. The motor has a worm gear on the armature shaft that drives one or two gears, and, in turn, operates the linkage to the wiper arms. The motor is a small, shunt wound dc motor. Resistors are placed in the control circuit from the switch to reduce the current and provide different operating speeds.

The WIPER LINKAGE and ARMS transfer motion from the wiper motor transmission to the wiper blades. The rubber wiper blades fit on the wiper arms.

The WIPER BLADE is a flexible rubber squeegee-type device. It may be steel or plastic backed and is designed to maintain total contact with the windshield throughout the stroke. Wiper blades should be inspected periodically. If they are hardened, cut, or split, they are to be replaced.

When electrical problems occur in the windshield wiper system, use the service manual and its wiring diagram of the circuit. First check the fuses, electrical connections, and all grounds. Then proceed with checking the components.

Ignition System

In spark ignition engines, a device is required to ignite the compressed air-fuel mixture at the end of compression stroke. Ignition system fulfills this requirement. It is a part of electrical system which carries the electric current at required voltage to the spark plug which generates spark at correct time. It consists of a battery, switch, distributor ignition coil, spark plugs and necessary wiring.

A compression ignition engine, i.e. a diesel engine does not require any ignition system. Because, self ignition of fuel air mixture takes place when diesel is injected in the compressed air at high temperature at the end of compression stroke.

Requirements Of An Ignition System

(a) The ignition system should be capable of producing high voltage current, as high as 25000 volts, so that spark plug can produce spark across its electrode gap.

(b) It should produce spark for sufficient duration so that mixture can be ignited at all operating speeds of automobile.

(c) Ignition system should function satisfactory at all engine speeds.

(d) Longer life of contact points and spark plug.

(e) Spark must generate at correct time at the end of compression stroke in every cycle of engine operation.

(f) The system must be easy to maintain, light in weight and compact in size.

(g) There should be provision of spark advance with speed and load.

(h) It should be able to function smoothly even when the spark plug electrodes are deposited with carbon lead or oil.

Types Of Ignition Systems

There are three types of ignition systems which are used in petrol engines.

(a) Battery ignition system or coil ignition system.

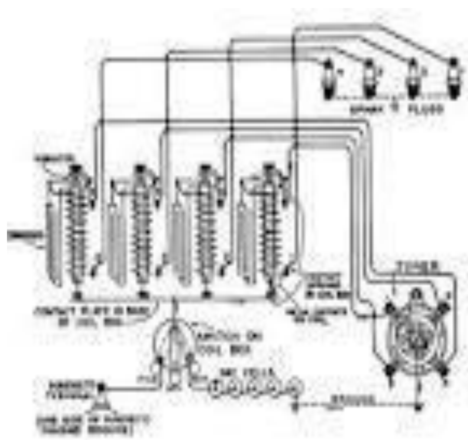
(b) Magneto ignition system.

(c) Electronic ignition system.

In battery ignition system, the current in the primary winding is supplied by a battery whereas it is supplied by a magneto in magneto ignition system.
Battery ignition system is used in cars and light truck. Magneto ignition system is used in some scooters. Both the systems work on the principle of mutual electromagnetic induction. Electronic ignition systems use solid state devices such as transistors and capacitors.

Battery or Coil Ignition System

Battery ignition system consists of a battery of 6 or 12 volts, ignition switch, induction coil, contact breaker, condenser, distributor and spark plugs. A typical battery ignition system for four cylinder SI engine has been shown in Figure



Battery or Coil Ignition System

The primary circuit consists of battery, switch, primary winding and contact breaker point which is grounded. A condenser is also connected in parallel to the contact breaker points. One end of the condenser is grounded and other connected to the contact breaker arm. It is provided to avoid sparking at contact breaker points so as to increase their life.

The secondary ignition circuit consists of secondary winding distributors and spark plugs. All spark plugs are grounded.

The ignition coil steps up 12 volts (or 6 volt) supply to a very high voltage which may range from 20,000 to 30,000 volts. A high voltage is required for the spark to jump across the spark plug gas. This spark ignites the air-fuel mixture as the end of compression stroke. The rotor of the distributor revolves and distributors the current to the four segments which send the current to different spark plugs. For a 4-cylinder engine the cam of the contact breaker has four lobes. Therefore, it makes and breaks the contact of the primary circuit four times in every revolution of cam. Because of which current is distributed to all the spark plugs in some definite sequence.

The primary winding of ignition coil has less number of turns (e.g. 200 turns) of thick wire. The secondary winding has relatively large number of turns (e.g. 20,000 turns) of thin wire.When ignition switch in turned on, the current flows from battery to the primary winding. This produces magnetic field in the coil. When the contact point is open, the magnetic field collapses and the movement of the magnetic field induces current in the secondary winding of ignition coil. As the number of turns in secondary winding are more, a very high voltage is produced across the terminals of secondary.

The distributor sends this high voltage to the proper spark plug which generates spark for ignition of fuel-air mixture. In this way, high voltage current is passed to all spark in a definite order so that combustion of fuel-air mixture takes place in all cylinders of the engine.

A ballast register is connected in series in primary circuit to regulate the current. At the time of starting this register is bypassed so that more current can flow in this circuit. The breaker points are held by a spring except when they are forced apart by lobes of the cam

. **Advantages**

(a) Low initial cost.

(b) Better spark at low speeds and better starting than magneto system.

(c) Reliable system.

(d) No problems due to adjustment of spark timings.

(e) Simpler than magneto system.

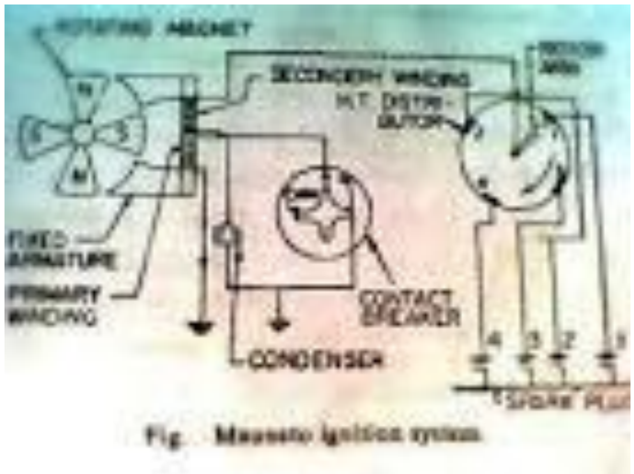
Disadvantages

(a) Battery requires periodical maintenance.

(b) In case of battery malfunction, engine cannot be started.

Magneto-ignition System

This system consists of a magneto in place of a battery. So, the magneto produces and supplies current in primary winding. Rest of the system is same as that in battery ignition system. A magneto ignition system for a four cylinder SI engine has been shown in Figure .
The magneto consists of a fixed armature having primary and secondary windings and a rotating magnetic assembly. This rotating assembly is driven by the engine.



Magneto Ignition System

Rotation of magneto generates current in primary winding having small number of turns. Secondary winding having large number of turns generates high voltage current which is supplied to distributor. The distributor sends this current to respective spark plugs. The magneto may be of rotating armature type or rotating magnet type. In rotating armature type magneto, the armature having primary and secondary windings and the condenser rotates between the poles of a stationary horse shoe magnet. In magneto, the magnetic field is produced by permanent magnets.

Advantages

(a) Better reliability due to absence of battery and low maintenance.

(b) Better suited for medium and high speed engines.

(c) Modern magneto systems are more compact, therefore require less space.

Disadvantages

(a) Adjustment of spark timings adversely affects the voltage.

(b) Burning of electrodes is possible at high engine speeds due to high voltage.

(c) Cost is more than that of magneto ignition systems.

FUNCTIONS OF COMPONENTS USED IN CIRCUITS

Functions of various components used in battery (coil) ignition and magneto-ignition systems are discussed here in brief.

Battery

It is an important component of electrical system. The battery supplies the necessary current to the primary winding of ignition coil which is converted into high voltage current to produce spark. It also supplied current to run the starting motor when engine is cranked for starting. A battery stores energy in the form of chemical energy and supplies it for running lights and other accessories of an automobile. Lead-acid battery is commonly used in most of the automobiles.

Ignition or Induction Coil

The ignition coil is step up transformer to increase the voltage form 12 volt or 6 volt to 20000-30000 volts. It consists of a primary winding and a secondary winding wound on a laminated soft iron core. Primary winding contains about 300 turns made of thick wire. Secondary consists of about 20000 turns of thin wire. In a can type coil, secondary is wound on the soft core over which primary is wound. This assembly is housed in a steel casing fitted with a cap. The cap is made of insulating material. The terminals for electrical connections are provided in cap. To save the windings from moisture and to improve insulation, windings are dipped in oil.

Contact Breakers

Contact breaker is required to make contact and break contact of the primary circuit of ignition system. It consists of two contact breaker points as shown in Figures. One point remains fixed while the other can move. A cam is sued to move the movable point. As cam moves, the contact is made and broken alternately. Primary circuit breaks when the breaker points open. Magnetic field collapses due to this. This produces high voltage current in the secondary winding which is supplied to the distributor. This current is distributor to proper spark plug where it produces spark for ignition of fuel-air mixture.

Condenser

The function of the condenser in the ignition system is to absorb and store the inductive current generated in the coil. If condenser is not provided, the induced current will cause arcing at the breaker points. This will cause burning of the breaker points.

Distributor

The distributor sends the high voltage current, generated in the secondary winding, to the proper spark plug at proper time. If the automobile is having a four cylinder engine, it will have four spark plugs.

The cap of the distributor is connected to the secondary winding of coil. It has a rotor which rotates and comes in contact with the terminals (4 in number for 4 spark plugs) placed around the rotor. As the rotor comes in contact with the terminals (numbered 1, 2, 3 and 4 in Figures), the current is passed to the respective spark plug at proper time when spark is needed.

Ignition Switch

The function of the ignition switch is to connect the battery and starting motor in the automobiles having self starting system.

Example : In car, jeep, etc.

Its function is to connect battery to induction coil in the battery ignition system.

Spark Plugs

The function of the spark plug is to produce spark between its electrodes. This spark is used to ignite the fuel-air mixture in the spark ignition (SI) engines.

Magneto

Magneto is used in magneto ignition system. Magneto is a kind of generator to provide electrical energy to run the ignition system. It is replacement of battery for ignition. When it is rotated by the engine, it produces high voltage current to be supplied to spark plugs through the distributor.

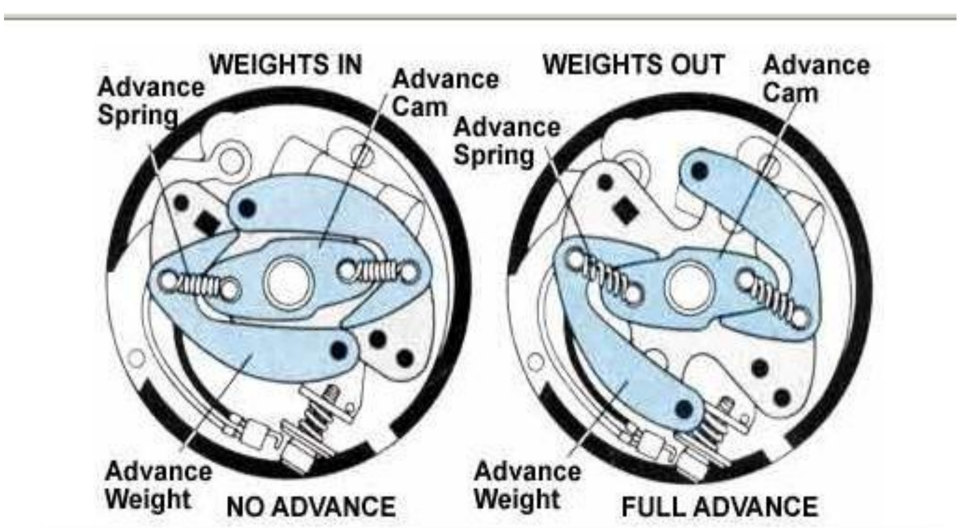
Ignition Advance

The purpose of spark advance mechanism is to assure that under every condition of engine operation, ignition takes place at the most favorable instant in time i.e. most favorable from a standpoint of engine power, fuel economy and minimum exhaust dilution. By means of these mechanisms the advance angle is accurately set so that ignition occurs before TDC point of the piston. The engine speed and the engine load are the control quantities required for the automatic adjustment of the ignition timing. Most of the engines are fitted with mechanisms which are integral with the distributor and automatically regulate the optimum spark advance to account for change of speed and load. The two mechanisms used are:

(a) Centrifugal advance mechanism, and
(b) Vacuum advance mechanism.

Centrifugal Advance Mechanism

The centrifugal advance mechanism controls the ignition timing for full- load operation. The adjustment mechanism is designed so that its operation results in the desired advance of the spark. The cam is mounted, movably, on the distributor shaft so that as the speed increases, the flyweights which are swung farther and farther outward, shaft the cam in the direction of shaft rotation. As a result, the cam lobes make contact with the breaker lever rubbing block somewhat earlier, thus shifting the ignition point in the early or advance direction.



Depending on the speed of the engine, and therefore of the shaft, the weights are swung outward a greater or a lesser distance from the center. They are then held in the extended position, in a state of equilibrium corresponding to the shifted timing angle, by a retaining spring which exactly balances the centrifugal force. The weights shift the cam either or a rolling contact or sliding contact basis; for this reasons we distinguish between the rolling contact type and the sliding contact type of centrifugal advance mechanism.

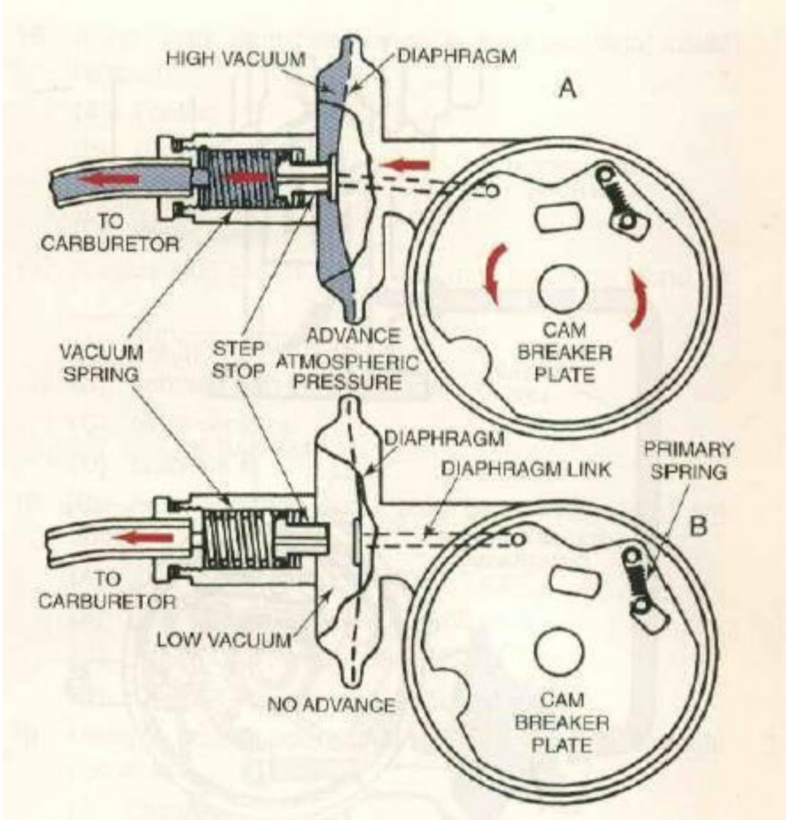
The beginning of the timing adjustment in the range of low engine speeds and the continues adjustment based on the full load curve are determined by the size of the weights by the shape of the contact mechanisms (rolling or sliding contact type), and by the retaining springs, all of which can be widely differing designs. The centrifugal force controlled cam is fitted with a lower limit stop for purposes of setting the beginning of the adjustment, and also with an upper limit stop to restrict the greatest possible full load adjustment.

Vacuum Advance Mechanism

Vacuum advance mechanism shifts the ignition point under partial load operation.
The adjustment system is designed so that its operation results in the prescribed partial load advance curve. In this mechanism the adjustment control quantity is the static vacuum prevailing in the carburetor, a pressure which depends on the position of the throttle valve at any given time and which is at a maximum when this valve is about half open. This explains the vacuum maximum.

The diaphragm of a vacuum unit is moved by changes in gas pressure. The position of this diaphragm is determined by the pressure differential at any given moment between the prevailing vacuum and atmospheric pressure. The beginning of adjustment is set by the pre-established tension on a compression spring. The diaphragm area, the spring force, and the spring rigidity are all selected in accordance with the partial –load advance curve which is to be followed and are all balanced with respect to each other. The diaphragm movement is transmitted through a vacuum advance arm connected to the movable breaker plate, and this movement shifts the breaker plate an additional amount under partial load Ignition Systems condition in a direction opposite to the direction of rotation of the distributor shaft. Limit stops on the vacuum advance arm in the base of the vacuum unit restrict the range of adjustment.

The vacuum advance mechanism operates independent of the centrifugal advance mechanism. The mechanical interplay between the two advance mechanisms, however, permits the total adjustment angle at any given time to be the result of the addition of the shifts provided by the two individual mechanisms operates in conjunction with the engine is operating under partial load.



SPARK PLUGS

The simple requirement of a spark plug is that it must allow a spark to form within the combustion chamber, to initiate burning. In order to do this the plug has to withstand a number of severe conditions. Consider, as an example, a four-cylinder four-stroke engine with a compression ratio of 9:1, running at speeds up to 5000 rev/min. The following conditions are typical. At this speed the four-stroke cycle will repeat every 24 ms.

_ End of induction stroke –0.9 bar at 65 ° C.

_ Ignition firing point –9 bar at 350 ° C.

_ Highest value during power stroke –45 bar at 3000 ° C.

_ Power stroke completed –4 bar at 1100 ° C.

Besides the above conditions, the spark plug must withstand severe vibration and a harsh chemical environment. Finally, but perhaps most important, the insulation properties must withstand voltage pressures up to 40kV.

Construction

The centre electrode is connected to the top terminal by a stud. The electrode is constructed of a nickel-based alloy. Silver and platinum are also used for some applications. If a copper core is used in the electrode this improves the thermal conduction properties.

The insulating material is ceramic-based and of a very high grade. Aluminium oxide, Al2O3 (95% pure), is a popular choice, it is bonded into the metal parts and glazed on the outside surface. The properties of this material, which make it most suitable, are as follows:

_ Young's modulus: 340kN/mm2.

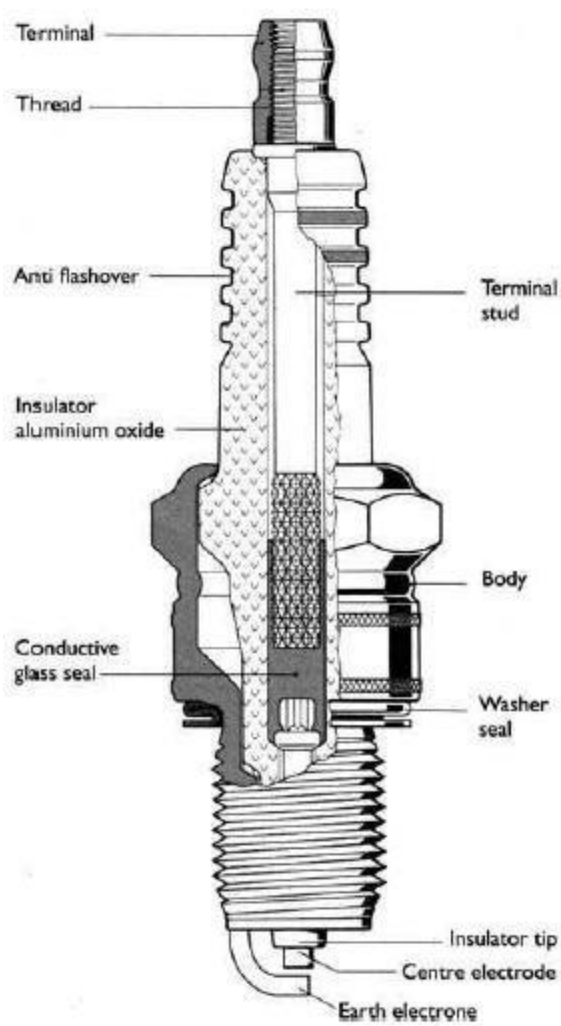
_ Coefficient of thermal expansion: 7.8 x10K^{-1}.

_ Thermal conductivity: 15–5W/mK (Range 200–900 ° C).

_ Electrical resistance: $10^{13}\Omega$/m.

The above list is intended as a guide only, as actual values can vary widely with slight manufacturing changes. The electrically conductive glass seal between the electrode and terminal stud is also used as a resistor. This resistor has two functions. First, to prevent burn-off of the centre electrode, and secondly to reduce radio interference. In both cases the desired effect is achieved because the resistor damps the current at the instant of ignition.

Flash-over, or tracking down the outside of the plug insulation, is prevented by ribs that effectively increase the surface distance from the terminal to the metal fixing bolt, which is of course earthed to the engine.



Electrode materials

The material chosen for the spark plug electrode must exhibit the following properties:

_ High thermal conductivity.

_ High corrosion resistance.

_ High resistance to burn-off.

For normal applications, alloys of nickel are used for the electrode material. Chromium, manganese, silicon and magnesium are examples of the alloying constituents. These alloys exhibit excellent properties with respect to corrosion and burn-off resistance.

To improve on the thermal conductivity, compound electrodes are used. These allow a greater nose projection for the same temperature range, as discussed in the last section. A common example of this type of plug is the copper-core spark plug.

Silver electrodes are used for specialist applications as silver has very good thermal and electrical properties. Again, with these plugs nose length can be increased within the same temperature range.

The thermal conductivity of some electrode materials is listed for comparison.

_ Silver 407 W/m K

_ Copper 384 W/m K

_ Platinum 70 W/m K

_ Nickel 59 W/m K

Compound electrodes have an average thermal conductivity of about 200 W/m K. Platinum tips are used for some spark plug applications due to the very high burn-off resistance of this material. It is also possible because of this to use much smaller diameter electrodes, thus increasing mixture accessibility. Platinum also has a catalytic effect, further accelerating the combustion process.

Electrode gap

Spark plug electrode gaps have, in general, increased as the power of the ignition systems driving the spark has increased. The simple relationship between plug gap and voltage required is that, as the gap increases so must the voltage (leaving aside engine operating conditions). Furthermore, the energy available to form a spark at a fixed engine speed is constant, which means that a larger gap using higher voltage will result in a shorter duration spark. A smaller gap will allow a longer duration spark. For cold starting an engine and for igniting weak mixtures, the duration of the spark is critical. Likewise the plug gap must be as large as possible to allow easy access for the mixture in order to prevent quenching of the flame.

The final choice is therefore a compromise reached through testing and development of a particular application. Plug gaps in the region of 0.6–1.2 mm seem to be the norm at present.

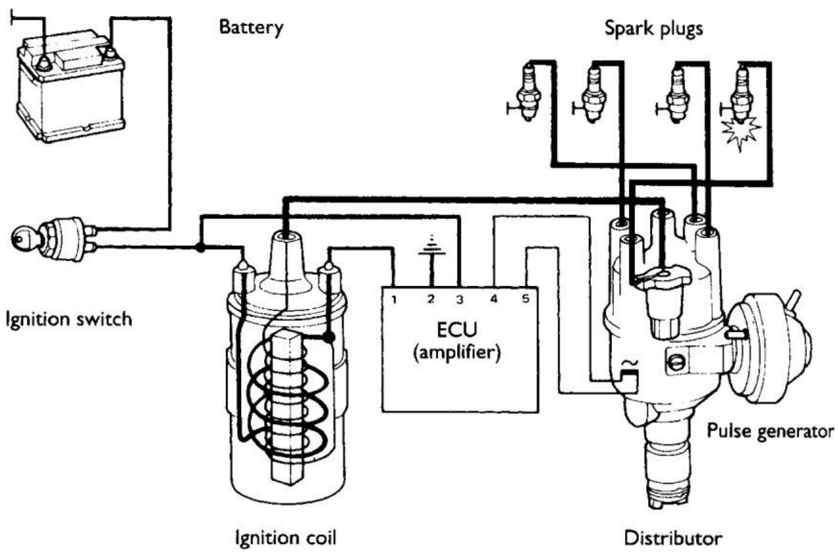
Electronic Ignition System

Electronic ignition is now fitted to almost all spark ignition vehicles. This is because the conventional mechanical system has some major disadvantages.

● Mechanical problems with the contact breakers, not the least of which is the limited lifetime.

● Current flow in the primary circuit is limited to about 4 A or damage will occur to the contacts – or at least the lifetime will be seriously reduced.

● Legislation requires stringent emission limits, which means the ignition timing must stay in tune for a long period of time.



Electronic Ignition System

● Weaker mixtures require more energy from the spark to ensure successful ignition, even at very high engine speed.

These problems can be overcome by using a power transistor to carry out the switching function and a pulse generator to provide the timing signal. Very early forms of electronic ignition used the existing contact breakers as the signal provider. This was a step in the right direction but did not overcome all the mechanical limitations, such as contact bounce and timing slip. Most (all?) systems nowadays are constant energy, ensuring high performance ignition even at high engine speed. Figure the circuit of a standard electronic ignition system.

Distributorless Ignition

Distributorless ignition systems (DIS) have been around for almost a decade now, and have eliminated much of the maintenance that used to be associated with the ignition system. No distributor means there's no distributor cap or rotor to replace, and no troublesome vacuum or mechanical advance mechanisms to cause timing problems. Consequently, DIS ignition systems are pretty reliable.

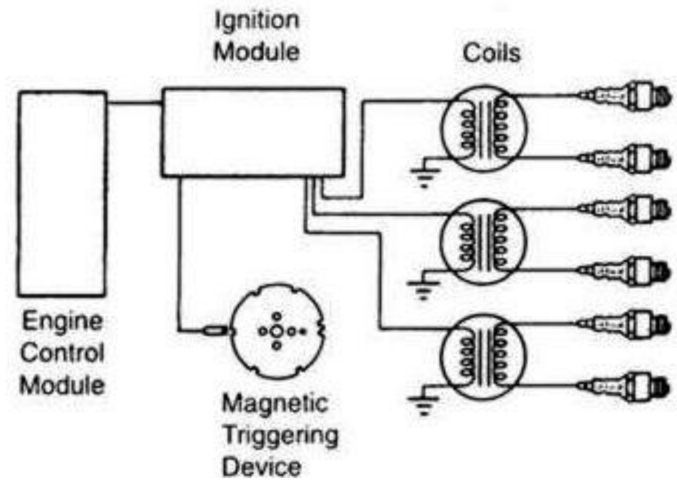
Even so, that doesn't mean they are trouble-free. Failures can and do occur for a variety of reasons. So knowing how to identify and diagnose common DIS problems can save you a lot of guesswork the next time you encounter an engine that cranks but refuses to start, or one that runs but is missing or misfiring on one or more cylinders.

If an engine cranks but won't start, is it fuel, ignition or compression? Ignition is usually the easiest of the three to check because on most engines, all you have to do is pull off a plug wire and check for spark when the engine is cranked. On coil-over-plug DIS systems, there are no plug wires so you have to remove a coil and use a plug wire or adapter to check for a spark.

If there's no spark in one cylinder, try another. No spark in any cylinder would most likely indicate a failed DIS module or crankshaft position sensor. Many engines that are equipped with electronic fuel injection also use the crankshaft position sensor signal to trigger the fuel injectors. So, if there's no spark and no injector activity, the problem is likely in the crank position sensor. No spark in only one cylinder or two cylinders that share a coil would tell you a coil has probably failed.

Principle of Operation

Distributorless ignition system used extensively by Ford incorporates all the features of electronic spark advance systems, except a special type of ignition coil is used in place of HT distributor. The system is generally used only on four- or six-cylinder engines, because the control system becomes highly complex for higher number of cylinders. It works on the principle of the lost spark. The spark distribution is achieved by the help of two double ended coils, fired alternately by the ECU. The ignition timing is obtained from a crankshaft speed and position sensor as well as through load and other corrections. When one of the coils is fired, a spark is delivered to two engine cylinders, either 1 and 4, or 2 and 3. The spark delivered to the cylinder on the compression stroke ignites the mixture as normal. Whereas the spark in other cylinder causes no effect, as this cylinder is just completing its exhaust stroke. Because of the low compression and the exhaust gases in the lost spark cylinder, the voltage only of about 3 kV is needed for the spark to jump the gap. This is similar to cap voltage of the more conventional rotor arm. The spark produced in the compression cylinder is therefore not affected. It may be noted that the spark on one of the cylinders jumps from the earth electrode to the spark plug centre, whereas in others it jumps from the centre electrode. This is because the energy available from modern constant energy systems produces a spark of suitable quality in either direction. However, the disadvantage is that the spark plugs may wear more quickly with this system.



System Components

The distributorless ignition system contains three main components such as the electronic module, a crankshaft position sensor and the distributorless ignition coil. Many systems use a manifold absolute pressure sensor, integrated in the module. The module functions almost in the same way as the electronic spark advance system.

The crankshaft position sensor operates in the similar way to the one described in the previous section. It is also a reluctance sensor positioned against the front of the flywheel or against a reluctor wheel just behind the front crankshaft pulley. The tooth pattern uses 36-1 teeth, which are spaced at 10 degree intervals, with a gap for the 36th tooth. The missing tooth is located at 90 degrees before TDC for numbers 1 and 4 cylinders. This reference position is located a fixed number of degrees before TDC for calculating the timing or ignition point as a fixed angle after the reference mark.

The distributorless ignition coil (Fig. 16.56) has a low tension winding, which is supplied with battery voltage to a centre terminal. The appropriate half of the winding is then connected to earth in the module. The high tension windings are separate and are specific to cylinders 1 and 4, or 2 and 3. Figure 16.57 shows a typical Ford distributorless ignition coil. The Citroen 2 CV has been using a double ended ignition coil together with contact breakers for many years.

Fault Diagnosis

The distributorless ignition system is highly reliable, specifically because it does not have any moving parts. The normal manufacturers servicing schedule should be adhered to for the replacement of spark plugs (often after 19,200 km operation). Some problems may be faced when trying to examine HT oscilloscope patterns, due to the lack of a king lead. This can be overcome by using a special adapter and shifting the sensing clip to each lead in turn. An ohmmeter can be used to test the distributorless ignition coil. The resistance of each primary winding should be 0.5 Q and the secondary windings between 11 and 16 kQ. The coil produces open circuit voltage in excess of 37 kV. The plug leads have integral retaining clips to prevent water ingress and vibration problems. The maximum resistance for the HT leads is 30 kQ per lead. Except for the octane adjustment on some models no service adjustments are possible with this system. This adjustment involves connecting two pins together on the module for normal operation, or earthing one pin or the other to change to a different fuel. The actual procedure as specified by the manufacturer for each particular model should be followed.

DIGITAL IGNITION SYTEM

Electronic Ignition System is as follow :

(a) Capacitance Discharge Ignition system

(b) Transistorized system

(c) Piezo-electric Ignition system

(d) The Texaco Ignition system

Capacitance Discharge Ignition System

It mainly consists of 6-12 V battery, ignition switch, DC to DC convertor, charging resistance, tank capacitor, Silicon Controlled Rectifier (SCR), SCR-triggering device, step up transformer, spark plugs.

A 6-12 volt battery is connected to DC to DC converter i.e. power circuit through the ignition switch, which is designed to give or increase the voltage to 250-350 volts. This high voltage is used to charge the tank capacitor (or condenser) to this voltage through the charging resistance. The charging resistance is also so designed that it controls the required current in the

SCR.

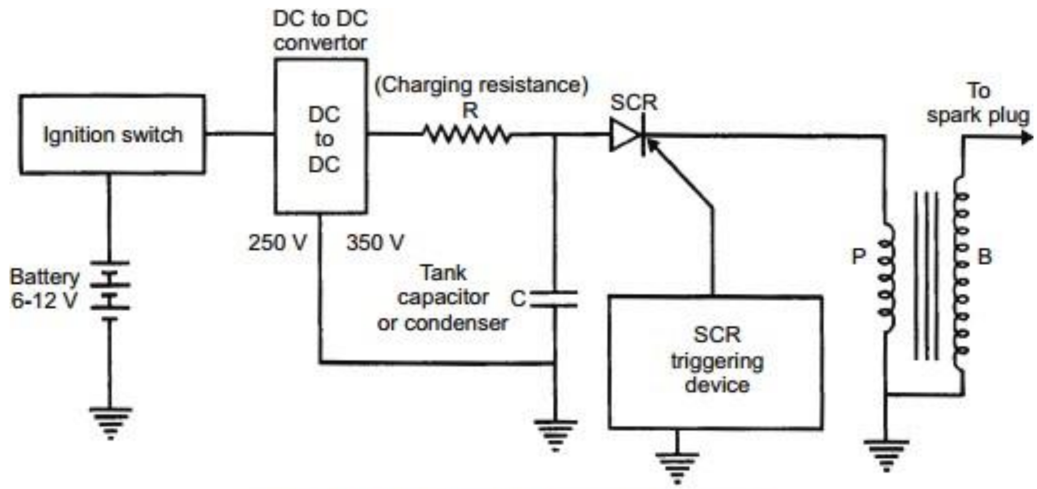
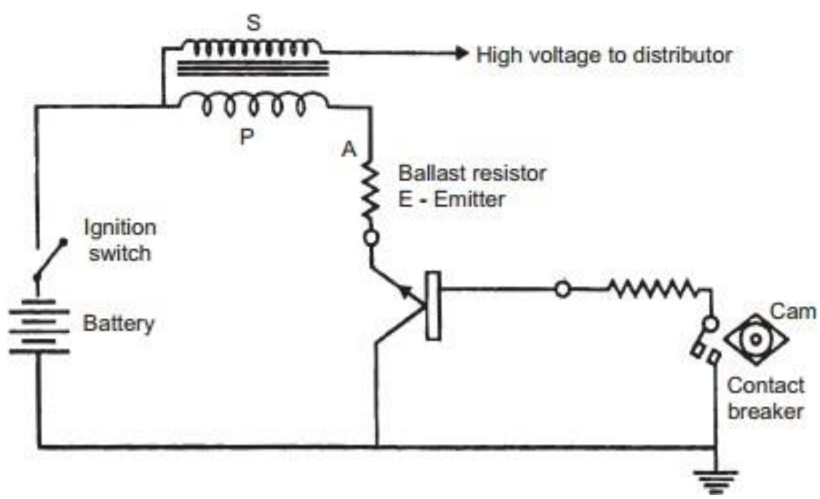


Figure 4.4 : Capacitance Discharge Ignition System

Depending upon the engine firing order, whenever the SCR triggering device, sends a pulse, then the current flowing through the primary winding is stopped. And the magnetic field begins to collapse. This collapsing magnetic field will induce or step up high voltage current in the secondary, which while jumping the spark plug gap produces the spark, and the charge of air fuel mixture is ignited.

Transistorized Assisted Contact (TAC) Ignition System



Advantages

(a) The low breaker-current ensures longer life.

(b) The smaller gap and lighter point assembly increase dwell time minimize contact bouncing and improve repeatability of secondary voltage.

(c) The low primary inductance reduces primary inductance reduces primary current drop-off at high speeds.

Disadvantages

(a) As in the conventional system, mechanical breaker points are necessary for timing the spark.

(b) The cost of the ignition system is increased.

(c) The voltage rise-time at the spark plug is about the same as before.

Piezo-electric Ignition System

The development of synthetic piezo-electric materials producing about 22 kV by mechanical loading of a small crystal resulted in some ignition systems for single cylinder engines. But due to difficulties of high mechanical loading need of the order of 500 kg timely control and ability to produce sufficient voltage, these systems have not been able to come up.

The Texaco Ignition System

Due to the increased emphasis on exhaust emission control, there has been a sudden interest in exhaust gas recirculation systems and lean fuel-air mixtures. To avoid the problems of burning of lean mixtures, the Texaco Ignition system has been developed. It provides a spark of controlled duration which means that the spark duration in crank angle degrees can be made constant at all engine speeds. It is a AC system. This system consists of three basic units, a power unit, a control unit and a distributor sensor. This system can give stable ignition up to A/F ratios as high as 24 : 1.

Non–Contact–type Ignition Triggering devices

Triggering is arranged so that the ignition coil is charged in sufficient time before the actual ignition point. This requires the formation of a dwell period (coil saturation time) in the ignition system. The energy to be released as a spark is usually stored in a coil as magnetic energy (with conventional systems). In other cases, this can be replaced with a capacitor as electrostatic energy, such as in a capacitive discharge ignition system (CDI), in which case the role of the coil changes to simply that of an energy transfer device. The high tension results from disconnecting the primary inductor from the power supply followed by transformation.

The high tension is then applied via the distributor to the cylinder currently performing the working stroke. All this combines to produce the required firing voltage, which is determined by the cylinder pressure, a byproduct of the inlet charge and compression, combined with the gap, temperature and shape of the spark plug electrode. The ignition system will then only deliver the voltage necessary to fire the spark plug. If all is well, the mixture will be successfully ignited. If insufficient energy is available, ignition does not occur, thus allowing a misfire. This is why adequate ignition must be provided.

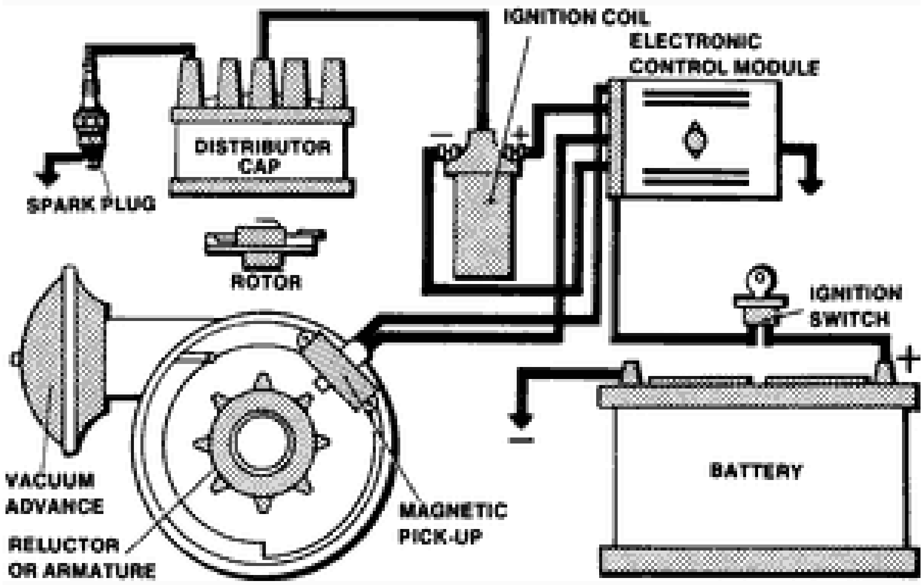
Electronically–Assisted and Full Electronic Ignition System

The need for higher mileage, reduced emissions and greater reliability has led to the development of the electronic ignition systems. These systems generate a much stronger spark which is needed to ignite leaner fuel mixtures. Breaker point systems needed a resistor to reduce the operating voltage of the primary circuit in order to prolong the life of the points. The primary circuit of the electronic ignition systems operates on full battery voltage which helps to develop a stronger spark. Spark plug gaps have widened due to the ability of the increased voltage to jump the larger gap. Cleaner combustion and less deposit have led to longer spark plug life.
On some systems, the ignition coil has been moved inside the distributor cap. This system is said to have an internal coil as opposed to the conventional external one.

Electronic Ignition systems are not as complicated as they may first appear. In fact, they differ only slightly from conventional point ignition systems. Like conventional ignition systems, electronic systems have two circuits: a primary circuit and a secondary circuit. The entire secondary circuit is the same as in a conventional ignition system. In addition, the section of the primary circuit from the battery to the battery terminal at the coil is the same as in a conventional ignition system.

Electronic ignition systems differ from conventional ignition systems in the distributor component area. Instead of a distributor cam, breaker plate, points, and condenser, an electronic ignition system has an armature (called by various names such as a trigger wheel, reluctor, etc.), a pickup coil (stator, sensor, etc.), and an electronic control module.
Essentially, all electronic ignition systems operate in the following manner: With the ignition switch turned on, primary (battery) current flows from the battery through the ignition switch to the coil primary windings. Primary current is turned on and off by the action of the armature as it revolves past the pickup coil or sensor. As each tooth of the armature nears the pickup coil, it

creates a voltage that signals the electronic mod



ule toturn off the coil primary current. A timing circuit in the module will turn the current on again

after the coil field has collapsed. When the current is off, however, the magnetic field built up in the coil is allowed to collapse, which causes a high voltage in the secondary windings of the coil. It is now operating on the secondary ignition circuit, which is the same as in a conventional ignition system.

Troubleshooting electronic ignition systems ordinarily requires the use of a voltmeter and/or an ohmmeter. Sometimes the use of an ammeter is also required. Because of differences in design and construction, troubleshooting is specific to each system. A complete troubleshooting guide for you particular application can be found in the Chilton's Total Car Care manual.

Throttle Body Injection(TBI):

TBI fuel injection system is a type of system where the fuel is injected into the throttle body. The throttle body fuel injection system operates by using a single or pair of injectors. The throttle looks like a carburetor without the fuel bowl, the metering jets or the float.

This type of fuel injection system consists of only two major castings the fuel body and the throttle body. The fuel body supplies the fuel while the throttle body has a valve that controls the flow of air. On the throttle, there are ports that gather signals to relay to the manifold absolute pressure sensor and to the emission control system.

TBI Fuel Injection Advantages:

It is less expensive than using other types of fuel injection systems.
It is easier to clean, maintain and service because there are fewer parts.
It is cheaper to manufacture than a port injection system and simpler to diagnose.
It also does not have the same level of injector balance problems that a port injection system might have when the injectors are clogged.

It greatly improves the fuel metering compared to a carburetor.

TBI Fuel Injection Disadvantages:

It is almost the same as a TBI carburetor wherein the fuel is not equally distributed to all the cylinders. This means that the air/fuel mixture injected differs for each cylinder.

It can cool the manifold much faster causing the fuel to puddle and condense in the manifold. The possibility of condensation is much higher since the fuel travels longer from the throttle body to the combustion chamber.

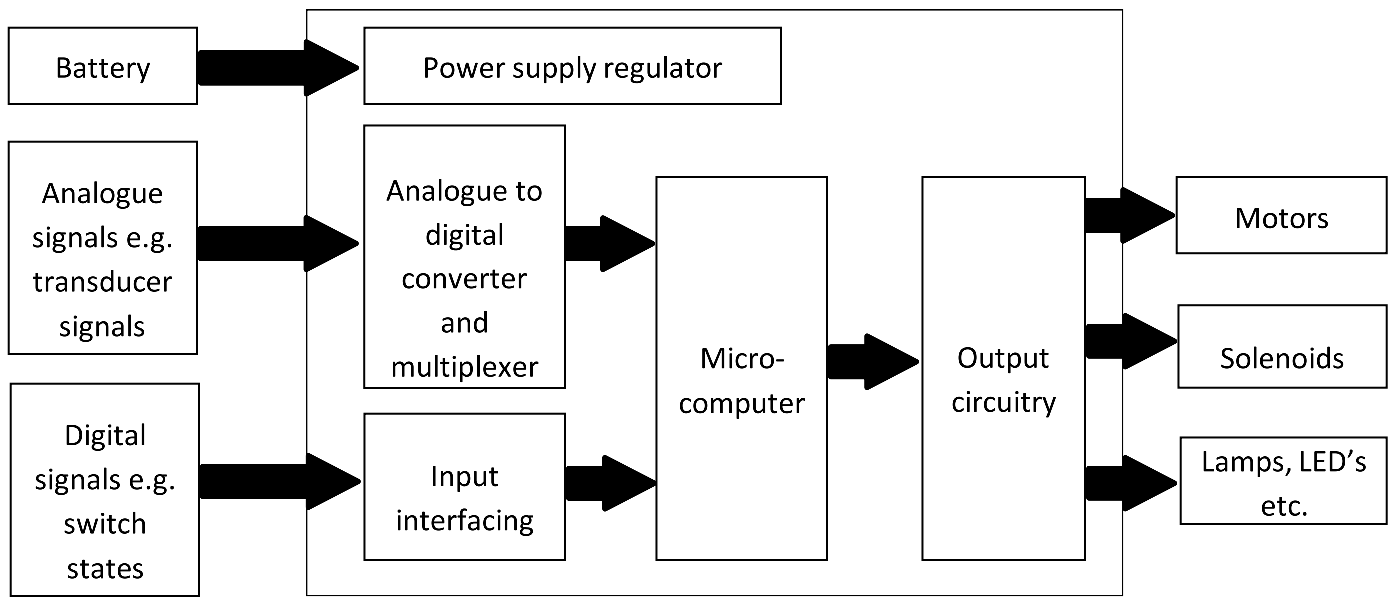
Since the system needs to be mounted on top of the combustion chamber, you're prevented from modifying the manifold design to improve your car's performance.

UNIT-3
Control Systems in Automobiles

Examples of Automotive Closed-loop Control Systems

Control System	**Indirectly controlled variable**	**Directly controlled variable**	**Manipulated variable**	**Sensor**	**Actuator**
Fuel injection system	Air-fuel ratio	Exhaust oxygen content	Quality of injection fuel	Zirconia or Titania based electro-chemical	Fuel injector
Knock control	Knock	Knock sensor output	Ignition timing	Piezo-electric accelerometer	Ignition coil switch. Transistor
Anti-lock braking system	Wheelslip limit	Wheelspeed	Brake time pressure	Magnetic reluctance	ABS solenoid valve

ECU (based on Micro Computers)



Engine Management Sensors

Measured variable	**Direct/indirect measurement**	**Sensor technology/ reference**	**Sensor mounting location**
Intake manifold absolute pressure	Indirect measurement of engine load or mass air-flow intake	Wheatstone bridge arrangement of thick film resistors bonded onto a thin alumina diaphragm	Within intake manifold
Mass airflow	Direct and indirect measurement of fuel injector basic pulse width	Various forms including ‘flap’ type, ‘hot-wire’, Karman vortex and thick-film diaphragm	Within air intake
Temperature	Direct measurement at various locations	Thermistor or thermocouple depending on temperature range	Intake air, outside air, catalytic converter, engine coolant, hydraulic oil
Engine speed and crankshft reference position	Direct measurement	Magnetic reluctance or Hall effect device	Flywheel on end of engine crankshaft

Engine Management Sensors (contd)

Measured variable	**Direct/indirect measurement**	**Sensor technology/ reference**	**Sensor mounting location**
Battery voltage	Direct measurement	Resistive attenuator	
Throttle position	Direct measurement	Potentiometer	Accelerator pedal
Knock (engine cylinder pressure oscillations during ignition)	Direct measurement	Piezoelectric accelerometer type.	Cylinder block or head
Oxygen concentration in exhaust gas (Lambda sensor)	Direct measurement	Zirconia or Titania based exhaust gas oxygen sensors	Exhaust manifold (normal operation above 300^0 C)

Chassis Control Sensors

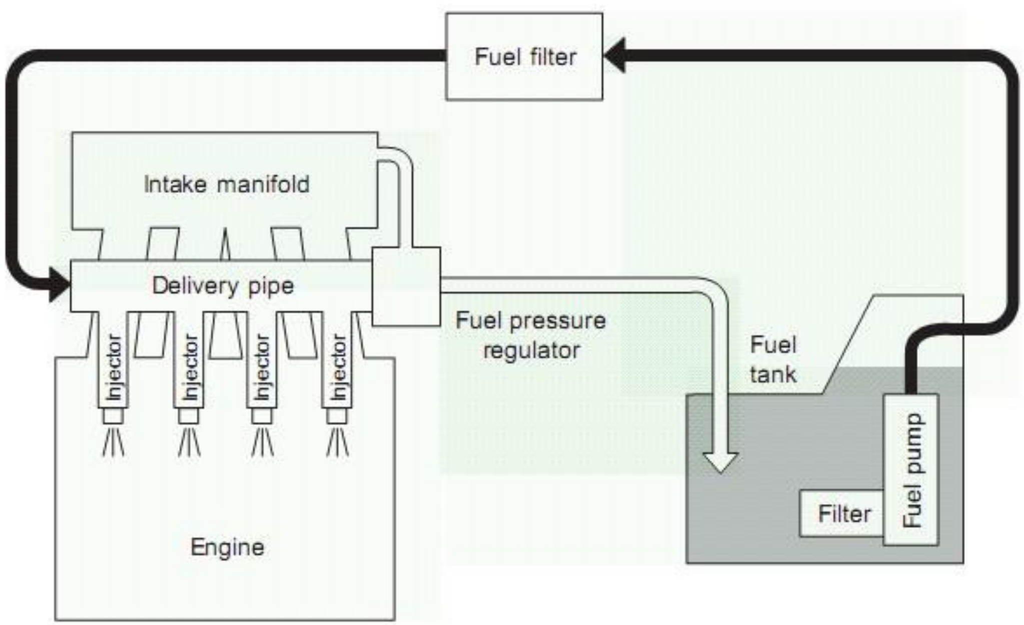
Measured variable and application	**Direct/indirect measurement**	**Sensor technology/ reference**	**Sensor mounting location**
Wheelspeed and engine speed, (ABS, TCS and electronic damping)	Direct measurement	Magnetic reluctance or Hall effect device	Brake assembly and crankshaft flywheel respectively
Steering wheel angle, (Electronic damping)	Direct measurement	Potentiometer or optical encoder	Steering shaft
Throttle position	Indirect measurement of vehicle accel.	Potentiometer	Accelerator pedal
Chassis and wheel acceleration, (electronic damping)	Direct	Piezo-electric accelerometer	Engine compart-ment and wheel assembly
Brake system pressure (electronic damping)	Indirect measurement of vehicle decelerat-ion	Flexing plate sensor with strain gauges mounted on plate	Brake master cylinder
Steering shaft torque (Electric power assisted steering)	Direct measurement	Optical device relying on steering shaft distortion under driver's twisting action	Steering shaft

Safety and Onboard navigation

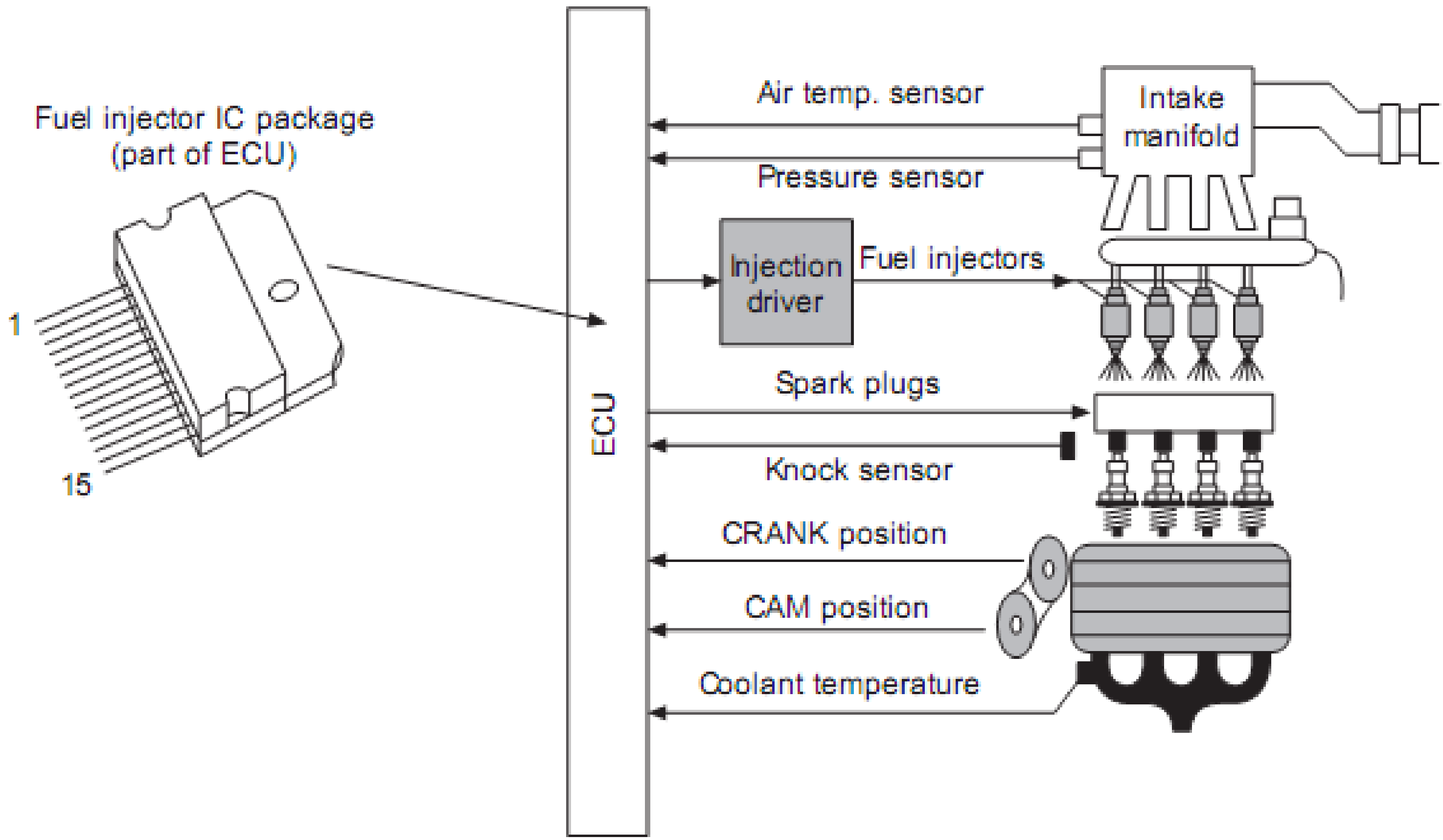
Measured variable	**Direct/indirect measurement**	**Sensor technology/ reference**	**Sensor mounting location**
Vehicle deceleration (air-bag systems)	Direct measurement	‘G’ sensor (Piezo-electric accelerometer)	Single-point electronic sensing, location in dashboard or steering wheel
Wheelspeed and engine speed (Vehicle nav. Systems)	Direct measurement	Magnetic reluctance or Hall effect device	Brake assembly.

Electronic fuel injection (EFI)

- allows precise and fast control of fuel injected
- by control of the 'on-time' period of the solenoid operated injectors (spray nozzle) and plunger.
- delivery pipe fuel pressure is maintained constant by a fuel pressure regulator
- opening and closing times of between 0.5 and 1 ms.
- engine operating speed of 6000 rpm (10 ms revolution time)
- injector on-time can be controlled between 1 and 10 ms.



Power driver application



- multi-point or sequential fuel injection, with one fuel injector near the intake valve (or valves) of each cylinder.
- At a device level, a fuel injector IC package
- provides the high solenoid drive current required
- Incorporates both over-voltage and short-circuit protection,
- fault reporting diagnostic routines also included

Two types of EFI System ----- Speed-density EFI

- inlet manifold absolute pressure (MAP) sensor has an important role
- **fuel injection opening period** or pulse width is related directly to the mass of air flowing into the engine as fuel-air ratio must be maintained constant in steady-state operation
- and the mass of air-flow is related to the manifold absolute pressure by the equation

$$m_a = \frac{V_d n_v P_i}{R T_i}$$

- where V_d is the displacement of the cylinder,
- n_v is the volumetric efficiency or the fraction of V_d actually filled on each stroke, [= f(speed)]
- p_i is manifold absolute pressure,
- R is a constant and
- T_i is the intake air temperature.

Mass air-flow EFI

- direct measurement of the quantity of air drawn into the engine (using an air-flow sensor (AFS)).
  - simple flap-type,
  - hot-wire and
  - Karman vortex devices,
- Direct measurement is better than feed-forward control in speed density EFI
- (factors like variation in volumetric efficiency, engine displacement due to speed and internal deposits need to be taken care of).

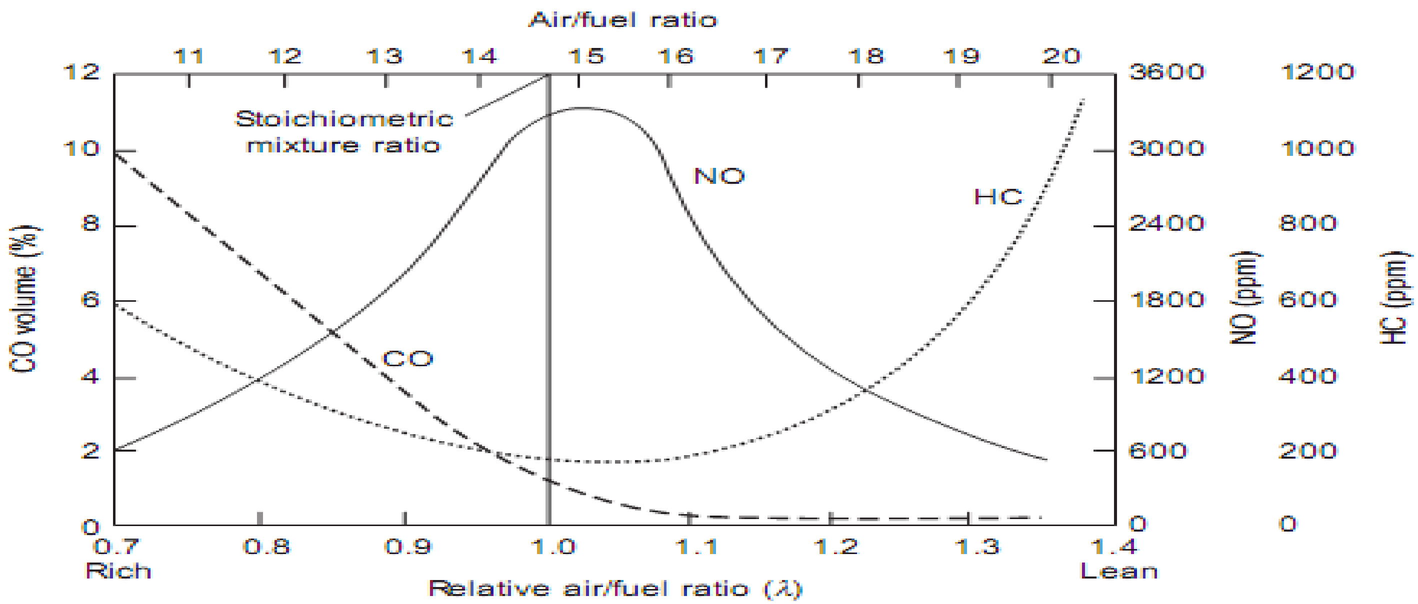
Both of these forms of EFI may be improved

- exhaust gas oxygen sensor for closed-loop control of the air–fuel ratio.
- if engine is to be controlled precisely air–fuel ratio must be controlled to within 1%.
- only possible with closed-loop control,

- ***Closed-loop control of air–fuel ratio***
- The objective of low exhaust-gas emission levels
  - maintain the air–fuel ratio at 14.7:1 [stoichiometrically / chemically perfect]
  - three-way catalytic converters to control emission

- In a closed loop system
- the fuel injection period computed by air intake measurement is modified
  - Based on measured exhaust gas oxygen (EGO) content.
  - injection period modification factor between 0.8 and 1.2.
- EGO tells whether $\lambda < 1$ or $\lambda > 1$
- Closed loop system has a limit cycle frequency between 0.5 to 2 Hz

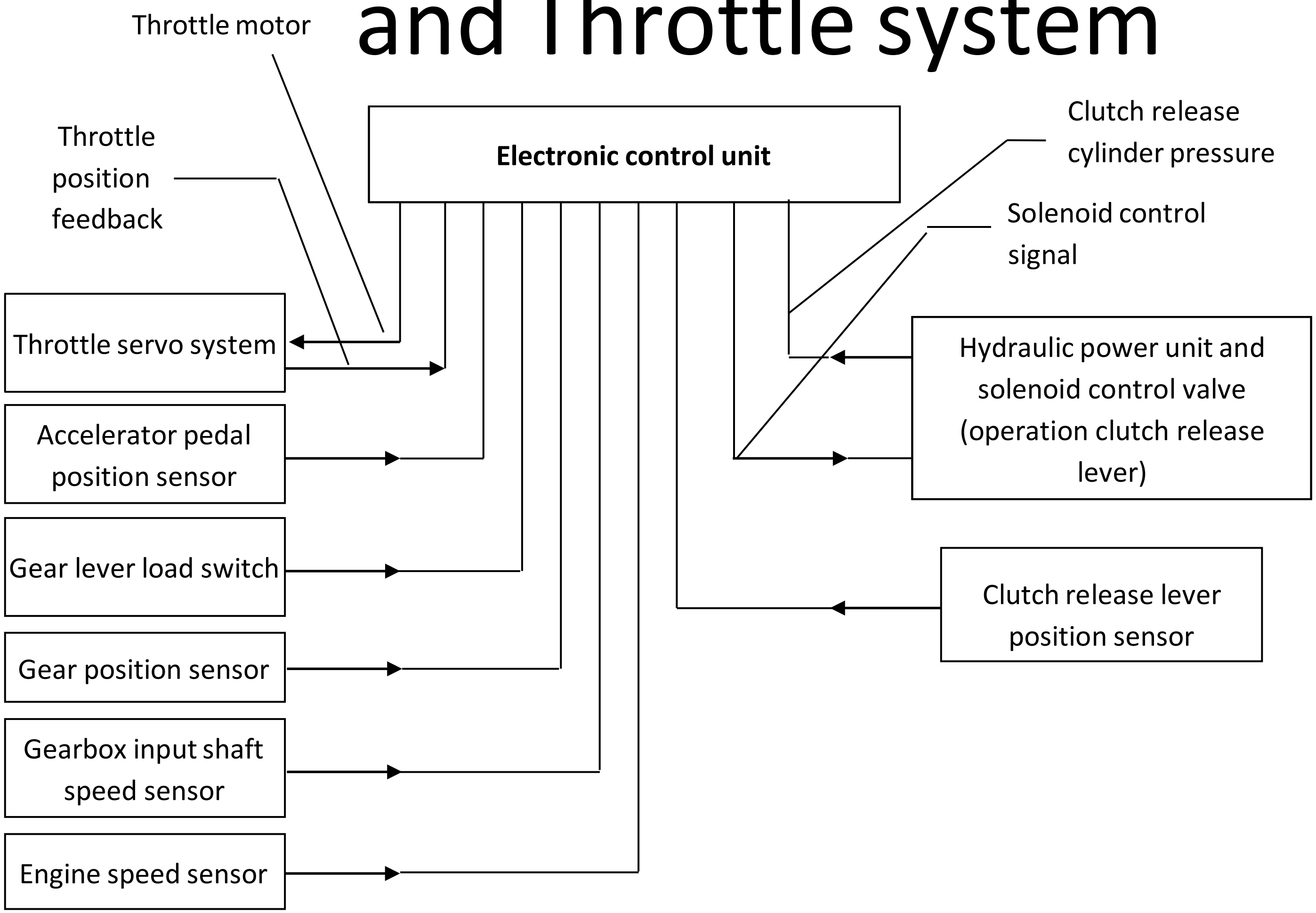
Pollutant emission as a function of relative air–fuel ratio, *l (Chowanietz, 1995)*



Electronic clutch control

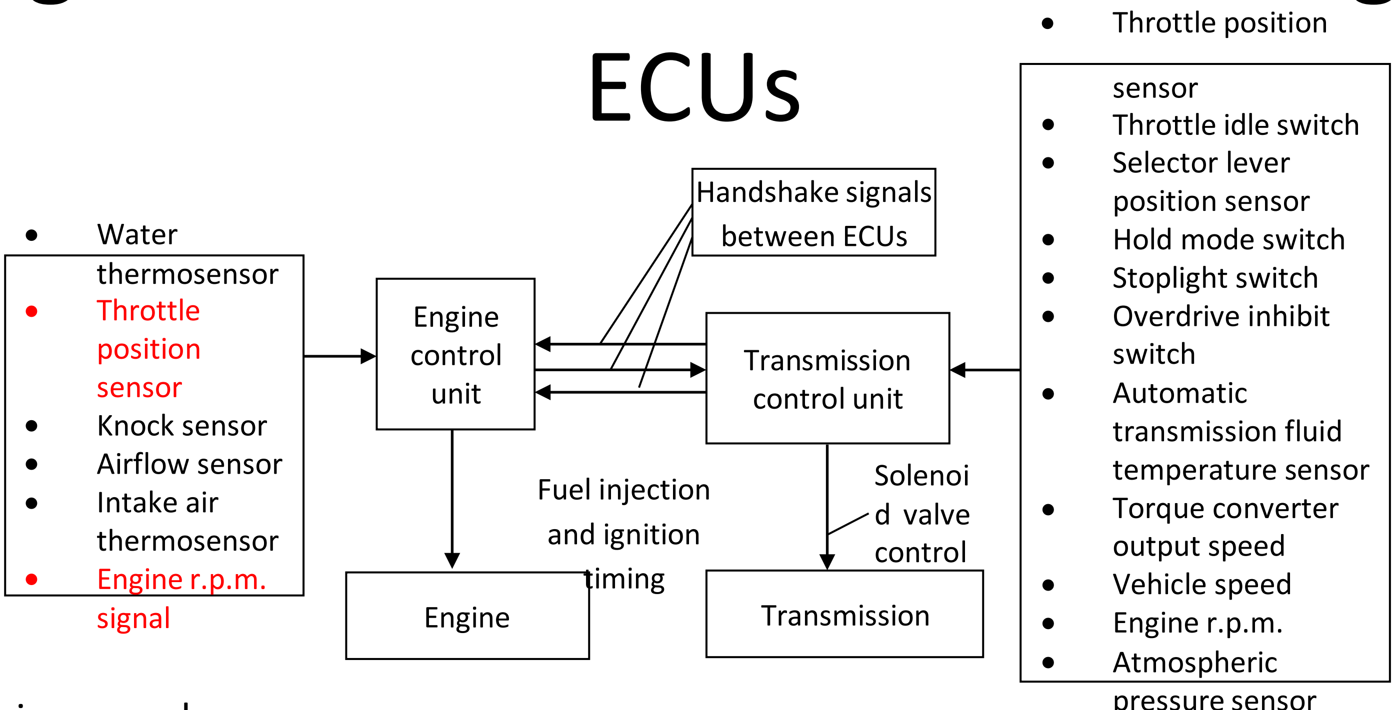
- To relieve pressing of clutch during gear change
- Throttle cable of accelerator pedal replaced by closed loop control system
  - Accelerator pedal position sensor and servomotor
  - Connected to an ECU for the gear change process

Block Diagram of an Automatic Clutch and Throttle system



- Control of clutch engagement and disengagement
- Improved safety
  - Prevention of engine starting when in gear
  - Inappropriate gear change

Integration of Transmission and Engine ECUs



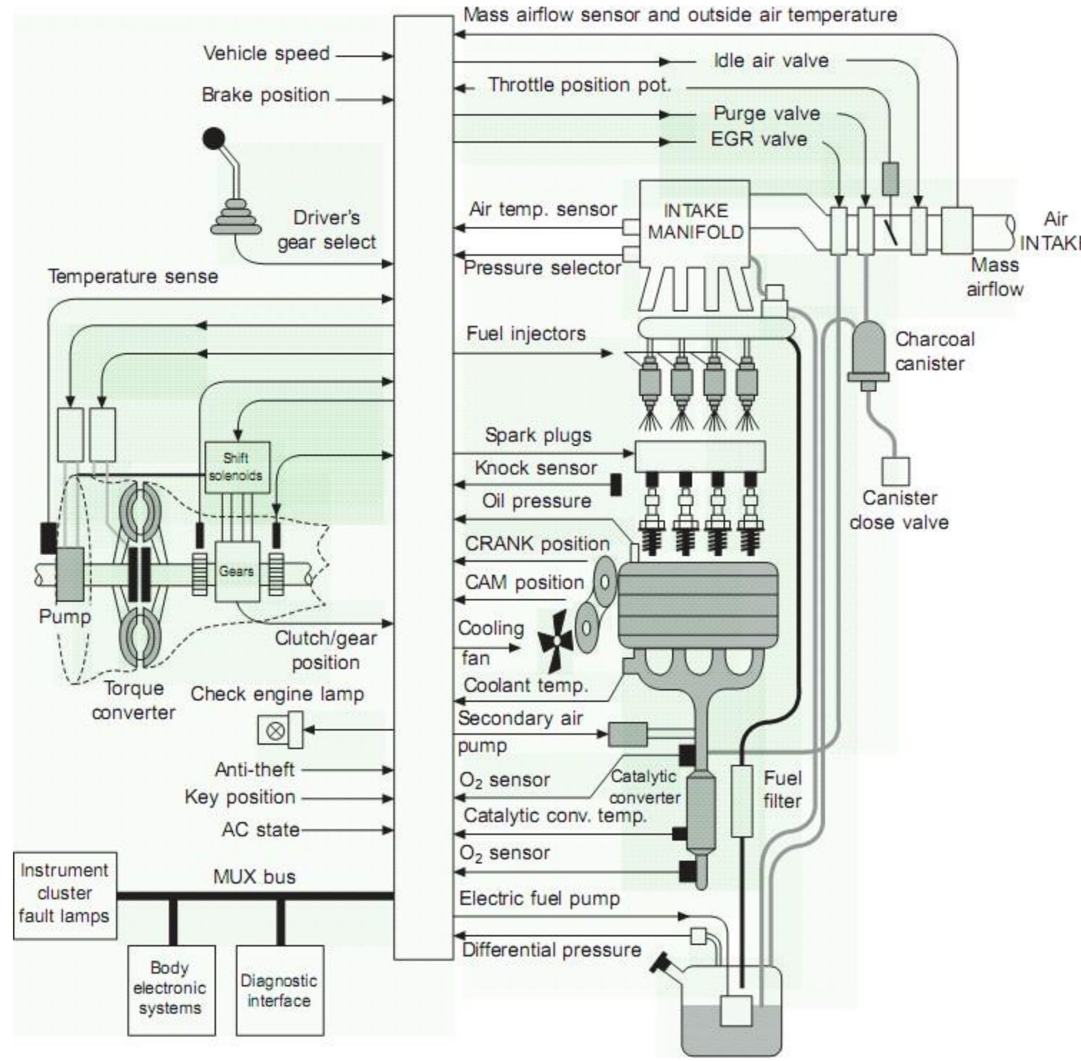
- During gear change up
  - transmission ECU signals the engine management ECU
    - cut off fuel injection and
    - Signals TECU to allow gear change
- During gear change down
  - TECU energizes signals E-ECU
    - Changes ignition timing a few degrees to reduce engine torque,
    - Signals TECU to allow gear change
- In the end both systems return to independent operation.

Integration of engine management and transmission control systems

- the multi-way switch reports position of the selector lever
  - If the lever is not in either Park or Neutral when starting, operation of the starter motor is inhibited
  - a warning buzzer sounded
- The hold switch (push-button switch on the selector lever)
  - instructs the ECU to hold the transmission in a current gear ratio
  - useful in descending a hill.
- The stoplight switch
  - When the brakes are applied and transmission is in a lockup condition,
  - the lockup clutch is then disengaged.
- The overdrive inhibit signal (O/D) (from a separate cruise control unit)
  - prevents the transmission from changing into overdrive (fourth gear) if
    - cruise control is activated and
    - vehicle speed is more than a certain amount below the set cruising speed.
- The Automatic Transmission Fluid (ATF) thermosensor
  - modify the line pressure at temperature extremes
  - To account for changes in fluid viscosity.
- Atmospheric pressure
  - If above 1500 m (engine develops less power at high altitudes)
  - the automatic gear change points are modified to suit the change in performance.

Powertrain Control System

- Also includes
  - Exhaust gas recirculation system (circulating exhaust into intake to reduce max combustion temp, and hence NO_x)
    - Controlled by powertrain ECU
    - Engine temp, load, speed
  - Evaporative Emission Control System (to circulate fuel vapour into intake and prevent leakage into atmosphere)

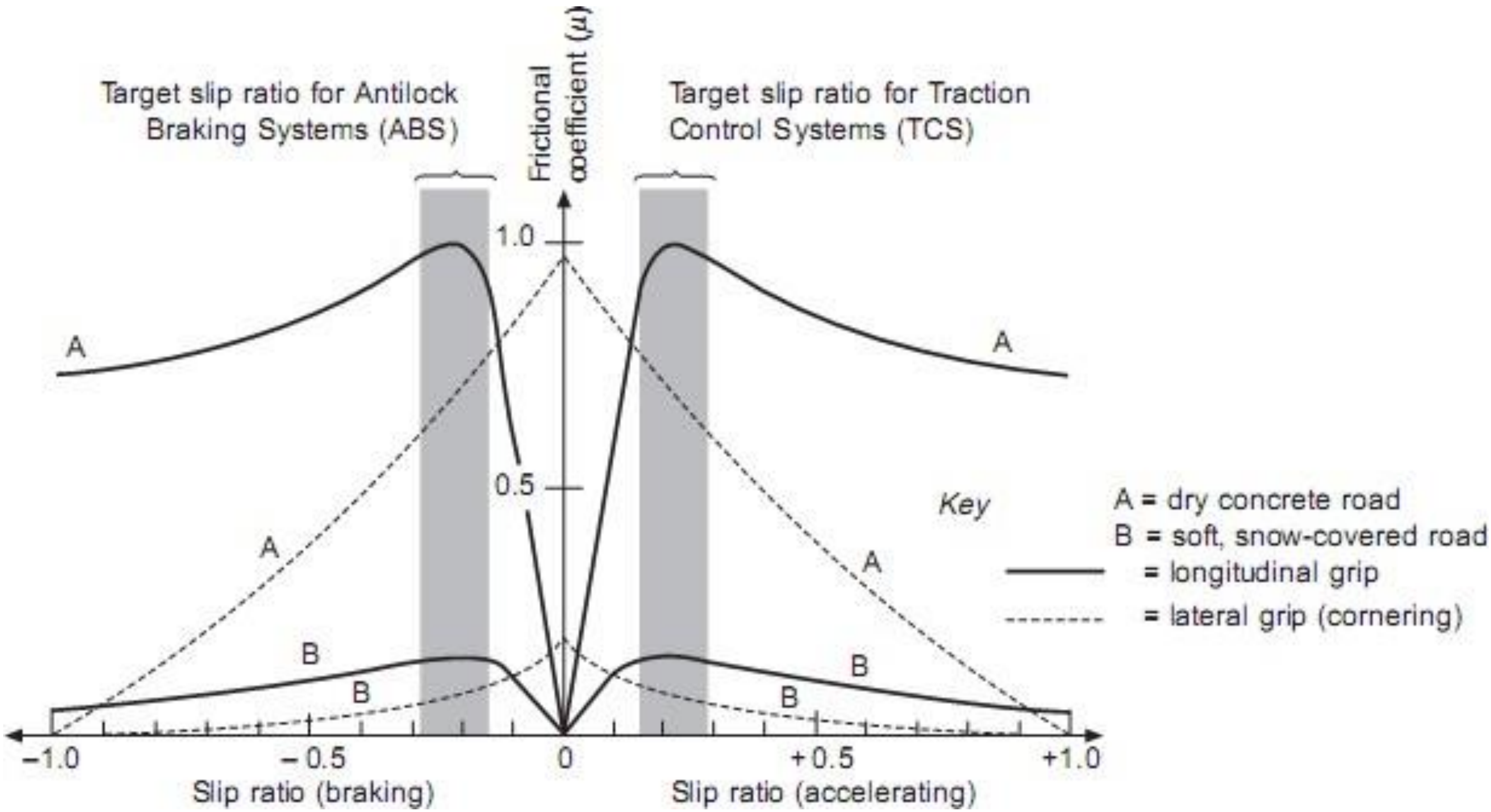


Chassis Control Systems

- Anti Lock Braking system
- Electronic Damping Control system
- Power Assisted Steering System
- Traction Control Systems

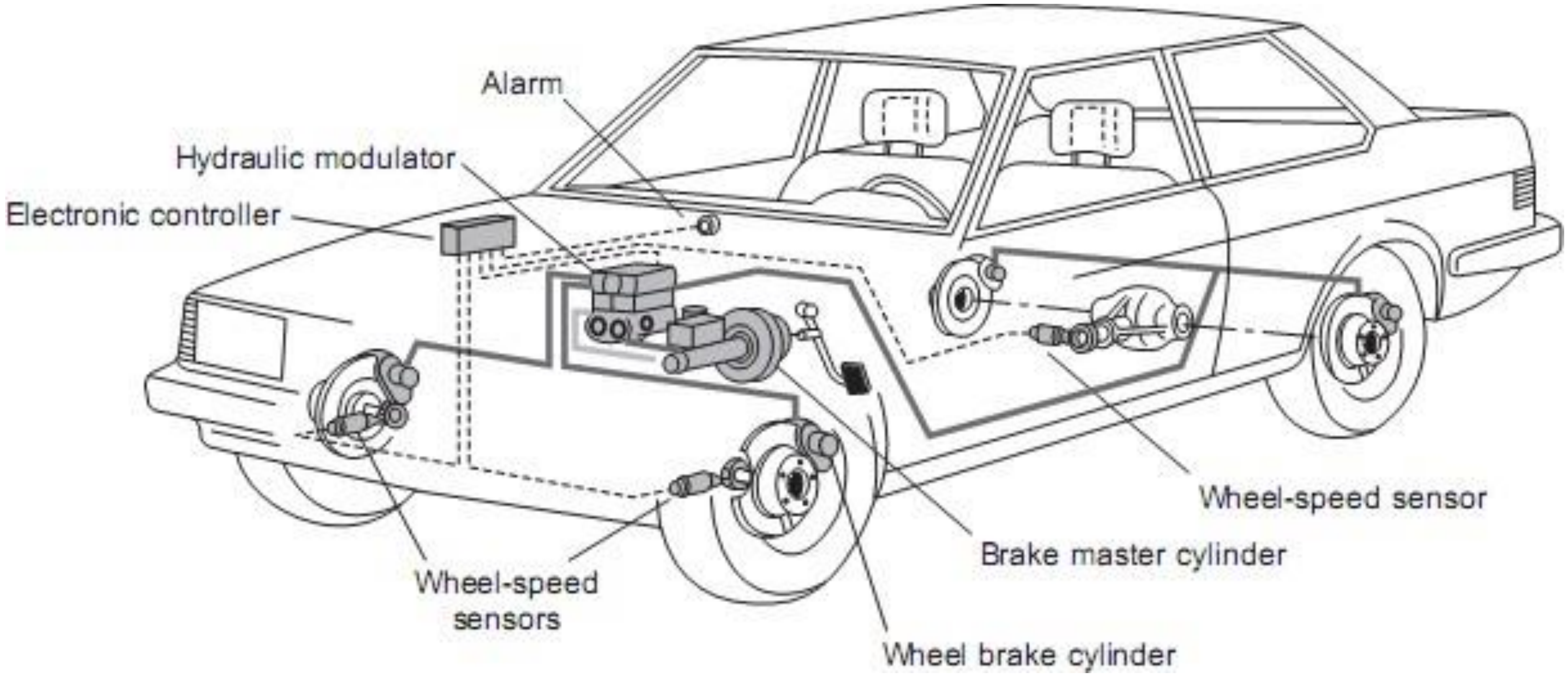
Anti-lock braking systems (ABS)

- The vehicle skids, the wheels lock and driving stability is lost so the vehicle cannot be steered;
- If a trailer or caravan is being towed it may jack-knife;
- The braking distance increases due to skidding;
- The tyres may burst due to excessive friction and forces being concentrated at the points where the locked wheels are in contact with the road surface;



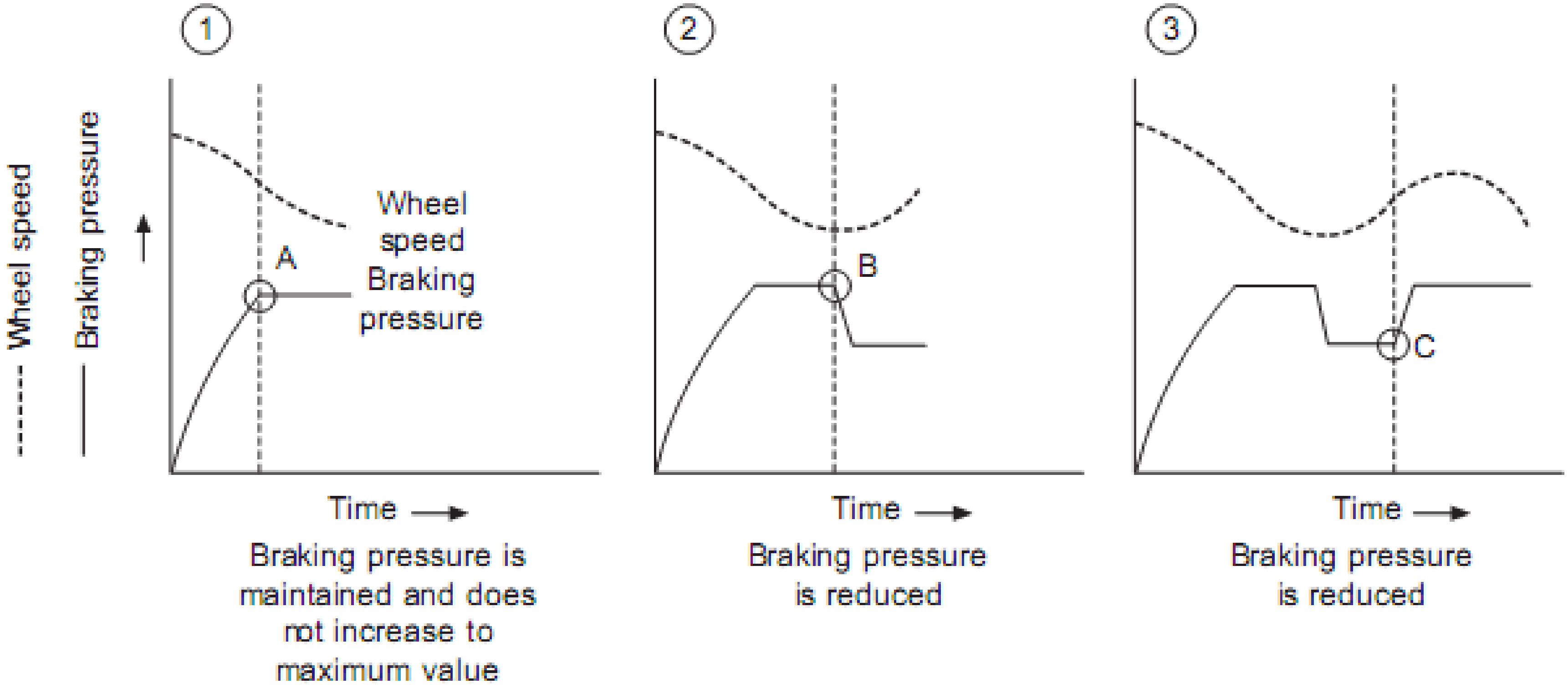
Variation of the coefficient of friction (μ) with slip ratio

- Induction type wheel-speed sensors on the wheel assembly or differential
- couple magnetically to a toothed wheel known as an impulse ring.



Antiskid braking system (ABS)

- All electronic signals come to the electronic controller (ECU)
- The ECU controls the hydraulic modulator
  - To control the Brake line pressure in Brake master cylinder



Wheel-speed and braking pressure during ABS-controlled braking

- **If wheel decelerates beyond a certain level, curtail brake pressure (1)**
- **If wheel decelerates further, reduce brake pressure further (2)**
- **If wheel accelerates, increase brake pressure (3)**

Traction control systems

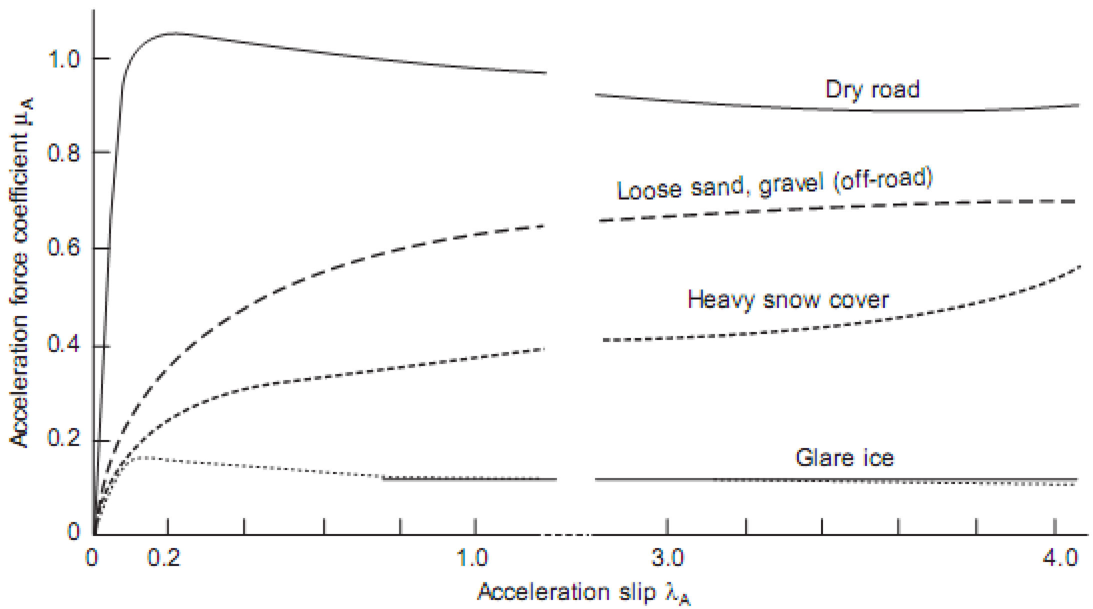
- prevent drive wheels from wheelspinning during starting or
  - accelerating on a wet or icy surface.
- avoid reduction of either steering response in front-wheel-drive (FWD) / vehicle stability on rear-wheel-drive (RWD) vehicles.

TCS operates

- to maximize adhesion to the road surface during acceleration
- Same sensors as in ABS
- The actuation uses fuel, ignition and driven wheel braking action

Traction Control Systems (TCS)

- to achieve reduction in driven wheel torque during wheelspin.
- maintain the acceleration slip of the driven wheels equal to the mean rotational velocity of the non-driven wheels + a specified speed difference known as the slip threshold.
- driven wheels are kept at a faster speed than the non-driven wheels
- the vehicle accelerates at a constant rate proportional to the difference in the two speeds. (if difference is not in limits (slip threshold), traction needs to be controlled)
- Control depends on road surface conditions or adhesion coefficient.
- on dry road surfaces, maximum acceleration at slip rates of 10 to 30%.
- On glare ice, maximum traction between 2 and 5 percent
- so TCS systems designed for a slip rate range between 2 and 20%.



Adhesion force coefficient μ_A as a function of acceleration λ_A(Jurgen, 1995)

- on loose sand or gravel and in deep snow the coefficient of adhesion increases continually with the slip rate
- TCS systems incorporate slip-threshold switches to allow the driver to select a higher slip threshold or switch off the TCS
- The control objectives of TCS are modified by vehicle speed and curve recognition.
- Both of these variables can be derived from the speeds of the non-driven wheels.
- coefficient of adhesion or friction decided on the basis of acceleration rate and engine torque
- The slip threshold is raised in response to higher friction coefficients to allow higher acceleration rates

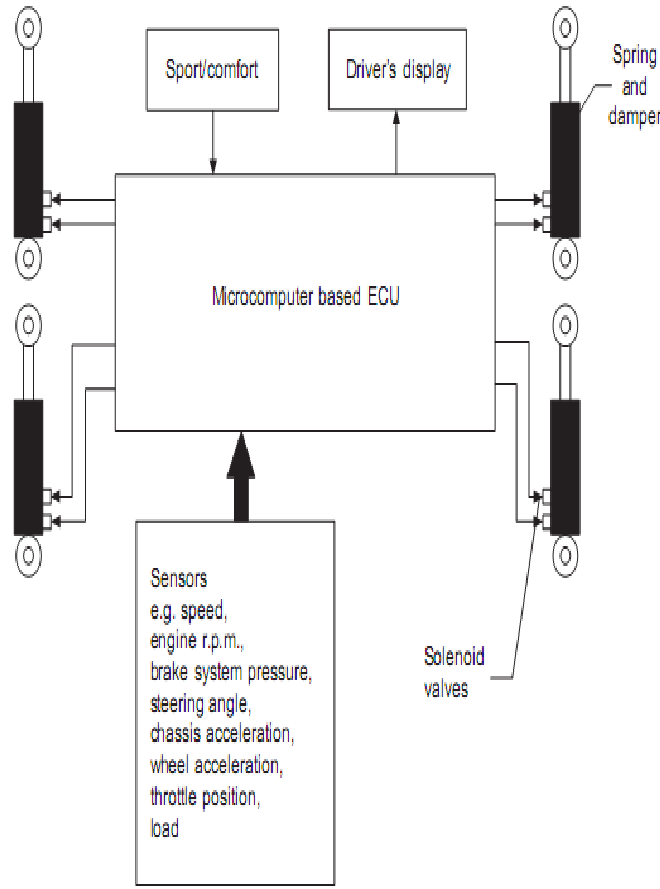
- Curve recognition or cornering detection also affects the control strategy for TCS.
- This strategy employs the difference in wheel speeds of the non-driven wheel speeds as a basis for reductions in the slip setpoint to enhance stability in curves.
- High vehicle speeds and low acceleration requirements on low coefficient of adhesion surfaces imply
  - a control strategy of progressively lower slip threshold setpoints as the vehicle speed increases,
  - gives maximum lateral adhesion on the surface.

Electronic damping control

- The primary function of a shock absorber
  - control vehicle movement against roll during turning and pitch during acceleration or braking.
  - Requires hard suspension
- secondary role
  - To prevent vehicle vibration caused by a poor road surface.
  - Requires a soft suspension
- Electronic damping control (EDC) used to attain these twin objectives
- altering the characteristics of spring and oil-filled damper arrangement
  - difficult and expensive
- Simple option - Suspensions with at least three settings; ‘soft’, ’medium’ and ‘firm’
- OR electronically controlled suspension systems using air, nitrogen gas and hydraulic oil as a suspension agent.

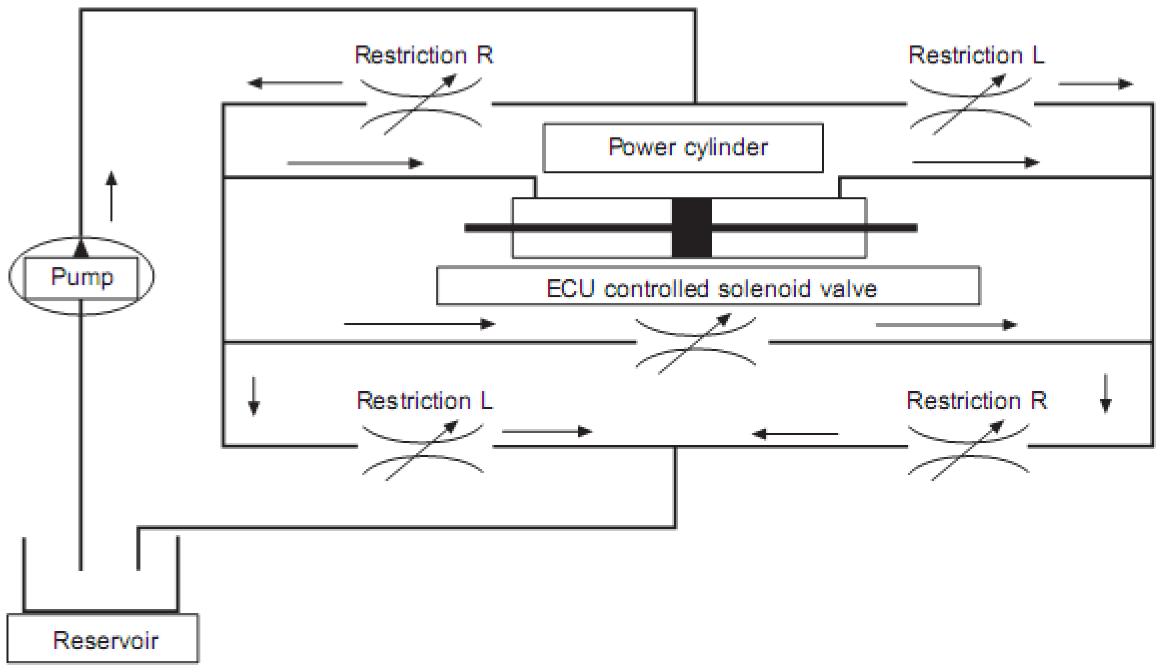
- sensors used
  - vehicle speed,
  - engine r.p.m.,
  - brake system pressure,
  - steering angle,
  - chassis and wheel acceleration,
  - throttle position,
  - vehicle load and
  - even road surface condition
- Road condition - implied by processing signals from front and rear height sensors rather than direct measurement.
  - if the height sensor signals a small high frequency but a large low frequency amplitude
    - a heaving or undulating road surface
    - Does not require a softening of damper.
  - A large high frequency component would suggest
    - a rough road surface and
    - Softening of damper action.
    - Conflicts with damper requirement to prevent rolling during cornering.
    - If the vehicle corners on a rough surface this must be resolved by the ECU.

- Longitudinal acceleration
  - measured directly using an acceleration sensor, or
  - inferred from brake system pressure and throttle opening angle.
  - Used to control pitching during acceleration / braking
- lateral forces
  - inferred by the rate at which the steering wheel is being turned and the vehicle speed.
  - used by the ECU to prevent rolling.
- the actuators are dampers fitted with two ON-OFF fluid control solenoids used to select one of four different damper settings (normal, soft, super-soft and firm).
- Driver can choose sport or smooth ride mode.
  - In sport mode soft or super-soft damper settings excluded
  - Result in a harder but more stable ride.



Electronically controlled damping system

Electronically controlled power-assisted steering (PAS)



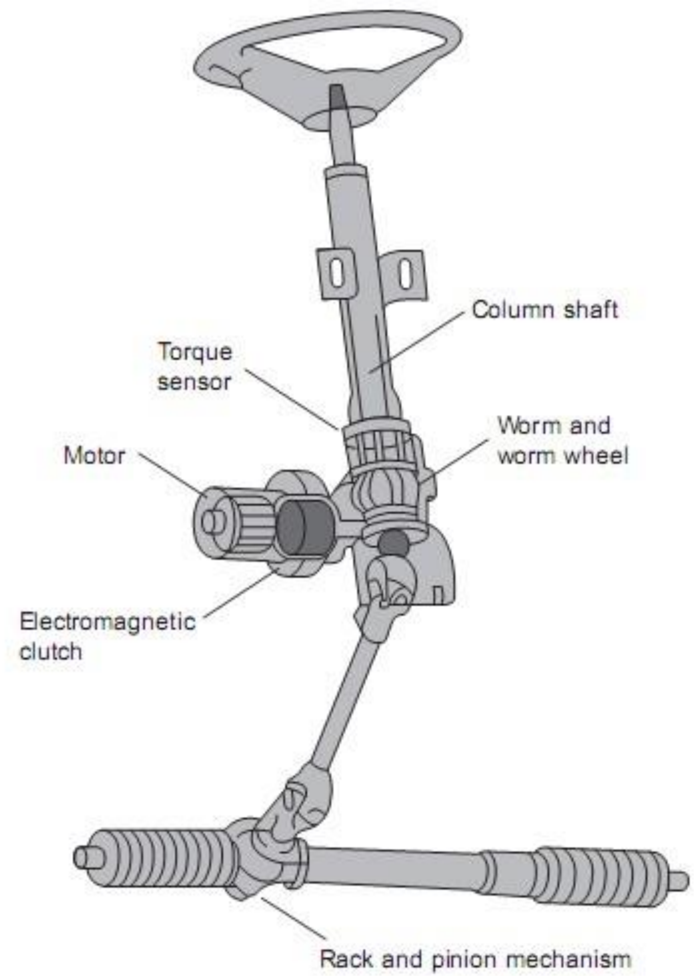
Hydraulic bridge circuit for electronically-controlled power steering showing flow paths

Electronically controlled hydraulic PAS

- the ports of a solenoid valve are connected across the rack and pinion steering hydraulic power cylinder.
- with increasing vehicle speed the valve opening is extended
  - reducing the hydraulic pressure in the power cylinder
  - increasing the steering effort.
- bridge-like restrictions for control of the power cylinder are formed by the paths through the pump to port connections of a rotary valve
- The valve is connected directly to the steering wheel and
  - a small movement of this controls the high pressure hydraulic fluid to reach the power cylinder/solenoid valve.

Electric PAS

- input to the rack and pinion steering system is from a motor/reduction gearbo
- motor torque is applied directly to either the pinion gear shaft or to the rack shaft.
- The steering effort range is greater than with hydraulic systems,
- installations are cheaper and reliable.
- Power is only consumed when steering wheel moves, (unlike hydraulic system)
- a torque sensor on the column shaft
- The electric motor coupled to the worm wheel mechanism through a reduction gearbox.
- The load torque T_L on the steering colum is the load presented by the worm mechanism and the rack and pinion assembly to which it is attached.



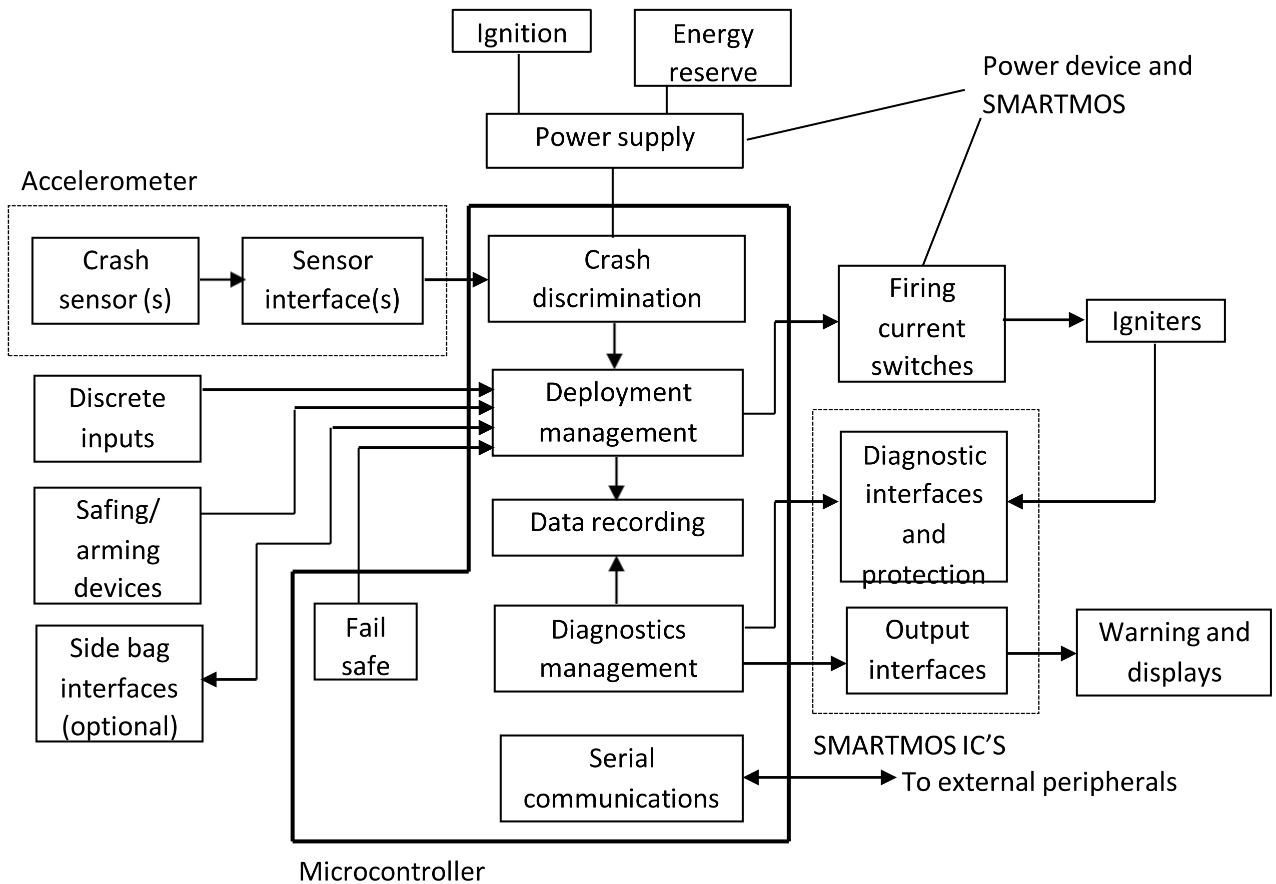
- The amount of motor torque is proportional to the motor current I_M.
- in a simple armature controlled d.c. motor the average current is given

$$I_M = \frac{V_M - k \times N}{R}$$

  - where *R is the armature resistance,*
  - *N is the speed of the motor and* V_M the motor voltage,
- the set point motor voltage depends on how much control effort is required from the d.c. motor.
- When a driver turning a steering wheel at a constant rate, say in cornering.
  - The d.c. motor, must turn at a speed proportional to this rate.
  - Controlling term --- motor voltage = k x N
- at high vehicle speeds the assistance given to the driver must decrease in proportion to speed
  - i.e., Decrease motor current or voltage as vehicle speed increases.
  - Motor voltage component = k_T x T_m (T_m is output from Torque sensor)
  - Inverse function of vehicle speed
- Add both components to get appropriate control

Air-bag and seat belt pre-tensioner systems

- systems consist of
  - crash detection sensors (typically piezoelectric) with a signal conditioning amplifier
  - a microcontroller distinguishing between crashes and normal vehicle dynamics,
  - igniter triggering for the pyrotechnic inflator
    - used for air-bag deployment and seat belt tightening.
- The allowable forward passenger travel with an air-bag system is 12.5 cm
  - with seat belt tensioning systems it is about 1 cm.
- Approximately 30 ms are required to inflate air-bags and
- time required to tension a seat belt with a retractor = ~10 ms.
- triggering must be done by the time forward displacement is reached minus the activation time of the respective restraining device.
- Often multiple sensors and sensor mounting positions
- When airbag is triggered
  - ECU turns on the firing current switches,
  - allows current through the igniter,
  - initiates a gas generation reaction inside the inflation module.
  - Capacitance based power maintained even if battery is disconnected



Air-bag electronics block diagram

UNIT-4

Engine Management and Types of Engine Sensors

Engine sensors in a vehicle are incorporated to provide the correct amount of fuel for all operating conditions. A large number of input sensors are monitored by the engine control unit. Today, sensor technology has become common in modern vehicles. Sensors enhance safety of the people - both on board and on road, control vehicle emissions and make vehicles more efficient. In this article, we will discuss different types of engine sensors used in modern vehicles.

Mass Air Flow Sensor (MAF)

The MAF sensor (electric sensor) is an integral part of the engine system. It is controlled by a computer. It is located in a plastic covering between the engine and the air filter. The purpose of MAF is to calculate the amount of air intake by the engine, in terms of volume and density. For measuring the volume and density of air, the sensor uses either a hot wire or a heated filament. After the measurement, it sends a voltage signal to the computer. With this, the computer can calculate the right amount of fuel needed to maintain the correct fuel mixture for every operating condition. If there is any fault in the MAF sensor, it may result in rough idle, stalling and poor fuel economy.

Throttle Position Sensor (TPS)

The Throttle Position Sensor (TPS) is a variable resistor attached or mounted on the throttle body and is operated by moving along with the throttle shaft or spindle. The TPS changes the resistances as the throttle opens and closes, and sends a voltage signal to the computer showing the angle or position of the throttle. Thus, the TPS causes the Electronic Control Unit (ECU) to use the data to measure the engine load, fuel delivery adjust timing, acceleration, deceleration when the engine is idle or in wide open throttle, and then makes the changes according to the operating conditions. Fuel rate is either increased ordecreased to achieve this.

Coolant Temperature Sensor (CTS)

The Coolant Temperature Sensor (CTS) is a temperature dependent variable resistor located on the cylinder head or intake manifold. The CTS is an important sensor and the operating strategy of the enginedepends on the signal it sends. So, it is called the “master” sensor.

The CTS measures the internal temperature of the engine coolant. It also senses the changes in temperature and sends a voltage signal to the Power train Control Module (PCM) for determining whetherthe engine is cold or warming up, is at normal operating temperature or is overheating.

Oxygen Sensor

The oxygen sensor is located on the exhaust manifold. This sensor monitors the amount of unburned oxygen present in the exhaust. When the fuel mixture is rich, most of the oxygen is exhausted during the combustion. So, only a little unburned oxygen will be left out in the exhaust. Difference in the oxygen levels creates an electrical potential, which causes the sensor to generate a voltage signal. This helps the ECU to check the quality of fuel mixture to make the changes accordingly. The sensor output will be high if the fuel mixture is rich, and the sensor output will be low if the fuel mixture is lean.

Manifold Absolute Pressure Sensor (MAP)

The MAP is a key sensor as it senses the engine load. It is mounted on the intake manifold. It monitors the difference between the air pressure in the intake manifold and outside. This sensor responds to the vacuum in the intake manifold and generates a voltage signal accordingly. It then sends the signal to the PCM. The input of the sensor is used for adjusting the fuel mixture and ignition timing, according to thechanges.

Engine Speed Sensor (ESS)

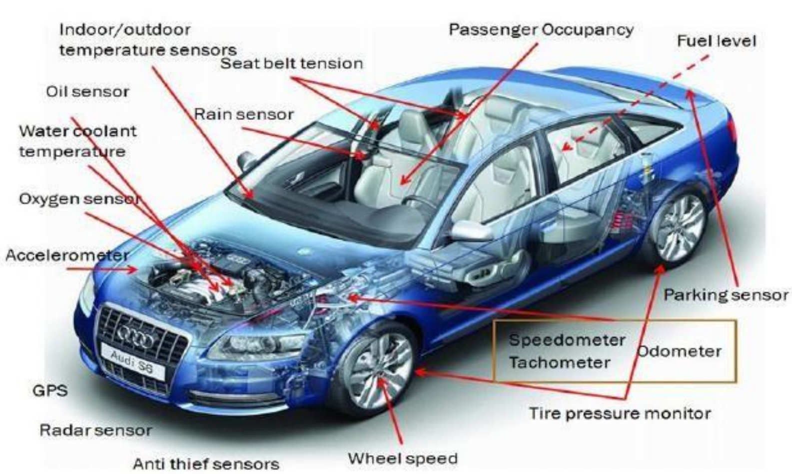
The ESS is a sensor attached to the crankshaft of the car's engine. It is different from vehicle speed sensor. The ESS is used for monitoring the engine speed. In other words, it is meant for assessing thespeed at which the crankshaft spins.

Voltage Sensor

The voltage sensor monitors the system voltage of the vehicle and reports it to PCM so that it can rise theidle speed of the vehicle, if the voltage is dropping.

Engine sensors are an important technological innovation. They lead to better performance, better qualityand more years of driving experience.

- **Sensors:**
- Sensors are the components of the system that provide the inputsthat enable the computer **(ECM)** to carry out the operations that make the system function correctly.
- In the case of vehicle sensors it is usually a voltage that isrepresented by a code at the computer's processor. If this voltage is incorrect the processor will probably take it as an invalid input and record a fault.



Types of Sensors used

Mass air flow (MAF) rate

Exhaust gas oxygen concentration (possibly heated)

Throttle plate angular position

Crankshaft angular position/RPM

Coolant temperature

Intake air temperature

Manifold absolute pressure (MAP)

Differential exhaust gas pressure

Vehicle speed

Transmission gear selector position

Mass air flow (MAF) rate sensor

- Airflow sensors are used on engines with multiport electronic fuel injection. This is because the amount of fuel delivered by an EFI system is controlled by a computer(powertrain control module or PCM) which turns the fuel injectors on and off.

- The airflow sensor keeps the computer informed about how much air is being pulled into the engine past the throttle plates. This input along with information from other engine sensors allows the computer to calculate how much fuel is needed.

- The computer then increases or decreases injector duration (on time) to provide the correct air/fuel ratio.

- Types: 1. Vane type Air Flow Rate Sensor
  2. Hot Wire type Air Flow Rate Sensor

- **Vane type air flow (MAF) rate sensor**
- An engine requires the correct air–fuel ratio to suit various conditions. With electronic fuel injection the ECM controls the air–fuel ratio and in order to do this it needs a constant flow of information about the amount of air flowing to the engine.
- With this information, and data stored in its memory, the ECM can then send out a signal to the injectors, so that they provide the correct amount of fuel.
- Air flow measurement is commonly performed by a 'flap'-type air flow sensor. The air flow sensor shown in Figure A. uses the principle of the potential divider (potentiometer).

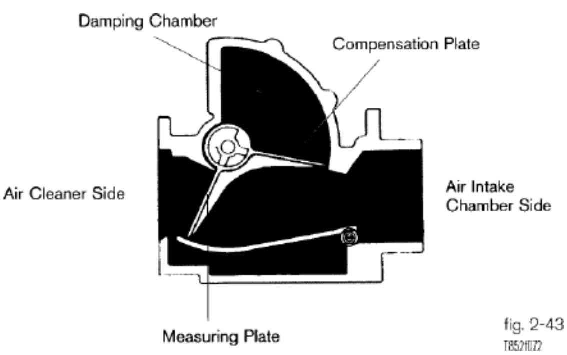
The theoretical form of a simple potential divider. A voltage, say 5 V, is applied across terminals *A* and *B*. *C* is a slider which is in contact with the resistor and a voltmeter is connected between *A* and

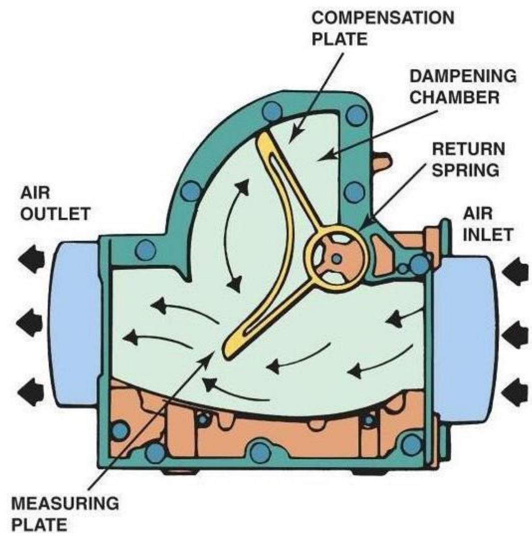
C. The voltage V_{AC} is related to the position of the slider

C in the form $V_{AC} = V_{AB} * x/l$.

- In the air flow sensor, the moving probe (wiper) of the potential divider is linked to the pivot of the measuring flap so that angular displacement of the measuring flapis registered as a known voltage at the potentiometer.

- Figure shows a simplified form of the air flow sensor. The closed position of the measuring flap will give a voltage ofapproximately zero, and when fully open the voltage will be 5 V.
- Intermediate positions will give voltages between these values. In practice, it is not quite as simple as this, because allowance must be made for other contingencies.
- A vane airflow sensor is located ahead of the throttle and monitors the volume of air entering the engine by means of a spring- loaded mechanical flap. The flap is pushed open by an **amount that is proportional to the volume of air entering the engine.**





- The flap has a wiper arm that rotates against a sealed potentiometer (variable resistor or rheostat), allowing the sensor's resistance and output voltage to change according toairflow.
- The greater the airflow, the further the flap is forced open. This lowers the potentiometer's resistance and increases the voltage return signal to the computer.
- A compensation plate acts as a shock absorber to prevent rapid movement or vibrations of the measuring plate.
- A sealed idle mixture screw is also located on the airflow sensor. This controls the amount of air that bypasses the flap,and consequently the richness or leanness of the fuel mixture.

VANE AIRFLOW SENSOR PROBLEMS

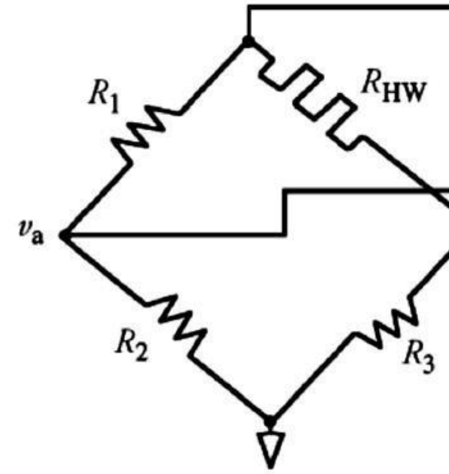
- Vane airflow sensors as well as all the other types of airflow sensors **can't tolerate air leaks.** Air leaks downstream ofthe sensor can allow "unmetered" or "false" air to enter the engine. The extra air can lean out the fuel mixture causing a variety of driveability problems, including lean misfire, hesitation and stumbling when accelerating, and a roughidle.
- **Dirt can also cause problems**. Unfiltered air passing through a torn or poor fitting air filter can allow dirt to build up on the flap shaft of a vane airflow sensor causing the flap to bind or stick. The operation of the flap can be tested by gently pushing it open with a finger. It should open and close smoothly with even resistance. If it binds or sticks, a shot of carburetor cleaner may loosen it up otherwise the sensor willhave to be replaced.

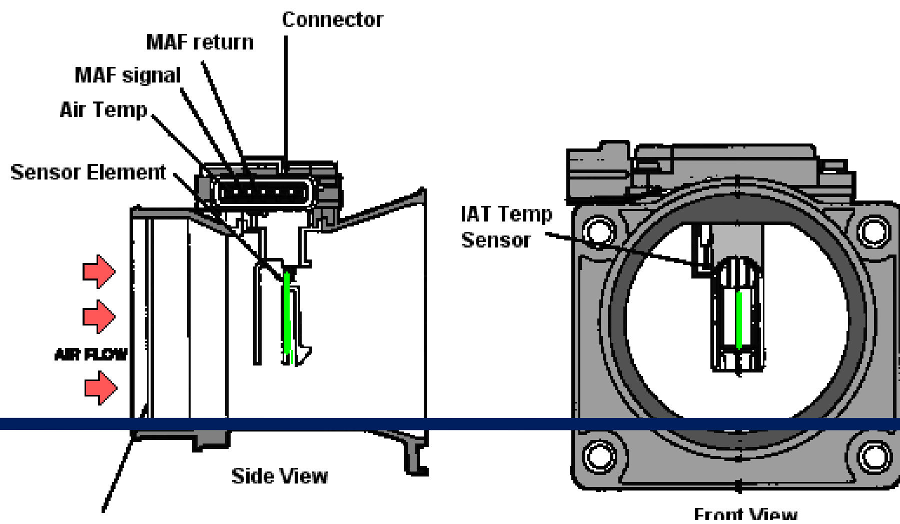
- **Backfiring in the intake manifold can force the flap backwards violently, often bending or breaking the flap. Some sensors have a "backfire" valve built into theflap to protect the flap in case of a backfire by venting the explosion. But the anti-backfire valve itself can become a source of trouble if it leaks. A leaky backfire valve will cause the sensor to read low and the engine torun rich.**

❖ Hot Wire Type MAF:

- The hot wire MAF sensor is a variation of a classic air flow sensor that was known as a hot wire anemometer and was used, for example, to measure wind velocity for weather forecasting.
- In this MAF, the hot-wire, or sensing, element is replaced by a hot-film structure mounted on a substrate.
- On the air inlet side is mounted a honeycomb flow straightener that "smooths" the air flow (causing nominally laminar air flow over the film element). At the lower portion of the structure is the signal processing circuitry.

- The film element is electrically heated to a constanttemperature above that of the inlet air.
- The hot-film element is incorporated in a Wheatstone bridgecircuit (Figure at top).
- The power supply for the bridge circuit comes from anamplifier





- The Wheatstone bridge consists of three fixed resistors R_1, R_2, and R_3 and a hot-film element having resistance R_{HW}. With no air flow the resistors R_1, R_2, and R_3 are chosen such that voltage v_a and v_b are equal (i.e., the bridge is said to be balanced).
- As air flows across the hot film, heat is carried away from the film by the moving air. The amount of heat carried away varies in **proportion to the mass flow rate of the air.**
- The heat lost by the film to the air tends to cause the resistance of the film to vary, which unbalances the bridge circuit, thereby producing an input voltage to the amplifier.
- The output of the amplifier is connected to the bridge circuit and provides the power for this circuit. The amplified voltage changes the resistance in such a way as to maintain afixed hot-film temperature relative to the inlet temperature.

- **The amplifier output voltage v_c varies with MAF and serves as a measure of R_m. Typically the conversion of MAF to voltage is slightly nonlinear, as indicated by the calibration curve depicted in Figure.**
- Fortunately, a modern digital engine controller can convert the analog bridge output voltage directly to mass air flow by simple computation.

- Hot Wire MAF sensors have no moving parts. Unlike a vane airflow meter that uses a spring-loaded flap, mass airflow sensors use electrical current to measure airflow.
- The sensing element, which is either a **platinum wire (hot wire) or nickel foil grid** (hot film), is heated electrically to keep it a certain number of degrees hotter than the incoming air.
- In the case of **hot film MAFs, the grid is heated to 75^{o}C** above incoming ambient air temperature. With the **hot wire sensors, the wire is heated to 100 ^{o}C** above ambient temperature.
- As air flows past the sensing element, it cools the element and increases the current needed to keep the element hot. Because the cooling effect varies directly with the temperature, density and humidity of the incoming air, the amount of current needed to keep the element hot is **directly proportional to the air "mass" entering the engine**

Positions Sensors:

The positions sensors are generally speed sensors of different working principle used for detecting the position of different parameters.

Parameters Measured,

- Crankshaft Position Sensor
- Camshaft Position Sensor
- ABS Wheel Sensors
- Vehicle Speed Sensor

Working Principles used,

- Magnetic Reluctance (Variable Reluctance) type
- Hall Effect type
- Optical Type

- Crankshaft Position Sensor:
- A crank position sensor is a component used in an internal combustion engine to monitor the position or rotational speed of the crankshaft. This information is used by engine management systems to control ignition system timing and other engine parameters**.**

Magnetic Reluctance (or Variable Reluctance)type sensor

- This type of sensor is used in many vehicle applications, such as ignition systems, engine speed sensors for fuelling, and wheel speedsensors for anti-lock braking etc.
- Air has a greater reluctance (resistance to magnetism) than iron and this fact is made use of in many sensors. The basic principle of operation of a variable reluctance type sensor (Fig. 5.1) may be understood from the following description.
- The principal elements of the sensor are:
  - an iron rotor with lobes on it;
  - a permanent magnet;
  - a metallic path (the pole piece) for carrying the magnetic flux;
  - a coil, wound around the metallic path, in which a voltageis induced.

- The reluctor disc has a number of tabs on it and these tabsare made to move through the air gap in the magnetic circuit.

- The movement of the reluctor tabs, through the air gap is achieved by rotation of the reluctor shaft. The voltage induced in the sensor coil is related to the rate of change of magnetic flux in the magnetic circuit.

- The faster the rate of change of magnetic flux the larger will be the voltage that is generated in the sensor coil. When the metal tab on the reluctor rotor is outside the air gap, the sensor voltage is zero.

- As the tab moves into the air gap the flow of magnetism (flux) increases rapidly. This causes the sensor voltage to increase, quite quickly, to a maximum positive value. Figure
  5.2 shows the approximate behavior of the voltage output as the reluctor is rotated.

- **Figure 5.2(a) shows the reluctor tab moving into the air gap. As the metal tab moves further into the gap the voltage begins to fall and, when the metal tab is exactly aligned with the pole piece, the sensor voltage falls back to zero (Although the magnetic flux is strongest at this point, it is not changing and this means that the voltage is zero.)**

- **As the metal tab continues to rotate out of the air gap and away from the pole piece, the rate of change of the magnetic flux is rapid, but opposite in direction to when the tab was moving into the air gap.**

- This results in the negative half of the voltage waveform as shown in Fig. 5.2(c). When the tab has moved out of the air gap the sensor voltage returns to zero. While the rotor shaft continues to turn another tab will enter the air gap and the above process will be repeated.

- **If the sensor coil is connected to an oscilloscope the pattern observed will be similar to that shown in Fig. 5.2(d).**

Unit 5

Geographical Information Systems (GIS)

Introduction

Geographical Information System (GIS) is a technology that provides the means to collect and use geographic data to assist in the development of Agriculture. A digital map is generally of much greater value than the same map printed on a paper as the digital version can be combined with other sources of data for analyzing information with a graphical presentation. The GIS software makes it possible to synthesize large amounts of different data, combining different layers of information to manage and retrieve the data in a more useful manner. GIS provides a powerful means for agricultural scientists to better service to the farmers and farming community in answering their query and helping in a better decision making to implement planning activities for the development of agriculture.

Overview of GIS

A Geographical Information System (GIS) is a system for capturing, storing, analyzing and managing data and associated attributes, which are spatially referenced to the Earth. The geographical information system is also called as a geographic information system or geospatial information system. It is an information system capable of integrating, storing, editing, analyzing, sharing, and displaying geographically referenced information. In a more generic sense, GIS is a software tool that allows users to create interactive queries, analyze the spatial information, edit data, maps, and present the results of all these operations. GIS technology is becoming essential tool to combine various maps and remote sensing information to generate various models, which are used in real time environment. Geographical information system is the science utilizing the geographic concepts, applications and systems.

Geographical Information System can be used for scientific investigations, resource management, asset management, environmental impact assessment, urban planning, cartography, criminology, history, sales, marketing, and logistics. For example, agricultural planners might use geographical data to decide on the best locations for a location specific crop planning, by combining data on soils, topography, and rainfall to determine the size and location of biologically suitable areas. The final output could include overlays with land ownership, transport, infrastructure, labour availability, and distance to market centers.

History of GIS development

The idea of portraying different layers of data on a series of base maps, and relating things geographically, has been around much older than computers invention. Thousands years ago, the early man used to draw pictures of the animals they hunted on the walls of caves. These animal drawings are track lines and tallies thought to depict migration routes. While simplistic in comparison to modern technologies, these early records mimic the two-element structure of modern geographic information systems, an image associated with attribute information.

Possibly the earliest use of the geographic method, in 1854 John Snow depicted a cholera outbreak in London using points to represent the locations of some individual cases. His study of the distribution of cholera led to the source of the disease, a contaminated water pump within the heart of the cholera outbreak. While the basic elements of topology and theme existed previously in cartography, the John Snow map was unique, using cartographic methods, not only to depict but also to analyze, clusters of geographically dependent phenomena for the first time.

The early 20th century saw the development of "photo lithography" where maps were separated into layers. Computer hardware development spurred by nuclear weapon research led to general-purpose computer "mapping" applications by the early 1960s. In the year 1962, the world's first true operational GIS was developed by the federal Department of Forestry and Rural Development in Ottawa, Canada by Dr. Roger Tomlinson. It was called the "Canada Geographic Information System" (CGIS) and was used to store, analyze, and manipulate data collected for the Canada Land Inventory (CLI). It is an initiative to determine the land capability for rural Canada by mapping information about soils, agriculture, recreation, wildlife, forestry, and land use at a scale of 1:50,000.

CGIS was the world's first "system" and was an improvement over "mapping" applications as it provided capabilities for overlay, measurement, and digitizing or scanning. It supported a national coordinate system that spanned the continent, coded lines as "arcs" having a true embedded topology, and it stored the attribute and location specific information in a separate files. Dr. Tomlinson is known as the "father of GIS," for his use of overlays in promoting the spatial analysis of convergent geographic data.

In 1964, Howard T Fisher formed the Laboratory for Computer Graphics and Spatial Analysis at the Harvard Graduate School of Design, where a number of important theoretical concepts in spatial data handling were developed. This lab had major influence on the development of GIS until early 1980s. Many pioneers of newer GIS "grew up" at the Harvard lab and had distributed seminal software code and systems, such as 'SYMAP', 'GRID', and 'ODYSSEY'.

By the early 1980s, M&S Computing (later Intergraph), Environmental Systems Research Institute (ESRI) and CARIS emerged as commercial vendors of GIS software, successfully incorporating many of the CGIS features, combining the first generation approach to separation of spatial and attribute information with a second generation approach to organizing attribute data into database structures. More functions for user interaction were developed mainly in a graphical way by a user friendly interface (Graphical User Interface), which gave to the user the ability to sort, select, extract, reclassify, reproject and display data on the basis of complex geographical, topological and statistical criteria. During the same time, the development of a public domain GIS begun by the U.S. Army Corp of Engineering Research Laboratory (USA-CERL) in Champaign, Illinois, a branch of the U.S. Army Corps of Engineers to meet the need of the United States military for software for land management and environmental planning.

In the years 1980s and 1990s industry growth were spurred on by the growing use of GIS on Unix workstations and the personal computers. By the end of the 20th century, the rapid growth in various systems had been consolidated and standardized on relatively few platforms and users were beginning to export the concept of viewing GIS data over the Internet, requiring uniform data format and transfer standards. More recently, there is a growing number of free, open source GIS packages, which run on a range of operating systems and can be customized to perform specific tasks. As computing power increased and hardware prices slashed down, the GIS became a viable technology for state development planning. It has become a real Management Information System (MIS), and thus able to support decision making processes.

Components of GIS

GIS enables the user to input, manage, manipulate, analyze, and display geographically referenced data using a computerized system. To perform various operations with GIS, the components of GIS such as software, hardware, data, people and methods are essential.

Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are (a) a database management system (DBMS) (b) tools for the input and manipulation of geographic information (c) tools that support geographic query, analysis, and visualization (d) a graphical user interface (GUI) for easy access to tools. GIS software are either commercial software or software developed on Open Source domain, which are available for free. However, the commercial software is copyright protected, can be expensive and is available in terms number of licensees.

Currently available commercial GIS software includes Arc/Info, Intergraph, MapInfo, Gram++ etc. Out of these Arc/Info is the most popular software package. And, the open source software are AMS/MARS etc.

Hardware

Hardware is the computer on which a GIS operates. Today, GIS runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations. Minimum configuration required to Arc/Info Desktop 9.0 GIS application is as follows:

Product: ArcInfo Desktop 9.0

Platform: PC-Intel

Operating System: Windows XP Professional Edition, Home Edition

Service Packs/Patches: SP 1

SP2 (refer to Limitations)

Shipping/Release Date: May 10, 2004

Hardware Requirements

CPU Speed: 800 MHz minimum, 1.0 GHz recommended or higher

Processor: Pentium or higher

Memory/RAM: 256 MB minumum, 512 MB recommended or higher

Display Properties: Greater than 256 color depth

Swap Space: 300 MB minimum

Disk Space: Typical 605 MB NTFS, Complete 695 MB FAT32 + 50 MB for installation

Browser: Internet Explorer 6.0 Requirement:

(Some features of ArcInfo Desktop 9.0 require a minimum installation of Microsoft Internet Explorer Version 6.0.)

Data

The most important component of a GIS is the data. Geographic data or Spatial data and related tabular data can be collected in-house or bought from a commercial data provider. Spatial data can be in the form of a map/remotely-sensed data such as satellite imagery and aerial photography. These data forms must be properly geo-referenced (latitude/longitude). Tabular data can be in the form attribute data that is in some way related to spatial data. Most GIS software comes with inbuilt Database Management Systems (DBMS) to create and maintain a database to help organize and manage data.

Users

GIS technology is of limited value without the users who manage the system and to develop plans for applying it. GIS users range from technical specialists who design and maintain the system to those who use it to help them do their everyday work.

These users are largely interested in the results of the analyses and may have no interest or knowledge of the methods of analysis. The user-friendly interface of the GIS software allows the nontechnical users to have easy access to GIS analytical capabilities without needing to know detailed software commands. A simple User Interface (UI) can consist of menus and pull-down graphic windows so that the user can perform required analysis with a few key presses without needing to learn specific commands in detail.

Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

Functions of GIS

General-purpose GIS software performs six major tasks such as input, manipulation, management, query and analysis, Visualization.

Input

The important input data for any GIS is digitized maps, images, spatial data and tabular data. The tabular data is generally typed on a computer using relational database management system software. Before geographic data can be used in a GIS it must be converted into a suitable digital format. The DBMS system can generate various objects such as index generation on data items, to speed up the information retrieval by a query. Maps can be digitized using a vector format in which the actual map points, lines, and polygons are stored as coordinates. Data can also be input in a raster format in which data elements are stored as cells in a grid structure (the technology details are covered in following section).

The process of converting data from paper maps into computer files is called digitizing. Modern GIS technology has the capability to automate this process fully for large projects; smaller jobs may require some manual digitizing. The digitizing process is labour intensive and time-consuming, so it is better to use the data that already exist.

Today many types of geographic data already exist in GIS-compatible formats. These data can be obtained from data suppliers and loaded directly into a GIS.

Manipulation

GIS can store, maintain, distribute and update spatial data associated text data. The spatial data must be referenced to a geographic coordinate systems (latitude/longitude). The tabular data associated with spatial data can be manipulated with help of data base management software. It is likely that data types required for a particular GIS project will need to be transformed or manipulated in some way to make them compatible with the system. For example, geographic information is available at different scales (scale of 1:100,000; 1:10,000; and 1:50,000). Before these can be overlaid and integrated they must be transformed to the same scale. This could be a temporary transformation for display purposes or a permanent one required for analysis. And, there are many other types of data manipulation that are routinely performed in GIS. These include projection changes, data aggregation, generalization and weeding out unnecessary data.

Management

For small GIS projects it may be sufficient to store geographic information as computer files. However, when data volumes become large and the number of users of the data becomes more than a few, it is advised to use a database management system (DBMS) to help store, organize, and manage data. A DBMS is a database management software package to manage the integrated collection of database objects such as tables, indexes, query, and other procedures in a database.

There are many different models of DBMS, but for GIS use, the relational model database management systems will be highly helpful. In the relational model, data are stored conceptually as a collection of tables and each table will have the data attributes related to a common entity. Common fields in different tables are used to link them together with relations. Because of its simple architecture, the relational DBMS software has been used so widely. These are flexible in nature and have been very wide deployed in applications both within and without GIS.

Query

The stored information either spatial data or associated tabular data can be retrieved with the help of Structured Query Language (SQL). Depending on the type of user interface, data can be queried using the SQL or a menu driven system can be used to retrieve map data. For example, you can begin to ask questions such as:

- Where are all the soils are suitable for sunflower crop?
- What is the dominant soil type for Paddy?
- What is the groundwater available position in a village/block/district?

Both simple and sophisticated queries utilizing more than one data layer can provide timely information to officers, analysts to have overall knowledge about situation and can take a more informed decision.

Analysis

GIS systems really come into their own when they are used to analyze geographic data. The processes of geographic analysis often called spatial analysis or geo-processing uses the geographic properties of features to look for patterns and trends, and to undertake "what if" scenarios. Modern GIS have many powerful analytical tools to analyse the data. The following are some of the analysis which are generally performed on geographic data.

A. Overlay Analysis

The integration of different data layers involves a process called overlay. At its simplest, this could be a visual operation, but analytical operations require one or more data layers to be joined physically. This overlay, or spatial join, can integrate data on soils, slope, and vegetation, or land ownership. For example, data layers for soil and land use can be combined resulting in a new map which contains both soil and land use information. This will be helpful to understand the different behaviour of the situation on different parameters.

B. Proximity Analysis

GIS software can also support buffer generation that involves the creation of new polygons from points, lines, and polygon features stored in the database. For example, to know answer to questions like; How much area covered within 1 km of water canal? What is area covered under different crops? And, for watershed projects, where is the boundary or delineation of watershed, slope, water channels, different types water harvesting structures are required, etc.

Visualization

GIS can provide hardcopy maps, statistical summaries, modeling solutions and graphical display of maps for both spatial and tabular data. For many types of geographic operation the end result is best visualized as a map or graph. Maps are very efficient at storing and communicating geographic information. GIS provides new and exciting tools to extend the art of visualization of output information to the users.

Technology used in GIS

Data creation

Modern GIS technologies use digital information, for which various digitized data creation methods are used. The most common method of data creation is digitization, where a hard copy map or survey plan is transferred into a digital medium through the use of a computer-aided design program with geo-referencing capabilities. With the wide availability of rectified imagery (both from satellite and aerial sources), heads-up digitizing is becoming the main avenue through which geographic data is extracted. Heads-up digitizing involves the tracing of geographic data directly on top of the aerial imagery instead of through the traditional method of tracing the geographic form on a separate digitizing tablet.

Relating information from different sources

If you could relate information about the rainfall of a state to aerial photographs of county, you might be able to tell which wetlands dry up at certain times of the year. A

GIS, which can use information from many different sources in many different forms, can help with such analyses. The primary requirement for the source data consists of knowing the locations for the variables. Location may be annotated by x, y, and z coordinates of longitude, latitude, and elevation, or by other geocode systems like postal codes. Any variable that can be located spatially can be fed into a GIS. Different kinds of data in map form can be entered into a GIS.

A GIS can also convert existing digital information, which may not yet be in map form, into forms it can recognize and use. For example, digital satellite images generated through remote sensing can be analyzed to produce a map-like layer of digital information about vegetative covers. Likewise, census or hydrologic tabular data can be converted to map-like form, serving as layers of thematic information in a GIS.

Data representation

GIS data represents real world objects such as roads, land use, elevation with digital data. Real world objects can be divided into two abstractions: discrete objects (a house) and continuous fields (rain fall amount or elevation). There are two broad methods used to store data in a GIS for both abstractions: Raster and Vector.

Raster

A raster data type is, in essence, any type of digital image. Anyone who is familiar with digital photography will recognize the pixel as the smallest individual unit of an image. A combination of these pixels will create an image, distinct from the commonly used scalable vector graphics, which are the basis of the vector model. While a digital image is concerned with the output as representation of reality, in a photograph or art transferred to computer, the raster data type will reflect an abstraction of reality. Aerial photos are one commonly used form of raster data, with only one purpose, to display a detailed image on a map or for the purposes of digitization. Other raster data sets will contain information regarding elevation, a DEM (digital Elevation Model), or reflectance of a particular wavelength of light.

Digital elevation model, map, and vector data, Raster data type consists of rows and columns of cells each storing a single value. Raster data can be images (raster images) with each pixel containing a color value. Additional values recorded for each cell may be a discrete value, such as land use, a continuous value, such as temperature, or a null value if no data is available. While a raster cell stores a single value, it can be extended by using raster bands to represent RGB (red, green, blue) colors, colormaps (a mapping between a thematic code and RGB value), or an extended attribute table with one row for each unique cell value. The resolution of the raster data set is its cell width in ground units.

Raster data is stored in various formats; from a standard file-based structure of TIF, JPEG formats to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes. Database storage, when properly indexed, typically allows for quicker retrieval of the raster data but can require storage of millions of significantly sized records.

Vector

A simple vector map, using each of the vector elements: points for wells, lines for rivers, and a polygon for the lake. In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. In the popular ESRI Arc series of programs, these are explicitly called shape files. Different geographical features are best expressed by different types of geometry:

Points

Zero-dimensional points are used for geographical features that can best be expressed by a single grid reference; in other words, simple location. For example, the locations of wells, peak elevations, features of interest or trailheads. Points convey the least amount of information of these file types.

Lines or polylines

One-dimensional lines or polylines are used for linear features such as rivers, roads, railroads, trails, and topographic lines.

Polygons

Two-dimensional polygons are used for geographical features that cover a particular area of the earth's surface. Such features may include lakes, park boundaries, buildings, city boundaries, or land uses. Polygons convey the most amount of information of the file types.

Each of these geometries are linked to a row in a database that describes their attributes. For example, a database that describes lakes may contain a lake's depth, water quality, pollution level. This information can be used to make a map to describe a particular attribute of the dataset. For example, lakes could be coloured depending on level of pollution. Different geometries can also be compared. For example, the GIS could be used to identify all wells (point geometry) that are within 1-mile (1.6 km) of a lake (polygon geometry) that has a high level of pollution.

Vector features can be made to respect spatial integrity through the application of topology rules such as 'polygons must not overlap'. Vector data can also be used to represent continuously varying phenomena. Contour lines and triangulated irregular networks (TIN) are used to represent elevation or other continuously changing values. TINs record values at point locations, which are connected by lines to form an irregular mesh of triangles. The face of the triangles represent the terrain surface.

Advantages and disadvantages

There are advantages and disadvantages to using a raster or vector data model to represent reality. Raster data sets record a value for all points in the area covered which may require more storage space than representing data in a vector format that can store data only where needed. Raster data also allows easy implementation of overlay operations, which are more difficult with vector data. Vector data can be displayed as vector graphics used on traditional maps, whereas raster data will appear as an image that may have a blocky appearance for object boundaries. Vector data can be easier to register, scale, and re-project. This can simplify combining vector layers from different sources. Vector data are more compatible with relational database environment. They can be part of a relational table as a normal column and processes using a multitude of operators.

The file size for vector data is usually much smaller for storage and sharing than raster data. Image or raster data can be 10 to 100 times larger than vector data depending on the resolution. Another advantage of vector data is it can be easily updated and maintained. For example, a new highway is added. The raster image will have to be completely reproduced, but the vector data, "roads," can be easily updated by adding the missing road segment. In addition, vector data allow much more analysis capability especially for "networks" such as roads, power, rail, telecommunications, etc. For example, with vector data attributed with the characteristics of roads, ports, and airfields, allows the analyst to query for the best route or method of transportation. In the vector data, the analyst can query the data for the largest port with an airfield within 60 miles and a connecting road that is at least two lane highway. Raster data will not have all the characteristics of the features it displays.

Voxel

Selected GIS additionally support the voxel data model. A voxel (a portmanteau of the words volumetric and pixel) is a volume element, representing a value on a regular grid in three dimensional space. This is analogous to a pixel, which represents 2D image data. Voxels can be interpolated from 3D point clouds (3D point vector data), or merged from 2D raster slices.

Non-spatial data

Additional non-spatial data can also be stored besides the spatial data represented by the coordinates of a vector geometry or the position of a raster cell. In vector data, the additional data are attributes of the object. For example, a forest inventory polygon may also have an identifier value and information about tree species. In raster data the cell value can store attribute information, but it can also be used as an identifier that can relate to records in another table.

Data capture

Data capture—entering information into the system—consumes much of the time of GIS practitioners. There are a variety of methods used to enter data into a GIS where it is stored in a digital format.

Existing data printed on paper or PET film maps can be digitized or scanned to produce digital data. A digitizer produces vector data as an operator traces points, lines, and polygon boundaries from a map. Scanning a map results in raster data that could be further processed to produce vector data.

Survey data can be directly entered into a GIS from digital data collection systems on survey instruments. Positions from a Global Positioning System (GPS), another survey tool, can also be directly entered into a GIS.

Remotely sensed data also plays an important role in data collection and consist of sensors attached to a platform. Sensors include cameras, digital scanners and LIDAR, while platforms usually consist of aircraft and satellites.

The majority of digital data currently comes from photo interpretation of aerial photographs. Soft copy workstations are used to digitize features directly from stereo pairs of digital photographs. These systems allow data to be captured in 2 and 3 dimensions, with elevations measured directly from a stereo pair using principles of photogrammetry. Currently, analog aerial photos are scanned before being entered into a soft copy system, but as high quality digital cameras become cheaper this step will be skipped.

Satellite remote sensing provides another important source of spatial data. Here satellites use different sensor packages to passively measure the reflectance from parts of the electromagnetic spectrum or radio waves that were sent out from an active sensor such as radar. Remote sensing collects raster data that can be further processed to identify objects and classes of interest, such as land cover.

When data is captured, the user should consider if the data should be captured with either a relative accuracy or absolute accuracy, since this could not only influence how information will be interpreted but also the cost of data capture.

In addition to collecting and entering spatial data, attribute data is also entered into a GIS. For vector data, this includes additional information about the objects represented in the system.

After entering data into a GIS, the data usually requires editing, to remove errors, or further processing. For vector data it must be made "topologically correct" before it can be used for some advanced analysis. For example, in a road network, lines must connect with nodes at an intersection. Errors such as undershoots and overshoots must also be removed. For scanned maps, blemishes on the source map may need to be removed from the resulting raster. For example, a fleck of dirt might connect two lines that should not be connected.

Raster-to-vector translation

Data restructuring can be performed by a GIS to convert data into different formats. For example, a GIS may be used to convert a satellite image map to a vector structure by generating lines around all cells with the same classification, while determining the cell spatial relationships, such as adjacency or inclusion.

More advanced data processing can occur with image processing, a technique developed in the late 1960s by NASA and the private sector to provide contrast enhancement, false colour rendering and a variety of other techniques including use of two dimensional Fourier transforms.

Since digital data are collected and stored in various ways, the two data sources may not be entirely compatible. So a GIS must be able to convert geographic data from one structure to another.

Projections, coordinate systems and registration

A property ownership map and a soils map might show data at different scales. Map information in a GIS must be manipulated so that it registers, or fits, with information gathered from other maps. Before the digital data can be analyzed, they may have to undergo other manipulations—projection and coordinate conversions for example, that integrate them into a GIS.

The earth can be represented by various models, each of which may provide a different set of coordinates (e.g., latitude, longitude, elevation) for any given point on the earth's surface. The simplest model is to assume the earth is a perfect sphere. As more measurements of the earth have accumulated, the models of the earth have become more sophisticated and more accurate. In fact, there are models that apply to different areas of the earth to provide increased accuracy (e.g., North American Datum, 1927 - NAD27 - works well in North America, but not in Europe). See Datum for more information.

Projection is a fundamental component of map making. A projection is a mathematical means of transferring information from a model of the Earth, which represents a three-dimensional curved surface, to a two-dimensional medium—paper or a computer screen. Different projections are used for different types of maps because each projection particularly suits certain uses. For example, a projection that accurately represents the shapes of the continents will distort their relative sizes. See Map projection for more information.

Since much of the information in a GIS comes from existing maps, a GIS uses the processing power of the computer to transform digital information, gathered from sources with different projections and/or different coordinate systems, to a common projection and coordinate system. For images, this process is called rectification.

Spatial Analysis with GIS

Data modeling

It is difficult to relate wetlands maps to rainfall amounts recorded at different points such as airports, television stations, and high schools. A GIS, however, can be used to depict two- and three-dimensional characteristics of the Earth's surface, subsurface, and atmosphere from information points. For example, a GIS can quickly generate a map with isopleths or contour lines that indicate differing amounts of rainfall.

Such a map can be thought of as a rainfall contour map. Many sophisticated methods can estimate the characteristics of surfaces from a limited number of point measurements. A two-dimensional contour map created from the surface modeling of rainfall point measurements may be overlaid and analyzed with any other map in a GIS covering the same area.

Additionally, from a series of three-dimensional points, or digital elevation model, isopleths lines representing elevation contours can be generated, along with slope analysis, shaded relief, and other elevation products. Watersheds can be easily defined for any given reach, by computing all of the areas contiguous and uphill from any given point of interest. Similarly, an expected thalweg of where surface water would want to travel in intermittent and permanent streams can be computed from elevation data in the GIS.

Topological modeling

In the past years, were there any gas stations or factories operating next to the swamp? Any within two miles (3 km) and uphill from the swamp? A GIS can recognize and analyze the spatial relationships that exist within digitally stored spatial data. These topological relationships allow complex spatial modeling and analysis to be performed. Topological relationships between geometric entities traditionally include adjacency (what adjoins what), containment (what encloses what), and proximity (how close something is to something else).

Networks

If all the factories near a wetland were accidentally to release chemicals into the river at the same time, how long would it take for a damaging amount of pollutant to enter the wetland reserve? A GIS can simulate the routing of materials along a linear network. Values such as slope, speed limit, or pipe diameter can be incorporated into network modeling in order to represent the flow of the phenomenon more accurately. Network modeling is commonly employed in transportation planning, hydrology modeling, and infrastructure modeling.

Cartographic modeling

The "cartographic modeling" was (probably) coined by Dana Tomlin in his PhD dissertation and later in his book which has the term in the title. Cartographic modeling refers to a process where several thematic layers of the same area are produced, processed, and analyzed. Tomlin used raster layers, but the overlay method (see below) can be used more generally. Operations on map layers can be combined into algorithms, and eventually into simulation or optimization models.

Map overlay

The combination of two separate spatial data sets (points, lines or polygons) to create a new output vector data set. These overlays are similar to mathematical Venn diagram overlays. A union overlay combines the geographic features and attribute tables of both inputs into a single new output. An intersect overlay defines the area where both inputs overlap and retains a set of attribute fields for each. A symmetric difference overlay defines an output area that includes the total area of both inputs except for the overlapping area.

Data extraction is a GIS process similar to vector overlay, though it can be used in either vector or raster data analysis. Rather than combining the properties and features of both data sets, data extraction involves using a "clip" or "mask" to extract the features of one data set that fall within the spatial extent of another data set.

In raster data analysis, the overlay of data sets is accomplished through a process known as "local operation on multiple rasters" or "map algebra," through a function that combines the values of each raster's matrix. This function may weigh some inputs more than others through use of an "index model" that reflects the influence of various factors upon a geographic phenomenon.

Automated cartography

Digital cartography and GIS both encode spatial relationships in structured formal representations. GIS is used in digital cartography modeling as a (semi) automated process of making maps, so called Automated Cartography. In practice, it can be a subset of a GIS, within which it is equivalent to the stage of visualization, since in most cases not all of the GIS functionality is used. Cartographic products can be either in a digital or in a hardcopy format. Powerful analysis techniques with different data representation can produce high-quality maps within a short time period. The main problem in Automated Cartography is to use a single set of data to produce multiple products at a variety of scales, a technique known as Generalization.

Geostatistics

Geostatistics is a point-pattern analysis that produces field predictions from data points. It is a way of looking at the statistical properties of those special data. It is different from general applications of statistics because it employs the use of graph theory and matrix algebra to reduce the number of parameters in the data. Only the second-order properties of the GIS data are analyzed.

When phenomena are measured, the observation methods dictate the accuracy of any subsequent analysis. Due to the nature of the data (e.g. traffic patterns in an urban environment; weather patterns over the Pacific Ocean), a constant or dynamic degree of precision is always lost in the measurement. This loss of precision is determined from the scale and distribution of the data collection.

To determine the statistical relevance of the analysis, an average is determined so that points (gradients) outside of any immediate measurement can be included to determine their predicted behavior. This is due to the limitations of the applied statistic and data collection methods, and interpolation is required in order to predict the behavior of particles, points, and locations that are not directly measurable.

Interpolation is the process by which a surface is created, usually a raster data set, through the input of data collected at a number of sample points. There are several forms of interpolation, each which treats the data differently, depending on the properties of the data set. In comparing interpolation methods, the first consideration should be whether or not the source data will change (exact or approximate). Next is whether the method is subjective, a human interpretation, or objective. Then there is the nature of transitions between points: are they abrupt or gradual. Finally, there is whether a method is global (it uses the entire data set to form the model), or local where an algorithm is repeated for a small section of terrain.

Interpolation is a justified measurement because of a Spatial Autocorrelation Principle that recognizes that data collected at any position will have a great similarity to, or influence of those locations within its immediate vicinity.

Digital elevation models (DEM), triangulated irregular networks (TIN), Edge finding algorithms, Theissen Polygons, Fourier analysis, Weighted moving averages, Inverse Distance Weighted, Moving averages, Kriging, Spline, and Trend surface analysis are all mathematical methods to produce interpolative data.

Address Geocoding

Geocoding is calculating spatial locations (X,Y coordinates) from street addresses. A reference theme is required to geocode individual addresses, such as a road centerline file with address ranges. The individual address locations are interpolated, or estimated, by examining address ranges along a road segment. These are usually provided in the form of a table or database. The GIS will then place a dot approximately where that address belongs along the segment of centerline. For example, an address point of 500 will be at the midpoint of a line segment that starts with address 1 and ends with address 1000. Geocoding can also be applied against actual parcel data, typically from municipal tax maps. In this case, the result of the geocoding will be an actually positioned space as opposed to an interpolated point.

It should be noted that there are several (potentially dangerous) caveats that are often overlooked when using interpolation. See the full entry for Geocoding for more information.

Various algorithms are used to help with address matching when the spellings of addresses differ. Address information that a particular entity or organization has data on, such as the post office, may not entirely match the reference theme. There could be variations in street name spelling, community name, etc. Consequently, the user generally has the ability to make matching criteria more stringent, or to relax those parameters so that more addresses will be mapped. Care must be taken to review the results so as not to erroneously map addresses incorrectly due to overzealous matching parameters.

Reverse geocoding

Reverse geocoding is the process of returning an estimated street address number as it relates to a given coordinate. For example, a user can click on a road centerline theme (thus providing a coordinate) and have information returned that reflects the estimated house number. This house number is interpolated from a range assigned to that road segment. If the user clicks at the midpoint of a segment that starts with address 1 and ends with 100, the returned value will be somewhere near 50. Note that reverse geocoding does not return actual addresses, only estimates of what should be there based on the predetermined range.

Data output and cartography

Cartography is the design and production of maps, or visual representations of spatial data. The vast majority of modern cartography is done with the help of computers, usually using a GIS. Most GIS software gives the user substantial control over the appearance of the data.

Cartographic work serves two major functions:

First, it produces graphics on the screen or on paper that convey the results of analysis to the people who make decisions about resources. Wall maps and other graphics can be generated, allowing the viewer to visualize and thereby understand the results of analyses or simulations of potential events. Web Map Servers facilitate distribution of generated maps through web browsers using various implementations of web-based application programming interfaces(AJAX, Java, Flash, etc).

Second, other database information can be generated for further analysis or use. An example would be a list of all addresses within one mile (1.6 km) of a toxic spill.

Graphic display techniques

Traditional maps are abstractions of the real world, a sampling of important elements portrayed on a sheet of paper with symbols to represent physical objects. People who use maps must interpret these symbols. Topographic maps show the shape of land surface with contour lines; the actual shape of the land can be seen only in the mind's eye.

Today, graphic display techniques such as shading based on altitude in a GIS can make relationships among map elements visible, heightening one's ability to extract and analyze information. For example, two types of data were combined in a GIS to produce a perspective view of a portion of San Mateo County, California.

The digital elevation model, consisting of surface elevations recorded on a 30-meter horizontal grid, shows high elevations as white and low elevation as black.
The accompanying Landsat Thematic Mapper image shows a false-color infrared image looking down at the same area in 30-meter pixels, or picture elements, for the same coordinate points, pixel by pixel, as the elevation information.

A GIS was used to register and combine the two images to render the three-dimensional perspective view looking down the San Andreas Fault, using the Thematic Mapper image pixels, but shaded using the elevation of the landforms. The GIS display depends on the viewing point of the observer and time of day of the display, to properly render the shadows created by the sun's rays at that latitude, longitude, and time of day.

Spatial ETL

Spatial ETL tools provide the data processing functionality of traditional Extract, Transform, Load (ETL) software, but with a primary focus on the ability to manage spatial data. They provide GIS users with the ability to translate data between different standards and proprietary formats, whilst geometrically transforming the data en-route.

GIS software

Geographic information can be accessed, transferred, transformed, overlaid, processed and displayed using numerous software applications. Within industry commercial offerings from companies such as ESRI and Mapinfo dominate, offering an entire suite of tools. Government and military departments often use custom software, open source products, such as Gram++, GRASS, or more specialized products that meet a well-defined need. Free tools exist to view GIS datasets and public access to geographic information is dominated by online resources such as Google Earth and interactive web mapping.

Originally up to the late 1990s, when GIS data was mostly based on large computers and used to maintain internal records, software was a stand-alone product. However with increased access to the Internet and networks and demand for distributed geographic data grew, GIS software gradually changed its entire outlook to the delivery of data over a network. GIS software is now usually marketed as combination of various interoperable applications and APIs.

Data creation

GIS processing software is used for the task of preparing data for use within a GIS. This transforms the raw or legacy geographic data into a format usable by GIS products. For example an aerial photograph may need to be stretched using photogrammetry so that its pixels align with longitude and latitude gradations. This can be distinguished from the transformations done within GIS analysis software by the fact that these changes are permanent, more complex and time consuming. Thus, a specialized high-end type of software is generally used by a skilled person in GIS processing aspects of computer science for digitization and analysis. Raw geographic data can be edited in many standard database and spreadsheet applications and in some cases a text editor may be used as long as care is taken to properly format data.

A geo-database is a database with extensions for storing, querying, and manipulating geographic information and spatial data.

Management and analysis

GIS analysis software takes GIS data and overlays or otherwise combines it so that the data can be visually analysed. It can output a detailed map, or image used to communicate an idea or concept with respect to a region of interest. This is usually used by persons who are trained in cartography, geography or a GIS professional as this type of application is complex and takes some time to master. The software performs transformation on raster and vector data sometimes of differing datums, grid system, or reference system, into one coherent image. It can also analyse changes over time within a region. This software is central to the professional analysis and presentation of GIS data. Examples include the ArcGIS family of ESRI GIS applications, Smallworld, Gram++ and GRASS.

Statistical

GIS statistical software uses standard database queries to retrieve data and analyse data for decision making. For example, it can be used to determine how many persons of an income of greater than 60,000 live in a block. The data is sometimes referenced with postal codes and street locations rather than with geodetic data. This is used by computer scientists and statisticians with computer science skills, with an objective of characterizing an area for marketing or governing decisions. Standard DBMS can be used or specialized GIS statistical software. These are many times setup on servers so that they can be queried with web browsers. Examples are MySQL or ArcSDE.

Readers

GIS readers are computer applications that are designed to allow users to easily view digital maps as well as view and query GIS-managed data. By definition, they usually allow very little if any editing of the map or underlying map data. Readers can be normal standalone applications that need to be installed locally, though they are often designed to connect to data servers over the Internet to access the relevant information. Readers can also be included as an embedded application within a web page, obviating the need for local installation. Readers are designed to be relatively simple and easy to use as well as free.

Web API

This is the evolution of the scripts that were common with most early GIS systems. An Application Programming Interface (API) is a set of subroutines designed to perform a specific task. GIS APIs are designed to manage GIS data for its delivery to a web browser client from a GIS server. They are accessed with commonly used scripting language such as VBA or JavaScript. They are used to build a server system for the delivery of GIS that is to make available over an Intranet.

Distributed GIS

Distributed GIS concerns itself with Geographical Information Systems that do not have all of the system components in the same physical location. This could be the processing, the database, the rendering or the user interface. Examples of distributed systems are web-based GIS, Mobile GIS, Corporate GIS and GRID computing.

Mobile GIS

GIS has seen many implementations on mobile devices. With the widespread adoption of GPS, GIS has been used to capture and integrate data in the field.

Open-source GIS software

Many GIS tasks can be accomplished with open-source GIS software, which are freely available over Internet downloads. With the broad use of non-proprietary and open data formats such as the Shape File format for vector data and the Geotiff format for raster data, as well as the adoption of OGC standards for networked servers, development of open source software continues to evolve, especially for web and web service oriented applications. Well-known open source GIS software includes GRASS GIS, Quantum GIS, MapServer, uDig, OpenJUMP, gvSIG and many others. PostGIS provides an open source alternative to geo-databases such as Oracle Spatial, and ArcSDE.

The future of GIS

Many disciplines can benefit from GIS technology. An active GIS market has resulted in lower costs and continual improvements in the hardware and software components of GIS. These developments will result in a much wider use of the technology throughout science, government, business, and industry. The GIS applications including public health, crime mapping, national defense, sustainable development, agriculture, rural development, natural resources, landscape architecture, archaeology, regional and community planning, transportation and logistics. GIS is also diverging into location-based services (LBS). LBS allows GPS enabled mobile devices to display their location in relation to fixed assets (nearest restaurant, gas station, police station), mobile assets (friends, children, police car) or to relay their position back to a central server for display or other processing. These services continue to develop with the increased integration of GPS functionality with increasingly powerful mobile electronics such as cell phones, PDAs, laptops.

Web Mapping

In recent years there has been an explosion of mapping applications on the web such as Google Maps, and Live Maps. These websites give the public access to huge amounts of geographic data with an emphasis on aerial photography. Some of them, like Google Maps, expose an API that enable users to create custom applications. These vendors' applications offer street maps and aerial/satellite imagery that support such features as geocoding, searches, and routing functionality.

Some GIS applications also exist for publishing geographic information on the web that include MapInfo's MapXtreme, Intergraph's GeoMedia WebMap, ESRI's ArcIMS, ArcGIS Server, AutoDesk's Mapguide and the open source MapServer.

Exploring Global Change with GIS

Maps have traditionally been used to explore the Earth and to exploit its resources. GIS technology, as an expansion of cartographic science, has enhanced the efficiency and analytic power of traditional mapping. Now, as the scientific community recognizes the environmental consequences of human activity, GIS technology is becoming an essential tool in the effort to understand the process of global change. Various map and satellite information sources can combine in modes that simulate the interactions of complex natural systems.

Through a function known as visualization, a GIS can be used to produce images - not just maps, but drawings, animations, and other cartographic products. These images allow researchers to view their subjects in ways that literally never have been seen before. The images often are equally helpful in conveying the technical concepts of GIS study-subjects to non-scientists.

Adding the dimension of time

The condition of the Earth's surface, atmosphere, and subsurface can be examined by feeding satellite data into a GIS. GIS technology gives researchers the ability to examine the variations in Earth processes over days, months, and years.

As an example, the changes in vegetation through a growing season can be animated to determine when drought was most extensive in a particular region. The resulting graphic, known as a normalized vegetation index, represents a rough measure of plant health. Working with two variables over time would then allow researchers to detect regional differences in the lag between a decline in rainfall and its effect on vegetation. GIS technology and the availability of digital data on regional and global scales enable such analyses. The satellite sensor output used to generate a vegetation graphic is produced by the Advanced Very High Resolution Radiometer (AVHRR). This sensor system detects the amounts of energy reflected from the Earth's surface across various bands of the spectrum for surface areas of about 1 square kilometer. The satellite sensor produces images of a particular location on the Earth twice a day. AVHRR is only one of many sensor systems used for Earth surface analysis

GIS and related technology will help greatly in the management and analysis of these large volumes of data, allowing for better understanding of terrestrial processes and better management of human activities to maintain world economic vitality and environmental quality.

Semantics and GIS

Tools and technologies emerging from the W3C's Semantic Web Activity are proving useful for data integration problems in information systems. Correspondingly, such technologies have been proposed as a means to facilitate interoperability and data reuse among GIS applications and also to enable new mechanisms for analysis.

Ontologies are a key component of this semantic approach as they allow a formal, machine-readable specification of the concepts and relationships in a given domain. This in turn allows a GIS to focus on the meaning of data rather than its syntax or structure. For example, reasoning that a land cover type classified as Deciduous Needle leaf Trees in one dataset is a specialization of land cover type Forest in another more roughly-classified dataset can help a GIS automatically merge the two datasets under the more general land cover classification. Very deep and comprehensive ontologies have been developed in areas related to GIS applications, for example the Hydrology Ontology developed by the Ordnance Survey in the United Kingdom. Also, simpler ontologies and semantic metadata standards are being proposed by the W3C Geo Incubator Group to represent geospatial data on the web.

Recent research results in this area can be seen in the International Conference on Geospatial Semantics and the Terra Cognita -- Directions to the Geospatial Semantic Web workshop at the International Semantic Web Conference.

Remote Sensing Technology

Introduction

Remote Sensing (RS) is a technology that provides the means to collect and use geographic data to assist in the development of Agriculture. Remote Sensing in the most generally accepted meaning refers to instrument-based techniques employed in the acquisition and measurement of spatially organized or geographically distributed data on some properties such as spectral, spatial, physical of an array of target points of objects and materials from a define distance from the observed target. Remote sensing of the environment by geographers is usually done with the help of mechanical devices known as remote sensors. These gadgets have a greatly improved ability to receive and record information about an object without any physical contact. Often, these sensors are positioned away from the object of interest by using helicopters, planes, and satellites. Most sensing devices record information about an object by measuring an object's transmission of electromagnetic energy from reflecting and radiating surfaces.

Remote sensing imagery has many applications in mapping land use and cover, agriculture, soils mapping, forestry, city planning, archaeological investigations, military observation, and geological surveying.

Overview of Remote Sensing Technology

Remote Sensing is the technology that is now the principal tool by which the Earth's surface and atmosphere, the planets, and the entire Universe are being observed, measured, and interpreted from such vantage points as the terrestrial surface, earth-orbit, and outer space. The term "remote sensing" was coined by Ms Evelyn Pruitt in the mid-1950's when she was working with the U.S. Office of Naval Research (ONR) outside Washington, D.C as a oceanographer.

Remote Sensing is the most generally accepted meaning refers to "Instrument-based techniques employed in the acquisition and measurement of spatially organized data/information on some properties such as spectral, spatial, physical of an array of target points within the sensed scene that correspond to features, objects, and materials, doing this by applying one or more recording devices not in physical, intimate contact with the item(s) from at a finite distance from the observed target, in which the spatial arrangement is preserved. Various techniques involve pertinent to the sensed scene (target) by utilizing electromagnetic radiation, force fields, or acoustic energy sensed by recording cameras, radiometers and scanners, lasers, radio frequency receivers, radar systems, sonar, thermal devices, sound detectors, seismographs, magnetometers, gravimeters, scintillometers, and other instruments.

In simpler terms, Remote Sensing can be defined as "gathering data and information about the physical 'world' by detecting and measuring signals composed of radiation, particles, and fields emanating from objects located beyond the immediate vicinity of the sensor devices".

In the broadest sense, remote sensing is the small or large-scale acquisition of information of an object or phenomenon, by the use of either recording or real-time sensing devices that is not in physical or intimate contact with the object such as by way of aircraft, spacecraft, satellite. In practice, remote sensing is the stand-off collection through the use of a variety of devices for gathering information on a given object or area. Thus, Earth observation or weather satellite collection platforms, ocean and atmospheric observing weather buoy platforms, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and space probes are all examples of remote sensing. In modern usage, the term generally refers to the use of imaging sensor technologies including but not limited to the use of instruments aboard aircraft and spacecraft, and is distinct from other imaging-related fields such as medical imaging.

There are two kinds of remote sensing. (1) Passive sensors detect natural energy / radiation that is emitted or reflected by the object or surrounding area being observed. Reflected sunlight is the most common source of radiation measured by passive sensors. Examples of passive remote sensors include film photography, infrared, and radiometers. (2) Active collection, on the other hand, emits energy in order to scan objects and areas whereupon a passive sensor then detects and measures the radiation that is reflected or backscattered from the target. RADAR is an example of active remote sensing where the time delay between emission and return is measured, establishing the location, height, speed and direction of an object.

Remote sensing makes it possible to collect data on inaccessible areas. Remote sensing applications include monitoring deforestation, the effects of climate change on Arctic and Antarctic regions, coastal and ocean depths, availability of water in the ground, and many more.

Orbital platforms collect and transmit data from different parts of the electromagnetic spectrum, which in conjunction with larger scale aerial or ground-based sensing and analysis, provides researchers with enough information to monitor trends such natural long and short term phenomena. Other uses include different areas of the earth sciences such as natural resource management, agricultural fields such as land usage and conservation, national security, ground-based and stand-off collection on border areas.

History of Remote Sensing

Beyond the primitive methods of remote sensing our earliest ancestors used to standing on a high mountains or tree to view the landscape. The modern discipline arose with the development of flight. The balloonist made photographs of cities from their balloons. The first tactical use was during the civil war. Messenger pigeons, kites, rockets and unmanned balloons were also used for early images. With the exception of balloons, these first, individual images were not particularly useful for map making or for scientific purposes.

Systematic aerial photography was developed for military use beginning in World War I and reaching a climax during the Cold War with the use of modified combat aircraft. A more recent development is that of increasingly smaller sensor pods such as those used by law enforcement and the military, in both manned and unmanned platforms. The advantage of this approach is that this requires minimal modification to a given airframe. Later imaging technologies would include Infra-red, conventional, doppler and synthetic aperture radar

The development of artificial satellites in the latter half of the 20th century allowed remote sensing to progress to a global scale as of the end of the cold war. Instrumentation aboard various Earth observing and weather satellites such as Landsat, the Nimbus and more recent missions such as RADARSAT and UARS provided global measurements of various data for civil, research, and military purposes. Space probes to other planets have also provided the opportunity to conduct remote sensing studies in extra-terrestrial environment, synthetic aperture radar aboard the Magellan spacecraft provided detailed topographic maps of Venus.

Recent developments include, beginning in the 1960s and 1970s with the development of image processing of satellite images. Several research groups in Silicon Valley including NASA, developed Fourier transform techniques leading to the first notable enhancement of imagery data.

The introduction of online web services for easy access to remote sensing data in the 21st century mainly low/medium-resolution images, like Google Earth, has made remote sensing more familiar to the every one and has popularized the science.

Data acquisition techniques

Electromagnetic Radiation

Remote sensing is the practice of measuring an object or a phenomenon without being in direct contact with it. It is non-intrusive. This requires the use of a sensor situated remotely from the target of interest. A sensor is the instrument (camera) that takes the remote measurements. There are many different types of sensors, but almost all of them share something what they "sense" or take measurements of is usually Electro-Magnetic Radiation (EMR) or light energy. EMR is energy propagated through space in the form of tiny energy packets called photons that exhibit both wave-like and particle-like properties. Unlike other modes of energy transport, such as conduction (heating a metal skillet) or convection (flying a hot air balloon), radiation (as in EMR) is capable of propagating through the vacuum of space. The speed of that EMR in a vacuum (outer space) is approximately 300,000 kilometers per second (3 x 108 meters/second-1 or 186,000 miles/second-1). This is an extremely fast communications medium with visible light with its red, green, and blue colors that we see daily are an example of EMR. But there is a much larger spectrum of such energy. We often characterize this spectrum or range in terms of the wavelengths of different kinds of EMR. For a variety of reasons, there are some wavelengths of EMR that are more commonly used in remote sensing than other wavelengths.

Digital Sensors

Digital sensors also measure patterns in incoming EMR using analog detectors. However, measurements of EMR taken by each detector element are recorded, not using an analog medium such as film, but using numbers. These measurements are digitized through a process called analog-to-digital (A-to-D) conversion. Possible values are in a pre-defined range, such as 0 to 255. Each recorded numerical value is then stored on some kind of digital medium, such as a hard disk, as part of a raster dataset. The value in each raster cell represents the amount of energy received at the sensor from a particular circular area, instantaneous-field-of-view (IFOV) on the ground. Digital sensors make use of the same basic technology as a computer document scanner or a digital camera. In fact, specialized digital cameras are often used to acquire remote sensor data and professional-grade document scanners are often used to convert analog remote sensing data to digital data.

The detectors in a digital sensor can be arranged in a number of different ways. One method utilizes a single detector for each frequency band. A scanning mirror is then used to capture EMR at each IFOV along a scan line. The forward motion of the sensor allows for additional scan lines and therefore a two dimensional image. This is type of instrument is often referred to as a scanning mirror sensor.

A second method is to have a linear array of detectors for each band. Each detector in an array records EMR for a single IFOV in the cross-track dimension i.e., perpendicular to the direction of flight. The forward motion of the sensor again allows for repeated measurements and two-dimensional imagery. This type of sensor system is often called a linear array push-broom scanner. Push-broom systems have several advantages over scanning mirror sensors. They have fewer moving parts, so they are generally more durable. Also, the process of assigning coordinates to push-broom data is much easier.

A third digital sensor configuration is the one that is most like the operation of analog film-based systems. In this case, an entire area array is placed at the back of the sensor. Energy is focused through a lens onto this bank of detectors. These types of sensors are called digital cameras, or area array sensors. They are often used in similar applications as film-based cameras.

Types of Resolution

Resolution quantifies how distinguishable the individual parts of an object or phenomenon are. When discussing the specifications of remote sensor systems, we generally speak of four different types of resolution.

A. Temporal Resolution

Temporal resolution is how often a sensor visits, or can visit, a particular site to collect data. This is important because many applications depend on observing change in phenomena over time. A remote sensing instrument is mounted on a platform such as a satellite, an aircraft, a hot air balloon. The platform on which a sensor is mounted is the greatest determinant of that sensor's temporal resolution.

Some satellites orbit Earth without ever approaching its shadow - that is, they are in Sun-synchronous orbit. Other satellites maintain a fixed position above the rotating Earth - these are in geo-synchronous orbit. In either case, these satellites have a regular and predictable temporal resolution (every 16 days). Some satellite-based sensors are more flexible than other ones because of their ability to point at various targets near their default field-of-view. These more flexible sensors may have a temporal resolution range (2-3 days). Sensors mounted on aircraft fly ad-hoc or on-demand missions with less predictable but more flexible temporal resolution (every hour).

B. Spatial Resolution

Spatial resolution describes the size of the individual measurements taken by the remote sensor system. This concept is closely related to scale. With an analog sensor, such as film, the spatial resolution is commonly expressed in the same terms as the scale (e.g., 1:500). Since digital sensor records information in raster format the spatial resolution is the cell size (e.g., 3 x 3 meters) in ground units.

C. Spectral Resolution

Spectral resolution describes the sensor systems' ability to distinguish different portions of the EMR spectrum. Some sensors are sensitive to visible light only, while others can also capture near-infrared energy. The portions of the spectrum to which an instrument is sensitive are referred to as its bands. A sensor can have multiple bands, and bands can be of varying widths. Spectral resolution refers both to the number and width of the bands for a given sensor.

A panchromatic band is a wide band that encompasses a large spectral range, often the entire visible spectrum. Commonly we call film that is sensitive to the entire visible range "black and white" film because often we print images from this sort of film in grayscale. However, there are analog and digital sensors that have wide panchromatic bands that also encompass the near infrared portion of the spectrum.

When a sensor records only a few portions of the spectrum i.e., contains only a few, relatively wide bands, it is said to be a multispectral system. A multispectral sensor might have two or three bands in the visible range i.e. red, green, and blue and it might also have a few near-infrared or middle infrared bands. Typical multispectral systems have between 4 and 10 bands.

Hyperspectral sensors have a large number of relatively narrow bands. By definition, hyperspectral sensors have a higher spectral resolution than multispectral sensors. Commonly a sensor is considered hyperspectral when it has at least 20 or 30 bands. Many such sensors have hundreds of bands. In general, a sensor with more spectral bands has a greater ability to distinguish between two objects with similar spectral properties.

Each band in a digital dataset can be thought of as an individual raster layer. Visualize an image in three dimensions, with rows, columns, and bands filling the x, y, and z coordinates of a cube.

D. Radiometric Resolution

Radiometric resolution describes the number of unique values that can be recorded by a sensor system when measuring reflected or emitted EMR. In a digital system this is easily quantified as a number. Since the digital numbers in remote sensor data are stored in a computer, they are often expressed in terms of how many bits are used to store that variety of numbers (Ex., 8-bits, 11-bits). An 8-bit sensor would store a value for each measurement in an integer range from 0 to 255. This range has 28-256 discrete values. With analog, or film-based, systems it is the quality of the film that determines its radiometric resolution.

Converting Remote Sensing Data into Geospatial Data

Remote sensing applications are rarely successful without at least some direct measurements / ground truth being taken within the area. However, "truth" is really a misnomer since there is always at least some error in measurements, even if they are taken directly. "Ground reference" would be a better descriptor. A correct term for measurements taken directly as opposed to remote measurements is in situ data collection. Several types of in situ measurements may be necessary for a given project or application. Almost all remote sensing projects require some amount of in situ data collection in order to perform geometric and radiometric calibration. Additional in situ data may be required to create reference maps of spatial variables, including biophysical properties.

Geometric Correction

When remote sensor data is initially collected it is not geospatial data. In order to make the transition to geospatial data, geometric correction must be applied to make the data into a real-world coordinate system. Beyond having no real-world coordinates assigned, the raw data also contains geometric distortion. This means that all of the objects or phenomena that can be seen in the data are not equally out of place relative to a desired coordinate system. Distortion generally increases away from the point in the data that were acquired at straight down. Distortion is therefore different depending on the sensor configuration (Ex., scanning mirror sensors vs. area array digital cameras). Another source of distortion are variations in the terrain and objects on the terrain. Tall objects and steeply sloping terrain lead to more distortion than flat objects on flat terrain.

A basic method for geometric correction involves the use of a GPS receiver in the field. GPS measurements are taken at locations that are also easily identifiable in the imagery. These types of locations will vary according to the spatial resolution of the remote sensor data. Ideally the smallest possible features that can be visualized in that data should be located in the field and their positions surveyed. These features should also be permanently situated. The recorded locations of these features in the study area are collectively known as control points. Road intersections typically make good control points. Features above the ground surface do not make good control points because they cause distortion. Control points should be collected at locations spaced evenly throughout the remote sensor image. In fact, the relative location of the control points is at least as important as the number of points.

Once enough control points have been collected, they can be used to adjust the data to its approximate spatial position within a coordinate system. Most geospatial software packages provide an interface for doing this. As part of the process, the software package will typically report a number indicating the degree to which the desired transformation was successfully implemented. The success rate depends on the amount of distortion present in the raw data. Once the remote sensor data has undergone this process it is said to be georectified data.

Photogrammetry Correction

In order to create an image that is free from all major distortions, the terrain and sensor-induced distortions must be accounted for explicitly. This is done by using a combination of GPS control points, a digital elevation model (DEM), and a detailed report of the distortion present in the sensor system. When data has been corrected in this manner it is said to be orthorectified. In an orthorectified image, all points are in their proper x, y position and aligned as they would appear if one were looking straight down at them.

The practice of orthorectification is part of photogrammetry — the art of taking direct measurements from photos and other remotely sensed data. Measurements derived using photogrammetric techniques include the height of objects on the terrain, their x, y location, and the ground distance between objects.

Radiometric Correction

In addition to geometric distortion, EMR that is received by the sensor contains radiometric distortions. The source of these distortions is primarily the atmosphere and its dynamic constituents. If there were no atmosphere with which to contend, EMR recorded by the sensor would be a much more perfect representation of EMR reflected or emitted from the target object or phenomena. However, along the path between the target and the sensor, EMR must interact twice with the atmosphere. Some of this energy is scattered and some of it is absorbed. Atmospheric constitutes such as water vapor and pollution vary across space and time, and therefore these distortions make it particularly difficult to compare datasets collected at different times.

There are various ways to minimize this distortion. Between-date radiometric differences can be minimized if the datasets are collected at similar times so that the

Sun's position is held constant. Also, acquiring data on a clear day will minimize the amount of water vapor and clouds.

Even after taking these measures, many applications require additional radiometric correction to account for differences and distortions in the EMR values recorded at the sensor. This can be done in a few different ways, each with some degree of difficulty and level of uncertainty in the results. Following are three examples, of many that one could give.

One simple radiometric correction technique is to rescale all of the pixel brightness values in an image by identifying one of the darkest pixels and one of the brightest pixels. The darkest pixel is re-assigned a value of 0, and the lightest a value of 255. The intermediate values are then rescaled to fit evenly in between. Although this method is very easy and requires no additional input data, it is the least reliable. This technique is known as a min-max contrast stretch.

A second, simple method is referred to as empirical line calibration. In this method, several in situ radiometric measurements are taken over various objects concurrently with the acquisition of the remote sensor data. The instrument used for these measurements is called a radiometer. Unlike the remote sensor system, the radiometer is used to take measurements in situ with almost no atmosphere with which to contend. The data collected using the radiometer is used to develop a simple linear mathematical function to predict what the radiometric values should be over the entire image.

A third method is more complex than either of the previous two. It relies on collecting explicit information on the environmental conditions at the time of the remote sensor data acquisition. This information might include a detailed profile of temperature and humidity within the atmospheric column, the Sun-Earth geometry, and the position of the sensor with respect to each pixel. This method is actually a group of methods, each requiring differing information. Automated computer algorithms are then used to process the remote sensor data along with the ancillary data to produce a radiometrically-corrected image.

Automated Classification

Although manual image interpretation is valuable and often provides highly detailed and accurate information, many applications require that objects on the ground be classified faster and more cost-effectively. In these cases it is necessary to automatically interpret, or classify, the image using computer algorithms.

There are primarily two different ways to approach this goal. Both are based on the simple concept that similar objects or phenomena have similar spectral reflectance properties. The first method is referred to as an unsupervised classification. In this method, the computer algorithm operates without any prior knowledge of the scene. Pixels are grouped together based on the similarity of their spectral characteristics. These clusters of similar pixels, representing unique spectral classes, are then reported to the user, who is responsible for transforming them into information classes. This process can be aided with in situ data and/or manual image interpretation.

Supervised classification, is the second method, requires that the user have some knowledge of the actual objects and phenomena within the image. This knowledge could have been acquired through in situ data collection or manual image interpretation. The user specifies the classes (Ex. water, forest, crops) and then instructs, or trains, a computer algorithm by feeding it the exact location of several training examples for each class throughout the image. The computer algorithm examines the properties of these areas and then seeks similar regions throughout the image, eventually classifying the entire image. Spectral data is often the primary data source considered in the process, although recently more effort has been made to incorporate more advanced aspects, such as object shape and relative position.

Mapping Spatial Variables

There are certain aspects of phenomena that must be sensed while in direct contact with an object. For example, it is impossible to directly measure the live biomass, the amount of living matter, present in a stand of vegetation without harvesting the vegetation, processing it to remove water and foreign substances, and then weighing it. However, it is possible for a trained person to estimate the biomass present in a particular stand of vegetation without coming in direct contact with it. In a similar manner, remote sensing principles provide a way to quantify what is "seen" and provide information, such as biomass or other biophysical variables, which are present in an image.

Mapping biomass requires taking some in situ measurements of the vegetation of interest. These measurements are used to build a mathematical model relating to the quantity of biomass to the spectral reflectance values in the remote sensor data. An example of this type of equation might be:

Biomass = Bias + (Constant A x Near-infrared) + (Constant B x Red)

In addition to using the band values directly, it has been shown that specific mathematical combinations of band values are effective for mapping various phenomena. For example, the Normalized Difference Vegetation Index (NDVI) is often highly related to a number of vegetation properties, including green biomass. There are many other band indices for use in vegetation, geologic, and other application areas.

Applications of Remote Sensing and GIS

Remote sensing is an important tool to provide important information on soils, land evaluation, land degradation, crop distribution, crop growth, availability of water resources etc. The information of Remote Sensing can be improved in its efficiency by combining with conventional technologies / ground surveys and also the advanced tools such as GIS for analysis and interpretation.

Remote sensing data is available in digital form and can be used as an input layer to GIS software. The software such as ArcInfo/ERDAS, which supports for both remote sensing and GIS data. The advent of technology in storage capacity, processing capabilities, relational databases, and enhanced graphical user interface has given more capabilities to work on remote sensing and GIS data for analysis and interpretation of data. Use of GIS in combination with remote sensing enhances the decision-making in following ways;

- Process identification to enable comparison of different acquisitions through time
- Identification of agricultural and other development problems
- Evaluation of possible technical interventions for conservation or reclamation measures.
- Monitoring of soils, water, and land degradation processes.

Crop Production Databases

Crop production database is used to know how many hectares have been cultivated, where the cultivation has occurred and how will be likely production of food i.e. Area and Production various crops can be assed with the help of remote sensing and GIS applications. Crop distribution help in modeling of climatic and other environmental changes and their effects on agriculture.

Crop growth and yield determination:

Crop growth and yield are determined by a number of factors such as genetic potential of crop cultivar, soil, weather, cultivation practices such as date of sowing, amount of irrigation and fertilizer and biotic stresses. However, generally for a given area, year-to-year yield variability has been mostly modeled through weather as a predictor using either empirical or crop simulation approach. With the launch and continuous availability of multi-spectral (visible, near-infrared) sensors on polar orbiting earth observation satellites remote sensing data has become an important tool for yield modeling. RS data provide timely, accurate, synoptic and objective estimation of crop growing conditions or crop growth for developing yield models and issuing yield forecasts at a range of spatial scales. RS data have certain advantage over meteorological observations for yield modeling, such as dense observational coverage, direct viewing of the crop and ability to capture effect of non-meteorological factors. An integration of the three technologies, viz., crop simulation models, RS data and GIS can provide an excellent solution to monitoring and modeling of crop at a range of spatial scales.

Crop monitoring

The use of GIS along with RS data for crop monitoring is an established approach in all phases of the activity, namely preparatory, analysis and output. In the preparatory phase GIS is used for (a) stratification/zonation using one or more input layers (climate, soil, physiolgraphy, crop dominance etc.), or (b) preparing input data (weather, soil and collateral data) which is available in different formats to a common format. In the analysis phase use of GIS is mainly through operations on raster layers of NDVI or computing VI profiles within specified administrative boundaries. The final output phase also involves GIS for aggregation and display of outputs for defined regions (e.g., administrative regions) and creating map output products with required data integration through overlays.

- Statistically sound sample points / grids will be identified for various land degradation classes from interpreted map for ground truth collection and for accuracy assessment.
- Field work has to be carried out by the interpreter including soil sample collection along with site details.
- During the field work the relationship between image elements and tentatively identified land degradation classes will be established that are delineated during preliminary interpretation. The sample points will be readjusted depending upon the variability in the field and sufficient points will be collected for finalisation of maps and accuracy assessment.
- The preliminarily interpreted land degradation map will be finalised in light of ground truth data and soil sample analysis (wherever done) to arrive at the final map. Existing legacy data on forests, wastelands, degraded lands, biodiversity, land use / land cover etc. canl be made use of for better delineation of land degradation classes.
- The minimum mapping polygon size of 3 mm x 3 mm on 1:50,000 scale equivalent to 2.25 ha area, would be followed while delineating the degraded lands from satellite data.
- Quality check has to be performed randomly and thematic maps are to be assessed for thematic as well as

location accuracies.

- Digital geo-database would be developed to address retrieval and storage of different data inputs and outputs. Meta data elements have to be designed those area relevant to different types of input data.

The land degradation databases thus developed can be used along with various other thematic data sets like land cover, digital elevation model, climatic data sets for developing suitable reclamation plans.

The Seat Belt System

Why Wear Seat Belts

Wearing seat belts, and wearing them properly, is fundamental to your safety and the safety of your passengers.

During a crash or emergency stop, seat belts can help keep you from being thrown against the inside of the car, against other occupants, or out of the car.

Of course, seat belts cannot completely protect you in every crash. But, in most cases, seat belts reduce your chance of serious injury, and can even save your life. That is why many

WARNING

states and all Canadian provinces require you to wear seat belts.

Not wearing a seat belt increases the chance of being killed or seriously hurt in a crash.

Be sure you and your passengers always wear seat belts and wear them properly.

Important Safety Reminders

Seat belts are designed for adults and larger children. All infants and small children must be properly restrained in child safety seats (see page 30).

- A pregnant woman needs to wear a seat belt to protect herself and her unborn child (see page 26).

- Two people should never use the same seat belt. If they do, they could be very seriously injured in a crash.

- Do not place the shoulder portion of a lap/shoulder belt under your arm or behind your back. This could increase the chance of serious injuries in a crash.

- Do not put shoulder belt pads or other accessories on seat belts. They can reduce the effectiveness of the belts and increase the chance of injury.

- On U.S. models, the front lap belt must always be worn in conjunction with the automatic front shoulder belt.

Seat Belt System Components

The following pages cover the seat belt system used in the U.S. Integras. For information about the seat belts in Canadian models, see page 20.

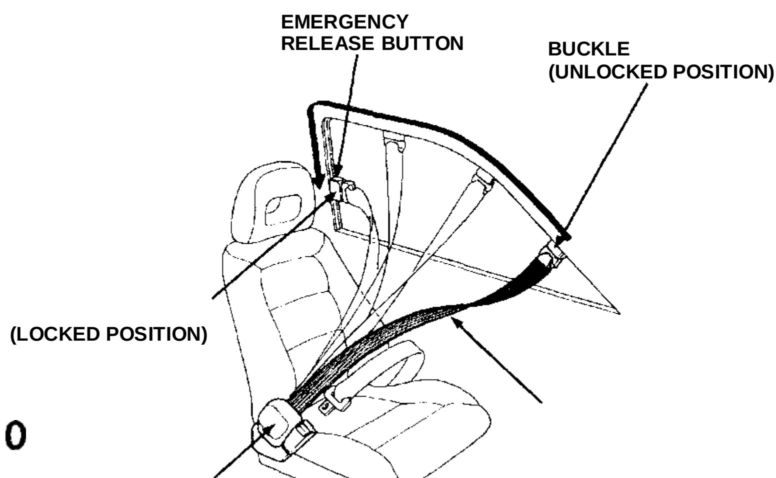
Your Integra has seat belts in all five seating positions. The front seats have automatic shoulder belts with separate lap belts. The outside seating positions in the rear have lap/shoulder belts, and the center position has a lap belt.

Your seat belt system also includes a seat belt indicator light and beeper to remind you to fasten your seat belt and make sure your passengers fasten theirs. The light and beeper are also used to indicate a malfunction of the automatic shoulder belt (see page 13).

Automatic Front Shoulder Belt

This belt automatically moves into position when the driver andfront seat passenger close their doors with the ignition ON (II). The belt has an emergency locking retractor. In normal driving, theretractor lets you move freely in your seat while it keeps sometension on the belt. During a collision or sudden stop, the retractor automatically locks the belt to help restrain your body.

Both automatic shoulder belts should always remain buckled. In an emergency, however, you can unlatch a shoulder belt by pushing the release button on the buckle.



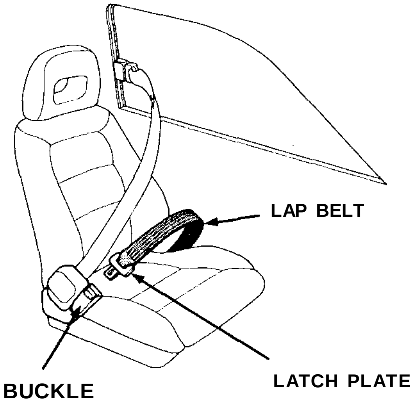
SHOULDER BELT

RETRACTOR

The Seat Belt System and How It Works (US) (cont'd)

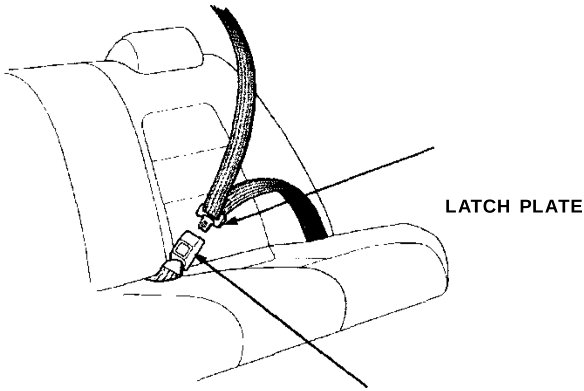
Front Lap Belt

Each front seat has a lap belt that must be latched and unlatched by hand. This belt also has an emergency locking retractor. In normal driving, the retractor lets you move freely in your seat while it keeps some tension on the belt. During a collision or sudden stop, the retractor automatically locks the belt to help restrain your body.



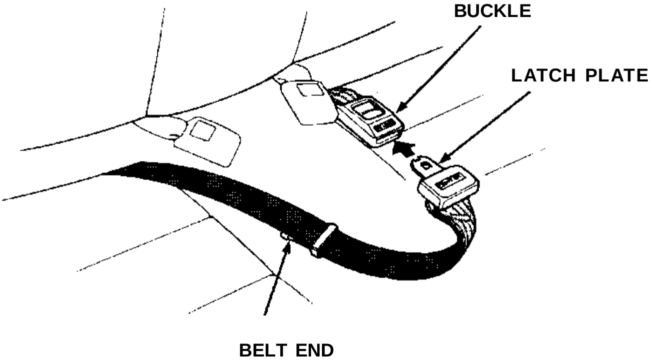
Rear Lap/Shoulder Belt

This style of seat belt has a single belt that goes over your shoulder, across your chest, and across your hips. This belt also has an emergency locking retractor. In normal driving, the retractor lets you move freely in your seat while it keeps some tension on the belt. During a collision or sudden stop, the retractor automatically locks the belt to help restrain your body.



Rear Lap Belt

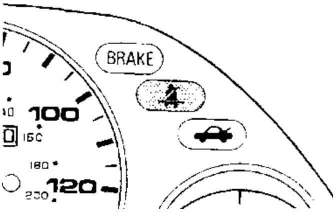
The rear lap belt has one manually-adjusted belt that fits across the hips. It is similar to safety belts used in airplanes.



Seat Belt Indicator Light and Beeper

The seat belt indicator light serves several functions. It comes on when you turn the ignition ON (II). It is a reminder to you and your passengers to protect yourselves by fastening the seat belts. A beeper also sounds if you have not fastened your lap belt. If you do not fasten your lap belt, the beeper will stop after a few seconds but the light will stay on until you latch your belt.

If either shoulder belt is unlatched while the ignition is on, the seat belt indicator light will come on and stay on until the belt is latched. The beeper will also sound for about six seconds.



The seat belt indicator light will flash and the beeper will sound rapidly if a shoulder belt stalls before completing its full forward or rearward travel. If this happens, you can recycle the system by opening and closing the door. If the buckle still does not move to the full forward or rearward position, see the manual operation procedure on page 150. If the problem persists after moving the buckle manually, see your Acura dealer.

If the seat belt indicator light and beeper do not function at all, or if they come on even when the belts are latched and the doors are fully closed, have the system checked immediately by an Acura dealer.

The Seat Belt System and How It Works (US) (cont'd)

Wearing Seat Belts Properly

You can increase the effectiveness of your seat belts if you take a little time to read the following pages and make sure you know how to wear seat belts properly.

⚠WARNING

Not wearing a seat belt properly increases the chance of serious injury or death in a crash.

Be sure you and your passengers always wear seat belts and wear them properly.

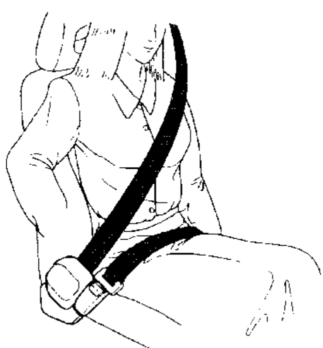
Wearing An Automatic Front Shoulder Belt

1. Enter the car, close the door, then make sure your seat is adjusted forward or backward to a good driving or riding position and your seat-back is upright (see page 27).

2. Before you turn on the ignition, make sure you and a front seat passenger are not in the path of the automatic shoulder belt buckle. Also check to see that the sunvisors are in their normal position above the front windshield.

3. Now turn the ignition ON (II). Your shoulder belt buckle will automatically move along the top of the window to the rearward position, and the length of the belt will automatically adjust to your body. A passenger's shoulder belt will also move to the rearward position.

   If a front door is not completely closed, that shoulder belt will not move. The seat belt indicator light and the door warning light will both come on. Make sure the door is completely closed by opening and closing it again.



4. Check that the belt rests smoothly over your shoulder and across your chest. Do NOT put the belt under your arm. This could increase your chances of serious injury in a crash.

5. Before driving away, don't forget to fasten your lap belt, and make sure a front passenger does the same. For the best protection, the lap belt and automatic shoulder belt should always be worn together.

When you remove the key from the ignition, your shoulder belt will automatically move to the forward position. Your belt will also move to the forward position if you open the door. The front passenger's shoulder belt will move to the forward position when that door is opened.

The automatic shoulder belt is designed to restrain the upper torso in a frontal impact. To reduce the risk of sliding under the shoulder belt and provide more protection in a crash or rollover, the front lap belt must always be worn in conjunction with the automatic shoulder belt.

WARNING

Not wearing a front lap belt in conjunction with an automatic shoulder belt increases the chance of serious injury or death in a crash.

Make sure you and a front passenger always wear BOTH a lap belt and an automatic shoulder belt.

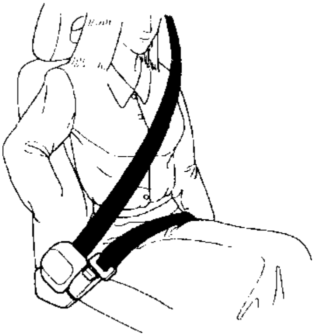
Wearing A Front Lap Belt

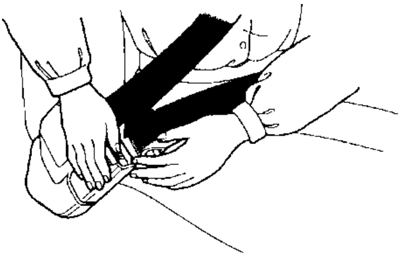
1. With the shoulder belt already in the rearward position (see above), grasp the lap belt latch plate and slowly pull the belt out of its retractor. If the belt locks, let it retract then pull it out again more slowly.

2. After making sure the belt is not twisted, insert the latch plate into the buckle. Then position the belt as low as possible across your hips, not across your stomach. This allows your strong pelvic bones to take the force of a crash.

The Seat Belt System and How It Works (US) (cont'd)

3. Make sure the lap belt and the automatic shoulder belt are not crossed.

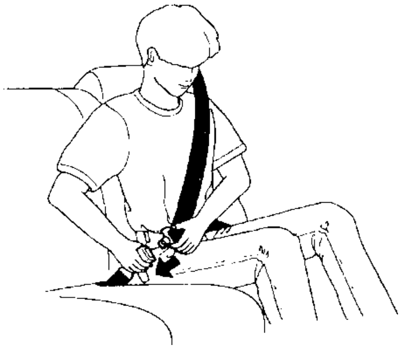




To unlatch a front lap belt, push the red PRESS button on the buckle.

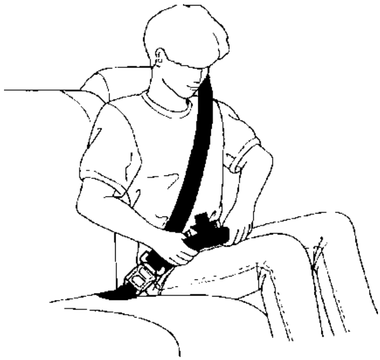
Wearing a Rear Lap/Shoulder Belt

1. Pull the latch plate across your body and insert it into the buckle. Tug on the belt to make sure the latch is securely locked.

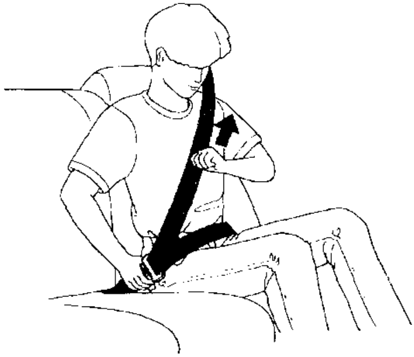


2. Check that the belt is not twisted.

3. Position the lap portion of the belt as low as possible across your hips, not across your stomach. This lets your strong pelvic bones take the force of a crash.

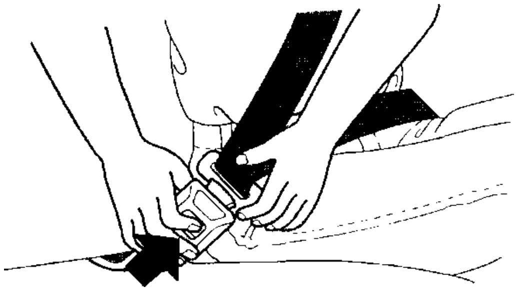


4. Pull up on the shoulder part of the belt to remove any slack. Make sure the belt goes over your collarbone and across your chest.



5. If the shoulder belt crosses your neck, you need to adjust yourseating position.
   Move toward the center of the seat until the belt fits over your collarbone.

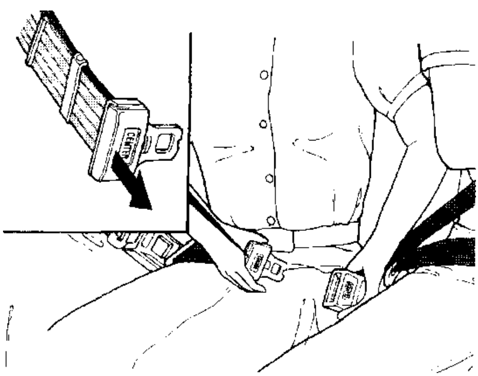
   To unlatch the seat belt, push the red PRESS button on the buckle. Guide the belt across your body to the door pillar. If the belt doesn't retract easily, pull it out and check for twists or kinks.



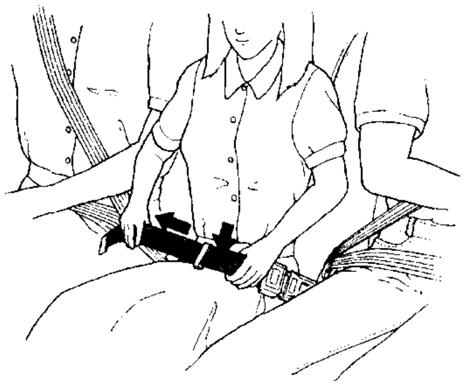
Wearing the Rear Lap Belt

1. Pull the latch plate across your hips and insert it into the buckle marked CENTER.

   If the belt is too short, hold the latch plate at a right angle and pull to extend the belt. Insert the latch plate into the buckle.



2. Position the belt as low as possible across your hips and pelvic bones, not across your stomach. Pull the loose end of the belt to adjust for a snug but comfortable fit.



To unlatch the belt, push the red PRESS button on the buckle.