Electrical Resistive and Electromagnetic Components

Lesley Redden

Guillermo Dutton

First Edition, 2012

ISBN 978-81-323-0868-3

Published by:
Academic Studio
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter 1

Electrical Resistance and Conductance

The **electrical resistance** of an electrical element measures its opposition to the passage of an electric current; the inverse quantity is **electrical conductance**, measuring how easily electricity flows along a certain path. Electrical resistance shares some conceptual parallels with the mechanical notion of friction. The SI unit of electrical resistance is the ohm (Ω), while electrical conductance is measured in siemens (S).

An object of uniform cross section has a resistance proportional to its resistivity and length and inversely proportional to its cross-sectional area. All materials show some resistance, except for superconductors, which have a resistance of zero.

The resistance of an object is defined as the ratio of voltage across it to current through it:

$$R = \frac{V}{I}$$

For a wide variety of materials and conditions, the electrical resistance R is constant for a given temperature; it does not depend on the amount of current through or the potential difference (voltage) across the object. Such materials are called Ohmic materials. For objects made of ohmic materials the definition of the resistance, with R being a constant for that resistor, is known as Ohm's law.

In the case of a nonlinear conductor (not obeying Ohm's law), this ratio can change as current or voltage changes; the inverse slope of a chord to an I–V curve is sometimes referred to as a "chordal resistance" or "static resistance".

Conductors and resistors

A 65-Ω resistor, as identified by its electronic color code (blue–green–black). An ohmmeter could be used to verify this value.

Objects such as wires that are designed to have low resistance so that they transfer current with the least loss of electrical energy are called conductors. Objects that are designed to have a specific resistance so that they can dissipate electrical energy or otherwise modify how a circuit behaves are called resistors. Conductors are made of highly conductive materials such as metals, in particular copper and aluminium. Resistors, on the other hand, are made of a wide variety of materials depending on factors such as the desired resistance, amount of energy that it needs to dissipate, precision, and cost.

DC resistance

The resistance of a given resistor or conductor grows with the length of conductor and decreases for larger cross-sectional area. The resistance R and conductance G of a conductor of uniform cross section, therefore, can be computed as

$$R = \rho \frac{\ell}{A},$$
$$G = \frac{\sigma A}{\ell}.$$

where ℓ is the length of the conductor, measured in metres [m], A is the cross-sectional area of the conductor measured in square metres [m²], and ρ (rho) is the electrical resistivity (also called *specific electrical resistance*) of the material, measured in ohm-metres (Ωm). Resistivity is a measure of the material's ability to oppose electric current. For purely resistive circuits conductance is related to resistance R by:

$$G = \frac{1}{R}$$

For practical reasons, any connections to a real conductor will almost certainly mean the current density is not totally uniform. However, this formula still provides a good approximation for long thin conductors such as wires.

AC resistance

A wire carrying alternating current has a reduced effective cross sectional area because of the skin effect. Adjacent conductors carryng alternating current have a higher resistance than they would in isolation or when carrying direct current, due to the proximity effect. At commercial power frequency, these effects are significant for large conductors carrying large currents, such as busbars in an electrical substation, or large power cables carrying more than a few hundred amperes.

When an alternating current flows through the circuit, its flow is not opposed only by the circuit resistance, but also by the opposition of electric and magnetic fields to the current change. That effect is measured by electrical reactance. The combined effects of reactance and resistance are expressed by electrical impedance.

Measuring resistance

An instrument for measuring resistance is called an ohmmeter. Simple ohmmeters cannot measure low resistances accurately because the resistance of their measuring leads causes a voltage drop that interferes with the measurement, so more accurate devices use four-terminal sensing.

Causes of resistance

In metals

A metal consists of a lattice of atoms, each with a shell of electrons. This is also known as a positive ionic lattice. The outer electrons are free to dissociate from their parent atoms and travel through the lattice, creating a 'sea' of electrons, making the metal a conductor. When an electrical potential difference (a voltage) is applied across the metal, the electrons drift from one end of the conductor to the other under the influence of the electric field.

Near room temperatures, the thermal motion of ions is the primary source of scattering of electrons (due to destructive interference of free electron waves on non-correlating potentials of ions), and is thus the prime cause of metal resistance. Imperfections of lattice also contribute into resistance, although their contribution in pure metals is negligible.

The larger the cross-sectional area of the conductor, the more electrons are available to carry the current, so the lower the resistance. The longer the conductor, the more scattering events occur in each electron's path through the material, so the higher the resistance. Different materials also affect the resistance.

In semiconductors and insulators

In metals, the Fermi level lies in the conduction band giving rise to free conduction electrons. However, in semiconductors the position of the Fermi level is within the band gap, approximately half-way between the conduction band minimum and valence band maximum for intrinsic (undoped) semiconductors. This means that at 0 kelvins, there are no free conduction electrons and the resistance is infinite. However, the resistance will continue to decrease as the charge carrier density in the conduction band increases. In extrinsic (doped) semiconductors, dopant atoms increase the majority charge carrier concentration by donating electrons to the conduction band or accepting holes in the valence band. For both types of donor or acceptor atoms, increasing the dopant density leads to a reduction in the resistance. Highly doped semiconductors hence behave metallic. At very high temperatures, the contribution of thermally generated carriers will dominate over the contribution from dopant atoms and the resistance will decrease exponentially with temperature.

Small-signal device conductances

The term conductance applies to electronic devices such as transistors and diodes, where it usually refers to a small-signal model that is a linearization of the underlying device equations

about a selected DC operating point or Q-point. This conductance is the reciprocal of the small-signal device resistance.

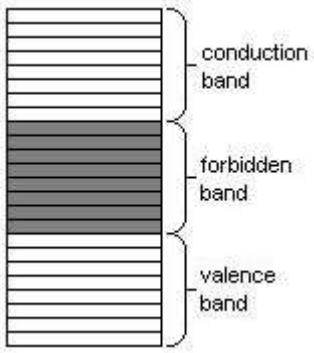
In ionic liquids/electrolytes

In electrolytes, electrical conduction happens not by band electrons or holes, but by full atomic species (ions) traveling, each carrying an electrical charge. The resistivity of ionic liquids varies tremendously by the concentration - while distilled water is almost an insulator, salt water is a very efficient electrical conductor. In biological membranes, currents are carried by ionic salts. Small holes in the membranes, called ion channels, are selective to specific ions and determine the membrane resistance.

Resistivity of various materials

Material	**Resistivity**, ρ ohm-metre
Metals	10^{-8}
Semiconductors	variable
Electrolytes	variable
Insulators	10^{16}
Superconductors	0 (exactly)

Band theory simplified



Electron energy levels in an insulator

Quantum mechanics states that the energy of an electron in an atom cannot be any arbitrary value. Rather, there are fixed energy levels which the electrons can occupy, and values in between these levels are impossible. The energy levels are grouped into two bands: the **valence band** and the **conduction band** (the latter is generally above the former). Electrons in the conduction band may move freely throughout the substance in the presence of an electrical field.

In insulators and semiconductors, the atoms in the substance influence each other so that between the valence band and the conduction band there exists a forbidden band of energy levels, which the electrons cannot occupy. In order for a current to flow, a relatively large

amount of energy must be furnished to an electron for it to leap across this forbidden gap and into the conduction band. Thus, even large voltages can yield relatively small currents.

Differential resistance

When the current–voltage dependence is not linear, **differential resistance**, **incremental resistance** or **slope resistance** is defined as the slope of the *V*-*I* graph at a particular point, thus:

$$R = \frac{\mathrm{d}V}{\mathrm{d}I}$$

This quantity is sometimes called simply *resistance*, although the two definitions are equivalent only for an ohmic component such as an ideal resistor. For example, a diode is a circuit element for which the resistance depends on the applied voltage or current.

If the *V*-*I* graph is not monotonic (i.e. it has a peak or a trough), the differential resistance will be negative for some values of voltage and current. This property is often known as *negative resistance*, although it is more correctly called *negative differential resistance*, since the absolute resistance *V*/*I* is still positive. An example of such an element is the tunnel diode.

Differential resistance is only useful to compare a nonlinear device with a linear source/load in some small interval; for example if it is necessary to evaluate a zener diode's voltage stability under different current values.

Temperature dependence

Near room temperature, the electric resistance of a typical metal increases linearly with rising temperature, while the electrical resistance of a typical semiconductor decreases with rising temperature. The amount of that change in resistance can be calculated using the temperature coefficient of resistivity of the material using the following formula:

$$R(T) = R_0[1 + \alpha(T - T_0)]$$

where T is its temperature, T_0 is a reference temperature (usually room temperature), R_0 is the resistance at T_0, and α is the percentage change in resistivity per unit temperature. The constant α depends only on the material being considered. The relationship stated is actually only an approximate one, the true physics being somewhat non-linear, or looking at it another way, α itself varies with temperature. For this reason it is usual to specify the temperature that α was measured at with a suffix, such as α_{15} and the relationship only holds in a range of temperatures around the reference.

At lower temperatures (less than the Debye temperature), the resistance of a metal decreases as T^5 due to the electrons scattering off of phonons. At even lower temperatures, the dominant scattering mechanism for electrons is other electrons, and the resistance decreases as T^2. At some point, the impurities in the metal will dominate the behavior of the electrical resistance which causes it to saturate to a constant value. Matthiessen's Rule (first formulated by Augustus

Matthiessen in the 1860s; the equation below gives its modern form) says that all of these different behaviors can be summed up to get the total resistance as a function of temperature,

$$R = R_{\mathrm{imp}} + aT^2 + bT^5 + cT$$

where R_{imp} is the temperature independent electrical resistivity due to impurities, and a, b, and c are coefficients which depend upon the metal's properties. This rule can be seen as the motivation to Heike Kamerlingh Onnes's experiments that led in 1911 to discovery of superconductivity.

Intrinsic semiconductors become better conductors as the temperature increases; the electrons are bumped to the conduction energy band by thermal energy, where they flow freely and in doing so leave behind holes in the valence band which also flow freely. The electric resistance of a typical intrinsic (non doped) semiconductor decreases exponentially with the temperature:

$$R = R_0 e^{-aT}$$

Extrinsic (doped) semiconductors have a far more complicated temperature profile. As temperature increases starting from absolute zero they first decrease steeply in resistance as the carriers leave the donors or acceptors. After most of the donors or acceptors have lost their carriers the resistance starts to increase again slightly due to the reducing mobility of carriers (much as in a metal). At higher temperatures it will behave like intrinsic semiconductors as the carriers from the donors/acceptors become insignificant compared to the thermally generated carriers.

The electric resistance of electrolytes and insulators is highly nonlinear, and case by case dependent, therefore no generalized equations are given.

Strain dependence

Just as the resistance of a conductor depends upon temperature, the resistance of a conductor depends upon strain. By placing a conductor under tension (a form of stress that leads to strain in the form of stretching of the conductor), the length of the section of conductor under tension increases and its cross-sectional area decreases. Both these effects contribute to increasing the resistance of the strained section of conductor. Under compression (strain in the opposite direction), the resistance of the strained section of conductor decreases.

Chapter 2

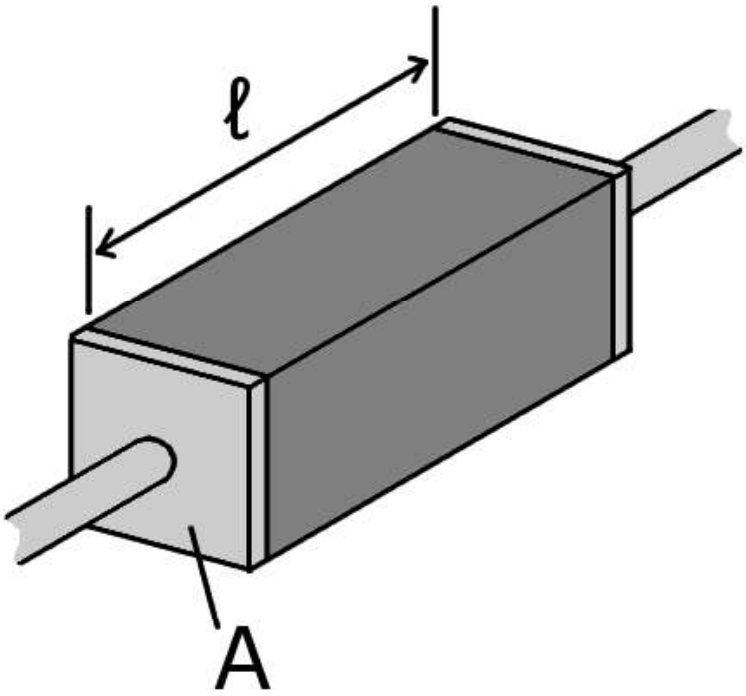
Electrical Resistivity and Conductivity

Electrical resistivity (also known as **resistivity**, **specific electrical resistance**, or **volume resistivity**) is a measure of how strongly a material opposes the flow of electric current. A low resistivity indicates a material that readily allows the movement of electric charge. The SI unit of electrical resistivity is the ohm metre [Ωm]. It is commonly represented by the Greek letter ρ (rho).

Electrical conductivity or **specific conductance** is the reciprocal quantity, and measures a material's ability to conduct an electric current. It is commonly represented by the Greek letter σ, but κ (esp. in electrical engineering) or γ are also occasionally used. Its SI unit is siemens per metre ($S \cdot m^{-1}$) and CGSE unit is reciprocal second (s^{-1}):

$$\sigma = \frac{1}{\rho}.$$

Definitions



A piece of resistive material with electrical contacts on both ends.

Electrical resistivity ρ (Greek: rho) is defined by,

$$\rho = \frac{E}{J}$$

where

ρ is the static resistivity (measured in ohm-metres, Ω-m)
E is the magnitude of the electric field (measured in volts per metre, V/m);
J is the magnitude of the current density (measured in amperes per square metre, A/m²).

Many resistors and conductors have a uniform cross section with a uniform flow of electric current and are made of one material. In this case, the above definition of ρ leads to:

$$\rho = R\frac{A}{\ell},$$

where

R is the electrical resistance of a uniform specimen of the material (measured in ohms, Ω)
ℓ is the length of the piece of material (measured in metres, m)
A is the cross-sectional area of the specimen (measured in square metres, m²).

The reason resistivity has the dimension units of ohm-metres can be seen by transposing the definition to make resistance the subject:

$$R = \rho \frac{\ell}{A}$$

The resistance of a given sample will increase with the length, but decrease with greater cross-sectional area. Resistance is measured in ohms. Length over area has units of 1/distance. To end up with ohms, resistivity must be in the units of "ohms × distance" (SI ohm-metre, US ohm-inch).

In a hydraulic analogy, increasing the cross-sectional area of a pipe reduces its resistance to flow, and increasing the length increases resistance to flow (and pressure drop for a given flow).

Resistivity of various materials

- A conductor such as a metal has high conductivity and a low resistivity.
- An insulator like glass has low conductivity and a high resistivity.
- The conductivity of a semiconductor is generally intermediate, but varies widely under different conditions, such as exposure of the material to electric fields or specific frequencies of light, and, most important, with temperature and composition of the semiconductor material.

The degree of doping in semiconductors makes a large difference in conductivity. To a point, more doping leads to higher conductivity. The conductivity of a solution of water is highly dependent on its concentration of dissolved salts, and other chemical species that ionize in the solution. Electrical conductivity of water samples is used as an indicator of how salt-free, ion-free, or impurity-free the sample is; the purer the water, the lower the conductivity (the higher the resistivity). Conductivity measurements in water are often reported as *specific conductance*, the conductivity of the water at 25 °C. An EC meter is normally used to measure conductivity in a solution.

This table shows the resistivity, conductivity and temperature coefficient of various materials at 20 °C (68 °F)

Material	**ρ [Ω·m] at 20 °C**	**σ [S/m] at 20 °C**	**Temperature coefficient [K^{-1}]**
Silver	1.59×10^{-8}	6.30×10^{7}	0.0038
Copper	1.68×10^{-8}	5.96×10^{7}	0.0039
Annealed Copper		5.80×10^{7}	
Gold	2.44×10^{-8}	4.52×10^{7}	0.0034
Aluminium	2.82×10^{-8}	3.5×10^{7}	0.0039
Calcium	$3.36\text{x}10^{-8}$		0.0041

Tungsten	5.60×10^{-8}		0.0045
Zinc	5.90×10^{-8}		0.0037
Nickel	6.99×10^{-8}		0.006
Iron	1.0×10^{-7}		0.005
Platinum	1.06×10^{-7}		0.00392
Tin	1.09×10^{-7}		0.0045
Lead	2.2×10^{-7}		0.0039
Titanium	$4.20x10^{-7}$		X
Manganin	4.82×10^{-7}		0.000002
Constantan	4.9×10^{-7}		0.000008
Mercury	9.8×10^{-7}		0.0009
Nichrome	1.10×10^{-6}		0.0004
Carbon (amorphous)	$5\text{-}8\times10^{-4}$		−0.0005
Carbon (graphite)	$2.5\text{-}5.0\times10^{-6}$ ⊥ basal plane 3.0×10^{-3} // basal plane		
Carbon (diamond)	$\sim10^{12}$		
Germanium	4.6×10^{-1}		−0.048
Sea water	2×10^{-1}	4.8	
Drinking water		0.0005 to 0.05	
Deionized water		5.5×10^{-6}	
Silicon	6.40×10^{2}		−0.075
Glass	10^{10} to 10^{14}		?
Hard rubber	approx. 10^{13}		?
Sulfur	10^{15}		?
Air		3 to 8×10^{-15}	
Paraffin	10^{17}		?
Quartz (fused)	7.5×10^{17}		?
PET	10^{20}		?
Teflon	10^{22} to 10^{24}		?

The effective temperature coefficient varies with temperature and purity level of the material. The 20 °C value is only an approximation when used at other temperatures. For example, the coefficient becomes lower at higher temperatures for copper, and the value 0.00427 is commonly specified at 0 °C.

The extremely low resistivity (high conductivity) of silver is characteristic of metals. George Gamow tidily summed up the nature of the metals' dealings with electrons in his science-

popularizing book, *One, Two, Three...Infinity* (1947): "The metallic substances differ from all other materials by the fact that the outer shells of their atoms are bound rather loosely, and often let one of their electrons go free. Thus the interior of a metal is filled up with a large number of unattached electrons that travel aimlessly around like a crowd of displaced persons. When a metal wire is subjected to electric force applied on its opposite ends, these free electrons rush in the direction of the force, thus forming what we call an electric current." More technically, the free electron model gives a basic description of electron flow in metals.

Temperature dependence

In general, electrical resistivity of metals increases with temperature, while the resistivity of semiconductors decreases with increasing temperature. In both cases, electron–phonon interactions can play a key role. At high temperatures, the resistance of a metal increases linearly with temperature. As the temperature of a metal is reduced, the temperature dependence of resistivity follows a power law function of temperature. Mathematically the temperature dependence of the resistivity ρ of a metal is given by the Bloch–Grüneisen formula:

$$\rho(T) = \rho(0) + A\left(\frac{T}{\Theta_R}\right)^n \int_0^{\frac{\Theta_R}{T}} \frac{x^n}{(e^x - 1)(1 - e^{-x})} dx$$

where ρ(0) is the residual resistivity due to defect scattering, A is a constant that depends on the velocity of electrons at the Fermi surface, the Debye radius and the number density of electrons in the metal. Θ_R is the Debye temperature as obtained from resistivity measurements and matches very closely with the values of Debye temperature obtained from specific heat measurements. n is an integer that depends upon the nature of interaction:

1. n=5 implies that the resistance is due to scattering of electrons by phonons (as it is for simple metals)
2. n=3 implies that the resistance is due to s-d electron scattering (as is the case for transition metals)
3. n=2 implies that the resistance is due to electron–electron interaction.

As the temperature of the metal is sufficiently reduced (so as to 'freeze' all the phonons), the resistivity usually reaches a constant value, known as the **residual resistivity**. This value depends not only on the type of metal, but on its purity and thermal history. The value of the residual resistivity of a metal is decided by its impurity concentration. Some materials lose all electrical resistivity at sufficiently low temperatures, due to an effect known as superconductivity.

An even better approximation of the temperature dependence of the resistivity of a semiconductor is given by the Steinhart–Hart equation:

$$1/T = A + B\ln(\rho) + C(\ln(\rho))^3$$

where *A*, *B* and *C* are the so-called **Steinhart–Hart coefficients**.

This equation is used to calibrate thermistors.

In non-crystalline semi-conductors, conduction can occur by charges quantum tunnelling from one localised site to another. This is known as variable range hopping and has the characteristic form of $\rho = Ae^{T^{-1/n}}$, where n=2,3,4 depending on the dimensionality of the system.

Complex resistivity and conductivity

When analyzing the response of materials to alternating electric fields, in applications such as electrical impedance tomography, it is necessary to replace resistivity with a complex quantity called **impeditivity** (in analogy to electrical impedance). Impeditivity is the sum of a real component, the resistivity, and an imaginary component, the **reactivity** (in analogy to reactance). The magnitude of Impeditivity is the square root of sum of squares of magnitudes of resistivity and reactivity.

Conversely, in such cases the conductivity must be expressed as a complex number (or even as a matrix of complex numbers, in the case of anisotropic materials) called the *admittivity*. Admittivity is the sum of a real component called the conductivity and an imaginary component called the susceptivity.

An alternative description of the response to alternating currents uses a real (but frequency-dependent) conductivity, along with a real permittivity. The larger the conductivity is, the more quickly the alternating-current signal is absorbed by the material (i.e., the more opaque the material is).

Resistivity density products

In some applications where the weight of an item is very important resistivity density products are more important than absolute low resistivity- it is often possible to make the conductor thicker to make up for a higher resistivity; and then a low resistivity density product material (or equivalently a high conductance to density ratio) is desirable. For example, for long distance overhead power lines— aluminium is frequently used rather than copper because it is lighter for the same conductance.

Material	Resistivity [nΩ·m]	Density [g/cm³]	Resistivity-density product [nΩ·m·g/cm³]
Sodium	47.7	0.97	46
Lithium	92.8	0.53	49
Calcium	33.6	1.55	52
Potassium	72.0	0.89	64
Aluminium	26.50	2.70	72
Copper	16.78	8.96	150
Silver	15.87	10.49	166

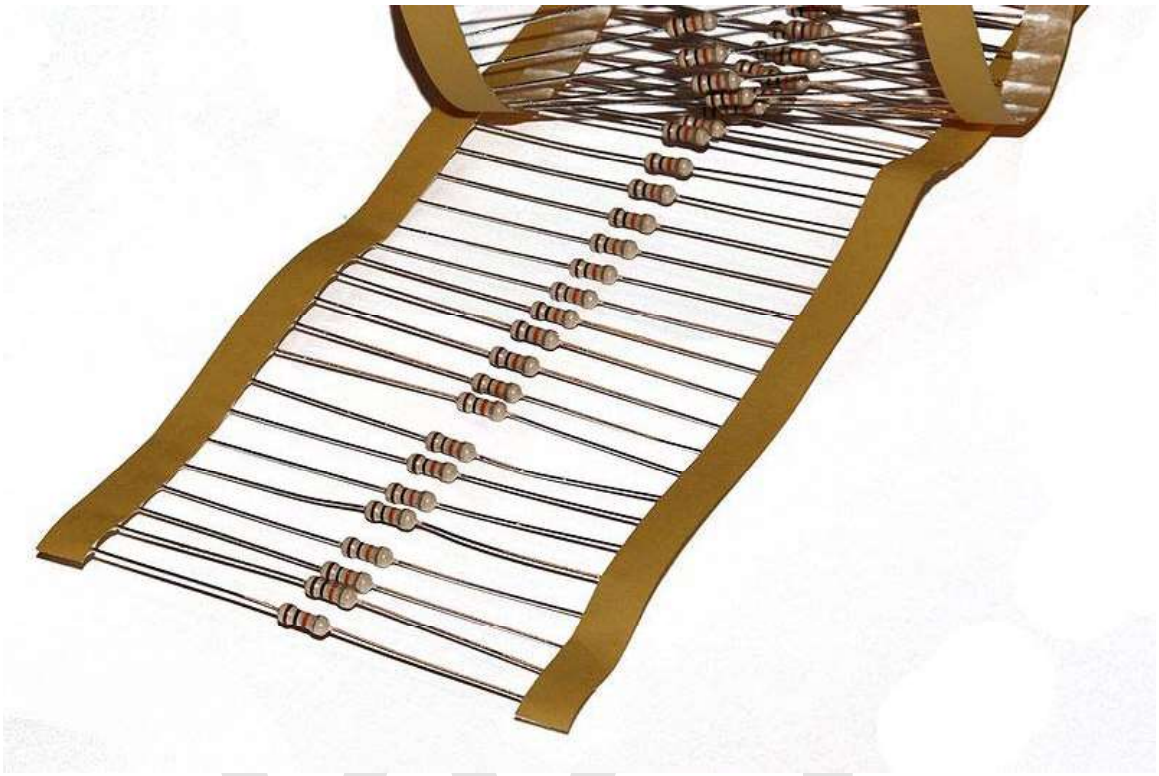
Silver, although it is the least resistive metal known, has a high density and does poorly by this measure. The calcium and the alkali metals have the best products, but are rarely used for conductors due to their high reactivity with water and oxygen. Aluminium is far more stable.

Chapter 3

Resistor

A typical axial-lead resistor

Partially exposed Tesla TR-212 1 kΩ carbon film resistor



Axial-lead resistors on tape. The tape is removed during assembly before the leads are formed and the part is inserted into the board.

Three carbon composition resistors in a 1960s valve (vacuum tube) radio

A **resistor** is a two-terminal passive electronic component which implements electrical resistance as a circuit element. When a voltage V is applied across the terminals of a resistor, a current I will flow through the resistor in direct proportion to that voltage. The reciprocal of the constant of proportionality is known as the resistance R, since, with a given voltage V, a larger value of R further "resists" the flow of current I as given by Ohm's law:

$$I = \frac{V}{R}$$

Resistors are common elements of electrical networks and electronic circuits and are ubiquitous in most electronic equipment. Practical resistors can be made of various compounds and films, as well as resistance wire (wire made of a high-resistivity alloy, such as nickel-chrome). Resistors are also implemented within integrated circuits, particularly analog devices, and can also be integrated into hybrid and printed circuits.

The electrical functionality of a resistor is specified by its resistance: common commercial resistors are manufactured over a range of more than 9 orders of magnitude. When specifying that resistance in an electronic design, the required precision of the resistance may require attention to the manufacturing tolerance of the chosen resistor, according to its specific application. The temperature coefficient of the resistance may also be of concern in some precision applications. Practical resistors are also specified as having a maximum power rating which must exceed the anticipated power dissipation of that resistor in a particular circuit: this is mainly of concern in power electronics applications. Resistors with higher power ratings are physically larger and may require heat sinking. In a high voltage circuit, attention must sometimes be paid to the rated maximum working voltage of the resistor.

The series inductance of a practical resistor causes its behavior to depart from ohms law; this specification can be important in some high-frequency applications for smaller values of resistance. In a low-noise amplifier or pre-amp the noise characteristics of a resistor may be an issue. The unwanted inductance, excess noise, and temperature coefficient are mainly dependent on the technology used in manufacturing the resistor. They are not normally specified individually for a particular family of resistors manufactured using a particular technology. A family of discrete resistors is also characterized according to its form factor, that is, the size of the device and position of its leads (or terminals) which is relevant in the practical manufacturing of circuits using them.

Units

The ohm (symbol: Ω) is the SI unit of electrical resistance, named after Georg Simon Ohm. An ohm is equivalent to a volt per ampere. Since resistors are specified and manufactured over a very large range of values, the derived units of milliohm (1 mΩ = 10^{-3} Ω), kilohm (1 kΩ = 10^{3} Ω), and megohm (1 MΩ = 10^{6} Ω) are also in common usage.

The reciprocal of resistance R is called conductance G = 1/R and is measured in Siemens (SI unit), sometimes referred to as a mho. Thus a Siemens is the reciprocal of an ohm: $S = \Omega^{-1}$.

Although the concept of conductance is often used in circuit analysis, practical resistors are always specified in terms of their resistance (ohms) rather than conductance.

Theory of operation

Ohm's law

The behavior of an ideal resistor is dictated by the relationship specified in Ohm's law:

$$V = I \cdot R$$

Ohm's law states that the voltage (V) across a resistor is proportional to the current (I) passing through it, where the constant of proportionality is the resistance (R).

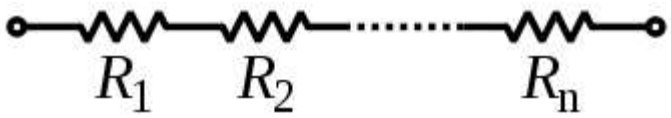
Equivalently, Ohm's law can be stated:

$$I = \frac{V}{R}$$

This formulation of Ohm's law states that, when a voltage (V) is present across a resistance (R), a current (I) will flow through the resistance. This is directly used in practical computations. For example, if a 300 ohm resistor is attached across the terminals of a 12 volt battery, then a current of 12 / 300 = 0.04 amperes (or 40 milliamperes) will flow through that resistor.

Series and parallel resistors

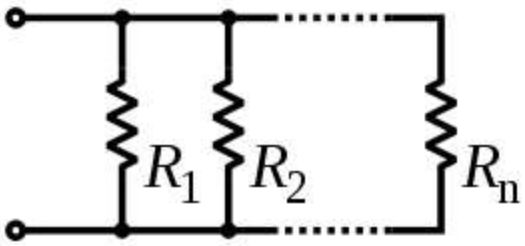
In a series configuration, the current through all of the resistors is the same, but the voltage across each resistor will be in proportion to its resistance. The potential difference (voltage) seen across the network is the sum of those voltages, thus the total resistance can be found as the sum of those resistances:



$$R_{\text{eq}} = R_1 + R_2 + \cdots + R_n$$

As a special case, the resistance of N resistors connected in series, each of the same resistance R, is given by NR.

Resistors in a parallel configuration are each subject to the same potential difference (voltage), however the currents through them add. The conductances of the resistors then add to determine the conductance of the network. Thus the equivalent resistance (R_{eq}) of the network can be computed:



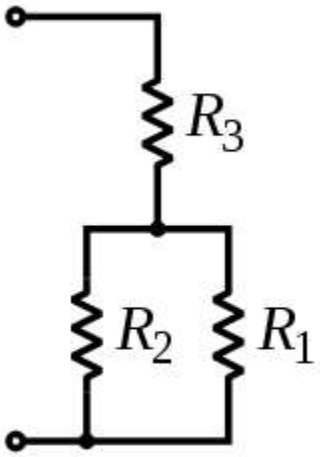
$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}$$

The parallel equivalent resistance can be represented in equations by two vertical lines "||" (as in geometry) as a simplified notation. For the case of two resistors in parallel, this can be calculated using:

$$R_{\text{eq}} = R_1 \| R_2 = \frac{R_1 R_2}{R_1 + R_2}$$

As a special case, the resistance of N resistors connected in parallel, each of the same resistance R, is given by R/N.

A resistor network that is a combination of parallel and series connections can be broken up into smaller parts that are either one or the other. For instance,



$$R_{\text{eq}} = (R_1 \| R_2) + R_3 = \frac{R_1 R_2}{R_1 + R_2} + R_3$$

However, some complex networks of resistors cannot be resolved in this manner, requiring more sophisticated circuit analysis. For instance, consider a cube, each edge of which has been replaced by a resistor. What then is the resistance that would be measured between two opposite vertices? In the case of 12 equivalent resistors, it can be shown that the corner-to-corner resistance is $^5/_6$ of the individual resistance. More generally, the Y-Δ transform, or matrix methods can be used to solve such a problem.

One practical application of these relationships is that a non-standard value of resistance can generally be synthesized by connecting a number of standard values in series and/or parallel. This can also be used to obtain a resistance with a higher power rating than that of the individual resistors used. In the special case of N identical resistors all connected in series or all connected in parallel, the power rating of the individual resistors is thereby multiplied by N.

Power dissipation

The power P dissipated by a resistor (or the equivalent resistance of a resistor network) is calculated as: $P = I^2R = IV = \frac{V^2}{R}$

The first form is a restatement of Joule's first law. Using Ohm's law, the two other forms can be derived.

The total amount of heat energy released over a period of time can be determined from the integral of the power over that period of time:

$$W = \int_{t_1}^{t_2} v(t)i(t)\,dt.$$

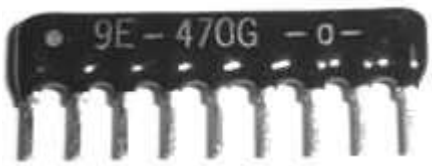
Practical resistors are rated according to their maximum power dissipation. The vast majority of resistors used in electronic circuits absorb much less than a watt of electrical power and require no attention to their power rating. Such resistors in their discrete form, including most of the packages detailed below, are typically rated as 1/10, 1/8, or 1/4 watt.

Resistors required to dissipate substantial amounts of power, particularly used in power supplies, power conversion circuits, and power amplifiers, are generally referred to as *power resistors*; this designation is loosely applied to resistors with power ratings of 1 watt or greater. Power resistors are physically larger and tend not to use the preferred values, color codes, and external packages described below.

If the average power dissipated by a resistor is more than its power rating, damage to the resistor may occur, permanently altering its resistance; this is distinct from the reversible change in resistance due to its temperature coefficient when it warms. Excessive power dissipation may raise the temperature of the resistor to a point where it can burn the circuit board or adjacent components, or even cause a fire. There are flameproof resistors that fail (open circuit) before they overheat dangerously.

Note that the nominal power rating of a resistor is not the same as the power that it can safely dissipate in practical use. Air circulation and proximity to a circuit board, ambient temperature, and other factors can reduce acceptable dissipation significantly. Rated power dissipation may be given for an ambient temperature of 25 °C in free air. Inside an equipment case at 60 °C, rated dissipation will be significantly less; a resistor dissipating a bit less than the maximum figure given by the manufacturer may still be outside the safe operating area and may prematurely fail.

Construction



A single in line (SIL) resistor package with 8 individual, 47 ohm resistors. One end of each resistor is connected to a separate pin and the other ends are all connected together to the remaining (common) pin - pin 1, at the end identified by the white dot.

Lead arrangements

Resistors with wire leads for through-hole mounting

Through-hole components typically have leads leaving the body axially. Others have leads coming off their body radially instead of parallel to the resistor axis. Other components may be SMT (surface mount technology) while high power resistors may have one of their leads designed into the heat sink.

Carbon composition

Carbon composition resistors consist of a solid cylindrical resistive element with embedded wire leads or metal end caps to which the lead wires are attached. The body of the resistor is protected with paint or plastic. Early 20th-century carbon composition resistors had uninsulated bodies; the lead wires were wrapped around the ends of the resistance element rod and soldered. The completed resistor was painted for color coding of its value.

The resistive element is made from a mixture of finely ground (powdered) carbon and an insulating material (usually ceramic). A resin holds the mixture together. The resistance is determined by the ratio of the fill material (the powdered ceramic) to the carbon. Higher concentrations of carbon, a weak conductor, result in lower resistance. Carbon composition resistors were commonly used in the 1960s and earlier, but are not so popular for general use now as other types have better specifications, such as tolerance, voltage dependence, and stress (carbon composition resistors will change value when stressed with over-voltages). Moreover, if internal moisture content (from exposure for some length of time to a humid environment) is significant, soldering heat will create a non-reversible change in resistance value. Carbon composition resistors have poor stability with time and were consequently factory sorted to, at best, only 5% tolerance. These resistors, however, if never subjected to overvoltage nor overheating were remarkably reliable considering the component's size

They are still available, but comparatively quite costly. Values ranged from fractions of an ohm to 22 megohms. Because of the high price, these resistors are no longer used in most applications. However, carbon resistors are used in power supplies and welding controls.

Carbon film

A carbon film is deposited on an insulating substrate, and a helix cut in it to create a long, narrow resistive path. Varying shapes, coupled with the resistivity of carbon, (ranging from 90 to 400 nΩ m) can provide a variety of resistances. Carbon film resistors feature a power rating range of 0.125 W to 5 W at 70 °C. Resistances available range from 1 ohm to 10 megohm. The carbon film resistor has an operating temperature range of -55 °C to 155 °C. It has 200 to 600 volts maximum working voltage range. Special carbon film resistors are used in applications requiring high pulse stability.

Thick and thin film

Thick film resistors became popular during the 1970s, and most SMD (surface mount device) resistors today are of this type. The principal difference between thin film and thick film resistors is not the actual thickness of the film, but rather how the film is applied to the cylinder (axial resistors) or the surface (SMD resistors).

Thin film resistors are made by sputtering (a method of vacuum deposition) the resistive material onto an insulating substrate. The film is then etched in a similar manner to the old (subtractive) process for making printed circuit boards; that is, the surface is coated with a photo-sensitive material, then covered by a pattern film, irradiated with ultraviolet light, and then the exposed photo-sensitive coating is developed, and underlying thin film is etched away.

Thick film resistors are manufactured using screen and stencil printing processes.

Because the time during which the sputtering is performed can be controlled, the thickness of the thin film can be accurately controlled. The type of material is also usually different consisting of one or more ceramic (cermet) conductors such as tantalum nitride (TaN), ruthenium dioxide (RuO_2), lead oxide (PbO), bismuth ruthenate ($Bi_2Ru_2O_7$), nickel chromium (NiCr), and/or bismuth iridate ($Bi_2Ir_2O_7$).

The resistance of both thin and thick film resistors after manufacture is not highly accurate; they are usually trimmed to an accurate value by abrasive or laser trimming. Thin film resistors are usually specified with tolerances of 0.1, 0.2, 0.5, or 1%, and with temperature coefficients of 5 to 25 ppm/K.

Thick film resistors may use the same conductive ceramics, but they are mixed with sintered (powdered) glass and some kind of liquid so that the composite can be screen-printed. This composite of glass and conductive ceramic (cermet) material is then fused (baked) in an oven at about 850 °C.

Thick film resistors, when first manufactured, had tolerances of 5%, but standard tolerances have improved to 2% or 1% in the last few decades. Temperature coefficients of thick film resistors are high, typically ±200 or ±250 ppm/K; a 40 kelvin (70 °F) temperature change can change the resistance by 1%.

Thin film resistors are usually far more expensive than thick film resistors. For example, SMD thin film resistors, with 0.5% tolerances, and with 25 ppm/K temperature coefficients, when bought in full size reel quantities, are about twice the cost of 1%, 250 ppm/K thick film resistors.

Metal film

A common type of axial resistor today is referred to as a metal-film resistor. Metal electrode leadless face (MELF) resistors often use the same technology, but are a cylindrically shaped resistor designed for surface mounting. Note that other types of resistors (e.g., carbon composition) are also available in MELF packages.

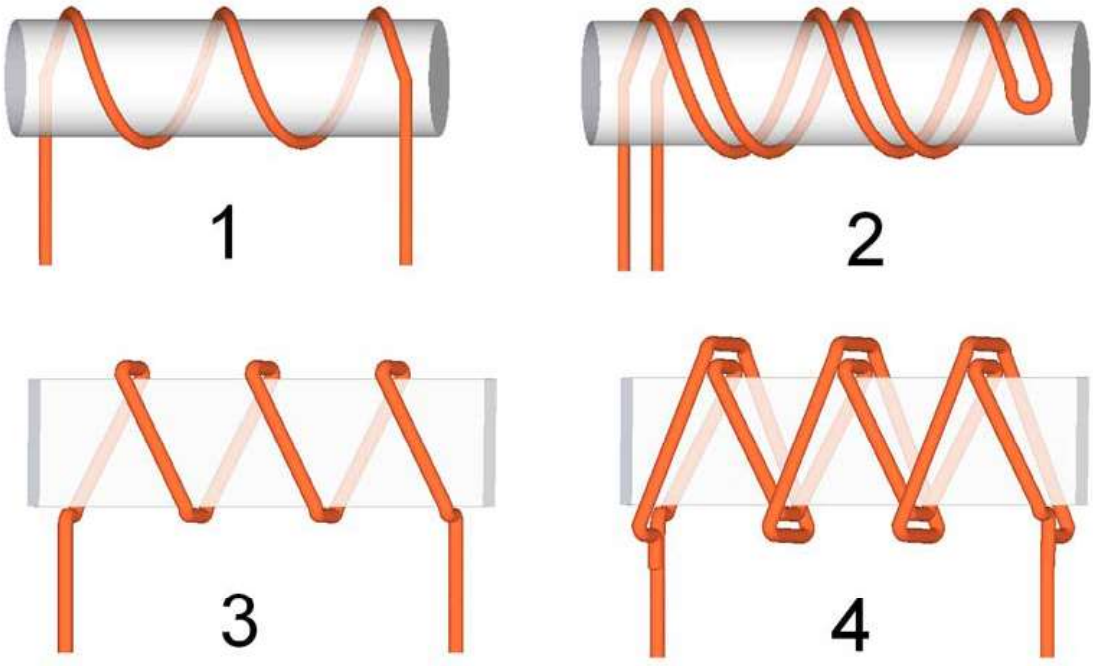
Metal film resistors are usually coated with nickel chromium (NiCr), but might be coated with any of the cermet materials listed above for thin film resistors. Unlike thin film resistors, the material may be applied using different techniques than sputtering (though that is one such technique). Also, unlike thin-film resistors, the resistance value is determined by cutting a helix through the coating rather than by etching. (This is similar to the way carbon resistors are

made.) The result is a reasonable tolerance (0.5, 1, or 2%) and a temperature coefficient that is generally between 50 and 100 ppm/K. Metal film resistors possess good noise characteristics and low non-linearity due to a low voltage coefficient. Also beneficial are the components efficient tolerance, temperature coefficient and stability.

Metal Oxide film

Metal-Oxide film resistors resemble Metal film types, but are made of metal oxides such as tin oxide. This results in a higher operating temperature and greater stability/reliability than Metal film. They are used in applications with high endurance demands.

Wirewound



Types of windings in wire resistors:
1 - common
2 - bifilar
3 - common on a thin former
4 - Ayrton-Perry

Wirewound resistors are commonly made by winding a metal wire, usually nichrome, around a ceramic, plastic, or fiberglass core. The ends of the wire are soldered or welded to two caps or rings, attached to the ends of the core. The assembly is protected with a layer of paint, molded plastic, or an enamel coating baked at high temperature. Because of the very high surface temperature these resistors can withstand temperatures of up to +450 °C. Wire leads in low power wirewound resistors are usually between 0.6 and 0.8 mm in diameter and tinned for ease of soldering. For higher power wirewound resistors, either a ceramic outer case or an aluminum outer case on top of an insulating layer is used. The aluminum-cased types are designed to be

attached to a heat sink to dissipate the heat; the rated power is dependent on being used with a suitable heat sink, e.g., a 50 W power rated resistor will overheat at a fraction of the power dissipation if not used with a heat sink. Large wirewound resistors may be rated for 1,000 watts or more.

Because wirewound resistors are coils they have more undesirable inductance than other types of resistor, although winding the wire in sections with alternately reversed direction can minimize inductance. Other techniques employ bifilar winding, or a flat thin former (to reduce cross-section area of the coil). For most demanding circuits resistors with Ayrton-Perry winding are used.

Applications of wirewound resistors are similar to those of composition resistors with the exception of the high frequency. The high frequency of wirewound resistors is substantially worse than that of a composition resistor

Foil resistor

The primary resistance element of a foil resistor is a special alloy foil several micrometres thick. Since their introduction in the 1960s, foil resistors have had the best precision and stability of any resistor available. One of the important parameters influencing stability is the temperature coefficient of resistance (TCR). The TCR of foil resistors is extremely low, and has been further improved over the years. One range of ultra-precision foil resistors offers a TCR of 0.14 ppm/°C, tolerance ±0.005%, long-term stability (1 year) 25 ppm, (3 year) 50 ppm (further improved 5-fold by hermetic sealing), stability under load (2000 hours) 0.03%, thermal EMF 0.1 μV/°C, noise -42 dB, voltage coefficient 0.1 ppm/V, inductance 0.08 μH, capacitance 0.5 pF.

Ammeter shunts

An ammeter shunt is a special type of current-sensing resistor, having four terminals and a value in milliohms or even micro-ohms. Current-measuring instruments, by themselves, can usually accept only limited currents. To measure high currents, the current passes through the shunt, where the voltage drop is measured and interpreted as current. A typical shunt consists of two solid metal blocks, sometimes brass, mounted on to an insulating base. Between the blocks, and soldered or brazed to them, are one or more strips of low temperature coefficient of resistance (TCR) manganin alloy. Large bolts threaded into the blocks make the current connections, while much-smaller screws provide voltage connections. Shunts are rated by full-scale current, and often have a voltage drop of 50 mV at rated current. Such meters are adapted to the shunt full current rating by using an appropriately marked dial face; no change need be made to the other parts of the meter.

Grid resistor

In heavy-duty industrial high-current applications, a grid resistor is a large convection-cooled lattice of stamped metal alloy strips connected in rows between two electrodes. Such industrial grade resistors can be as large as a refrigerator; some designs can handle over 500 amperes of

current, with a range of resistances extending lower than 0.04 ohms. They are used in applications such as dynamic braking and load banking for locomotives and trams, neutral grounding for industrial AC distribution, control loads for cranes and heavy equipment, load testing of generators and harmonic filtering for electric substations.

The term *grid resistor* is sometimes used to describe a resistor of any type connected to the control grid of a vacuum tube. This is not a resistor technology; it is an electronic circuit topology.

Special varieties

- Metal oxide varistor
- Cermet
- Phenolic
- Tantalum
- Water resistor

Variable resistors

Adjustable resistors

A resistor may have one or more fixed tapping points so that the resistance can be changed by moving the connecting wires to different terminals. Some wirewound power resistors have a tapping point that can slide along the resistance element, allowing a larger or smaller part of the resistance to be used.

Where continuous adjustment of the resistance value during operation of equipment is required, the sliding resistance tap can be connected to a knob accessible to an operator. Such a device is called a rheostat and has two terminals.

Potentiometers

A common element in electronic devices is a three-terminal resistor with a continuously adjustable tapping point controlled by rotation of a shaft or knob. These variable resistors are known as potentiometers when all three terminals are present, since they act as a continuously adjustable voltage divider. A common example is a volume control for a radio receiver.

Accurate, high-resolution panel-mounted potentiometers (or "pots") have resistance elements typically wirewound on a helical mandrel, although some include a conductive-plastic resistance coating over the wire to improve resolution. These typically offer ten turns of their shafts to cover their full range. They are usually set with dials that include a simple turns counter and a graduated dial. Electronic analog computers used them in quantity for setting coefficients, and delayed-sweep oscilloscopes of recent decades included one on their panels.

Resistance decade boxes

A resistance decade box or resistor substitution box is a unit containing resistors of many values, with one or more mechanical switches which allow any one of various discrete resistances offered by the box to be dialed in. Usually the resistance is accurate to high precision, ranging from laboratory/calibration grade accuracy of 20 parts per million, to field grade at 1%. Inexpensive boxes with lesser accuracy are also available. All types offer a convenient way of selecting and quickly changing a resistance in laboratory, experimental and development work without needing to attach resistors one by one, or even stock each value. The range of resistance provided, the maximum resolution, and the accuracy characterize the box. For example, one box offers resistances from 0 to 24 megohms, maximum resolution 0.1 ohm, accuracy 0.1%.

Special devices

There are various devices whose resistance changes with various quantities. The resistance of thermistors exhibit a strong negative temperature coefficient, making them useful for measuring temperatures. Since their resistance can be large until they are allowed to heat up due to the passage of current, they are also commonly used to prevent excessive current surges when equipment is powered on. Similarly, the resistance of a humistor varies with humidity. Metal oxide varistors drop to a very low resistance when a high voltage is applied, making them useful for protecting electronic equipment by absorbing dangerous voltage surges. One sort of photodetector, the photoresistor, has a resistance which varies with illumination.

The strain gauge, invented by Edward E. Simmons and Arthur C. Ruge in 1938, is a type of resistor that changes value with applied strain. A single resistor may be used, or a pair (half bridge), or four resistors connected in a Wheatstone bridge configuration. The strain resistor is bonded with adhesive to an object that will be subjected to mechanical strain. With the strain gauge and a filter, amplifier, and analog/digital converter, the strain on an object can be measured.

A related but more recent invention uses a Quantum Tunnelling Composite to sense mechanical stress. It passes a current whose magnitude can vary by a factor of 10^{12} in response to changes in applied pressure.

Measurement

The value of a resistor can be measured with an ohmmeter, which may be one function of a multimeter. Usually, probes on the ends of test leads connect to the resistor. A simple ohmmeter may apply a voltage from a battery across the unknown resistor (with an internal resistor of a known value in series) producing a current which drives a meter movement. The current flow, in accordance with Ohm's Law, is inversely proportional to the sum of the internal resistance and the resistor being tested, resulting in an analog meter scale which is very non-linear, calibrated from infinity to 0 ohms. A digital multimeter, using active electronics, may instead pass a specified current through the test resistance. The voltage generated across the test resistance in that case is linearly proportional to its resistance, which is measured and displayed.

In either case the low-resistance ranges of the meter pass much more current through the test leads than do high-resistance ranges, in order for the voltages present to be at reasonable levels (generally below 10 volts) but still measurable.

Measuring low-value resistors, such as fractional-ohm resistors, with acceptable accuracy requires four-terminal connections. One pair of terminals applies a known, calibrated current to the resistor, while the other pair senses the voltage drop across the resistor. Some laboratory quality ohmmeters, especially milliohmmeters, and even some of the better digital multimeters sense using four input terminals for this purpose, which may be used with special test leads. Each of the two so-called Kelvin clips has a pair of jaws insulated from each other. One side of each clip applies the measuring current, while the other connections are only to sense the voltage drop. The resistance is again calculated using Ohm's Law as the measured voltage divided by the applied current.

Standards

Production resistors

Resistor characteristics are quantified and reported using various national standards. In the US, MIL-STD-202 contains the relevant test methods to which other standards refer.

There are various standards specifying properties of resistors for use in equipment:

- BS 1852
- EIA-RS-279
- MIL-PRF-26
- MIL-PRF-39007 (Fixed Power, established reliability)
- MIL-PRF-55342 (Surface-mount thick and thin film)
- MIL-PRF-914
- MIL-R-11
- MIL-R-39017 (Fixed, General Purpose, Established Reliability)
- MIL-PRF-32159 (zero ohm jumpers)

There are other United States military procurement MIL-R- standards.

Resistance standards

The primary standard for resistance, the "mercury ohm" was initially defined in 1884 in as a column of mercury 106 mm long and 1 square millimeter in cross-section, at 0 degrees Celsius. Difficulties in precisely measuring the physical constants to replicate this standard result in variations of as much as 30 ppm. From 1900 the mercury ohm was replaced with a precision machined plate of manganin. Since 1990 the international resistance standard has been based on the quantized Hall effect discovered by Klaus von Klitzing, for which he won the Nobel Prize in Physics in 1985.

Resistors of extremely high precision are manufactured for calibration and laboratory use. They may have four terminals, using one pair to carry an operating current and the other pair to measure the voltage drop; this eliminates errors caused by voltage drops across the lead resistances, because no current flows through voltage sensing leads. It is important in small value resistors (100–0.0001 ohm) where lead resistance is significant or even comparable with respect to resistance standard value.

Resistor marking

Most axial resistors use a pattern of colored stripes to indicate resistance. Surface-mount resistors are marked numerically, if they are big enough to permit marking; more-recent small sizes are impractical to mark. Cases are usually tan, brown, blue, or green, though other colors are occasionally found such as dark red or dark gray.

Early 20th century resistors, essentially uninsulated, were dipped in paint to cover their entire body for color coding. A second color of paint was applied to one end of the element, and a color dot (or band) in the middle provided the third digit. The rule was "body, tip, dot", providing two significant digits for value and the decimal multiplier, in that sequence. Default tolerance was ±20%. Closer-tolerance resistors had silver (±10%) or gold-colored (±5%) paint on the other end.

Four-band resistors

Four-band identification is the most commonly used color-coding scheme on resistors. It consists of four colored bands that are painted around the body of the resistor. The first two bands encode the first two significant digits of the resistance value, the third is a power-of-ten multiplier or number-of-zeroes, and the fourth is the tolerance accuracy, or acceptable error, of the value. The first three bands are equally spaced along the resistor; the spacing to the fourth band is wider. Sometimes a fifth band identifies the thermal coefficient, but this must be distinguished from the true 5-color system, with 3 significant digits.

For example, green-blue-yellow-red is 56×10^4 Ω = 560 kΩ ± 2%. An easier description can be as followed: the first band, green, has a value of 5 and the second band, blue, has a value of 6, and is counted as 56. The third band, yellow, has a value of 10^4, which adds four 0's to the end, creating 560,000 Ω at ±2% tolerance accuracy. 560,000 Ω changes to 560 kΩ ±2% (as a kilo- is 10^3).

Each color corresponds to a certain digit, progressing from darker to lighter colors, as shown in the chart below.

Color	1st band	2nd band	3rd band (multiplier)	4th band (tolerance)	Temp. Coefficient
Black	0	0	$\times10^0$		
Brown	1	1	$\times10^1$	±1% (F)	100 ppm
Red	2	2	$\times10^2$	±2% (G)	50 ppm
Orange	3	3	$\times10^3$		15 ppm

Yellow	4	4	$\times 10^4$		25 ppm
Green	5	5	$\times 10^5$	±0.5% (D)	
Blue	6	6	$\times 10^6$	±0.25% (C)	
Violet	7	7	$\times 10^7$	±0.1% (B)	
Gray	8	8	$\times 10^8$	±0.05% (A)	
White	9	9	$\times 10^9$		
Gold			$\times 10^{-1}$	±5% (J)	
Silver			$\times 10^{-2}$	±10% (K)	
None				±20% (M)	

There are many mnemonics for remembering these colors.

Preferred values

Early resistors were made in more or less arbitrary round numbers; a series might have 100, 125, 150, 200, 300, etc. Resistors as manufactured are subject to a certain percentage tolerance, and it makes sense to manufacture values that correlate with the tolerance, so that the actual value of a resistor overlaps slightly with its neighbors. Wider spacing leaves gaps; narrower spacing increases manufacturing and inventory costs to provide resistors that are more or less interchangeable.

A logical scheme is to produce resistors in a range of values which increase in a geometrical progression, so that each value is greater than its predecessor by a fixed multiplier or percentage, chosen to match the tolerance of the range. For example, for a tolerance of ±20% it makes sense to have each resistor about 1.5 times its predecessor, covering a decade in 6 values. In practice the factor used is 1.4678, giving values of 1.47, 2.15, 3.16, 4.64, 6.81, 10 for the 1-10 decade (a decade is a range increasing by a factor of 10; 0.1-1 and 10-100 are other examples); these are rounded in practice to 1.5, 2.2, 3.3, 4.7, 6.8, 10; followed, of course by 15, 22, 33, … and preceded by … 0.47, 0.68, 1. This scheme has been adopted as the **E6** range of the IEC 60063 preferred number series. There are also **E12**, **E24**, **E48**, **E96** and **E192** ranges for components of ever tighter tolerance, with 12, 24, 96, and 192 different values within each decade. The actual values used are in the IEC 60063 lists of preferred numbers.

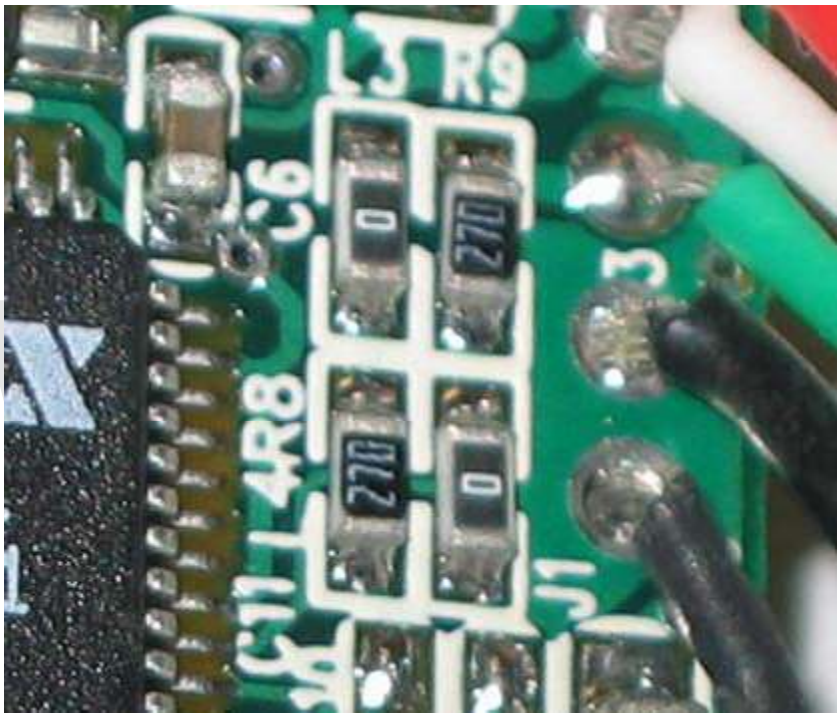
A resistor of 100 ohms ±20% would be expected to have a value between 80 and 120 ohms; its E6 neighbors are 68 (54-82) and 150 (120-180) ohms. A sensible spacing, E6 is used for ±20% components; E12 for ±10%; E24 for ±5%; E48 for ±2%, E96 for ±1%; E192 for ±0.5% or better. Resistors are manufactured in values from a few milliohms to about a gigaohm in IEC60063 ranges appropriate for their tolerance.

Earlier power wirewound resistors, such as brown vitreous-enameled types, however, were made with a different system of preferred values, such as some of those mentioned in the first sentence of this section.

5-band axial resistors

5-band identification is used for higher precision (lower tolerance) resistors (1%, 0.5%, 0.25%, 0.1%), to specify a third significant digit. The first three bands represent the significant digits, the fourth is the multiplier, and the fifth is the tolerance. Five-band resistors with a gold or silver 4th band are sometimes encountered, generally on older or specialized resistors. The 4th band is the tolerance and the 5th the temperature coefficient.

SMD resistors



This image shows four surface-mount resistors (the component at the upper left is a capacitor) including two zero-ohm resistors. Zero-ohm links are often used instead of wire links, so that they can be inserted by a resistor-inserting machine. Of course, their resistance is non-zero, although quite low. *Zero* is simply a brief description of their function.

Surface mounted resistors are printed with numerical values in a code related to that used on axial resistors. Standard-tolerance surface-mount technology (SMT) resistors are marked with a three-digit code, in which the first two digits are the first two significant digits of the value and the third digit is the power of ten (the number of zeroes). For example:

334 = 33 × 10^4 ohms = 330 kilohms
222 = 22 × 10^2 ohms = 2.2 kilohms
473 = 47 × 10^3 ohms = 47 kilohms
105 = 10 × 10^5 ohms = 1.0 megohm

Resistances less than 100 ohms are written: 100, 220, 470. The final zero represents ten to the power zero, which is 1. For example:

100 = 10 × 10^0 ohm = 10 ohms
220 = 22 × 10^0 ohm = 22 ohms

Sometimes these values are marked as *10* or *22* to prevent a mistake.

Resistances less than 10 ohms have 'R' to indicate the position of the decimal point (radix point). For example:

4R7 = 4.7 ohms
R300 = 0.30 ohms
0R22 = 0.22 ohms
0R01 = 0.01 ohms

Precision resistors are marked with a four-digit code, in which the first three digits are the significant figures and the fourth is the power of ten. For example:

1001 = 100 × 10^1 ohms = 1.00 kilohm
4992 = 499 × 10^2 ohms = 49.9 kilohm
1000 = 100 × 10^0 ohm = 100 ohms

000 and *0000* sometimes appear as values on surface-mount zero-ohm links, since these have (approximately) zero resistance.

More recent surface-mount resistors are too small, physically, to permit practical markings to be applied.

Industrial type designation

Format: *[two letters]<space>[resistance value (three digit)]<nospace>[tolerance code(numerical - one digit)]*

	Power Rating at 70 °C		
Type No.	**Power rating (watts)**	**MIL-R-11 Style**	**MIL-R-39008 Style**
BB	⅛	RC05	RCR05
CB	¼	RC07	RCR07
EB	½	RC20	RCR20
GB	1	RC32	RCR32

HB	2	RC42	RCR42
GM	3	-	-
HM	4	-	-

Tolerance Code

Industrial type designation	Tolerance	MIL Designation
5	±5%	J
2	±20%	M
1	±10%	K
-	±2%	G
-	±1%	F
-	±0.5%	D
-	±0.25%	C
-	±0.1%	B

The operational temperature range distinguishes commercial grade, industrial grade and military grade components.

- Commercial grade: 0 °C to 70 °C
- Industrial grade: −40 °C to 85 °C (sometimes −25 °C to 85 °C)
- Military grade: −55 °C to 125 °C (sometimes -65 °C to 275 °C)
- Standard Grade -5 °C to 60 °C

Electrical and thermal noise

In amplifying faint signals, it is often necessary to minimize electronic noise, particularly in the first stage of amplification. As dissipative elements, even an ideal resistor will naturally produce a randomly fluctuating voltage or "noise" across its terminals. This Johnson–Nyquist noise is a fundamental noise source which depends only upon the temperature and resistance of the resistor, and is predicted by the fluctuation–dissipation theorem. Using a larger resistor produces a larger voltage noise, whereas with a smaller value of resistance there will be more current noise, assuming a given temperature. The thermal noise of a practical resistor may also be somewhat larger than the theoretical prediction and that increase is typically frequency-dependent.

However the "excess noise" of a practical resistor is an additional source of noise observed only when a current flows through it. This is specified in unit of μV/V/decade - μV of noise per volt applied across the resistor per decade of frequency. The μV/V/decade value is frequently given in dB so that a resistor with a noise index of 0 dB will exhibit 1 μV (rms) of excess noise for each volt across the resistor in each frequency decade. Excess noise is thus an example of 1/*f* noise. Thick-film and carbon composition resistors generate more excess noise than other types at low frequencies; wire-wound and thin-film resistors, though much more expensive, are often utilized for their better noise characteristics. Carbon composition resistors can exhibit a noise index of 0 dB while bulk metal foil resistors may have a noise index of -40 dB, usually making the excess noise of metal foil resistors insignificant. Thin film surface mount resistors typically have lower noise and better thermal stability than thick film surface mount resistors. However,

the design engineer must read the data sheets for the family of devices to weigh the various device tradeoffs.

While not an example of "noise" per se, a resistor may act as a thermocouple, producing a small DC voltage differential across it due to the thermoelectric effect if its ends are at somewhat different temperatures. This induced DC voltage can degrade the precision of instrumentation amplifiers in particular. Such voltages appear in the junctions of the resistor leads with the circuit board and with the resistor body. Common metal film resistors show such an effect at a magnitude of about 20 μV/°C. Some carbon composition resistors can exhibit thermoelectric offsets as high as 400 μV/°C, whereas specially constructed resistors can reduce this number to 0.05 μV/°C. In applications where the thermoelectric effect may become important, care has to be taken (for example) to mount the resistors horizontally to avoid temperature gradients and to mind the air flow over the board.

Failure modes

The failure rate of resistors in a properly designed circuit is low compared to other electronic components such as semiconductors and electrolytic capacitors. Damage to resistors most often occurs due to overheating when the average power delivered to it (as computed above) greatly exceeds its ability to dissipate heat (specified by the resistor's *power rating*). This may be due to a fault external to the circuit, but is frequently caused by the failure of another component (such as a transistor that shorts out) in the circuit connected to the resistor. Operating a resistor too close to its power rating can limit the resistor's lifespan or cause a change in its resistance over time which may or may not be noticeable. A safe design generally uses overrated resistors in power applications to avoid this danger.

When overheated, carbon-film resistors may decrease or increase in resistance. Carbon film and composition resistors can fail (open circuit) if running close to their maximum dissipation. This is also possible but less likely with metal film and wirewound resistors.

There can also be failure of resistors due to mechanical stress and adverse environmental factors including humidity. If not enclosed, wirewound resistors can corrode.

Variable resistors degrade in a different manner, typically involving poor contact between the wiper and the body of the resistance. This may be due to dirt or corrosion and is typically perceived as "crackling" as the contact resistance fluctuates; this is especially noticed as the device is adjusted. This is similar to crackling caused by poor contact in switches, and like switches, potentiometers are to some extent self-cleaning: running the wiper across the resistance may improve the contact. Potentiometers which are seldom adjusted, especially in dirty or harsh environments, are most likely to develop this problem. When self-cleaning of the contact is insufficient, improvement can usually be obtained through the use of contact cleaner (also known as "tuner cleaner") spray. The crackling noise associated with turning the shaft of a dirty potentiometer in an audio circuit (such as the volume control) is greatly accentuated when an undesired DC voltage is present, often implicating the failure of a DC blocking capacitor in the circuit.

Chapter 4

Electrical Ballast

"Choke ballast" (inductor) used in older lighting. This example is from a tanning bed. Requires a lamp starter (below) and capacitor.

Lamp starter, required with some inductor type ballasts. Connects both ends of the lamp together to "preheat" the lamp ends for 1 second before lighting.

An **electrical ballast** (sometimes called **control gear**) is a device intended to limit the amount of current in an electric circuit.

Ballasts vary greatly in complexity. They can be as simple as a series resistor as commonly used with small neon lamps or light-emitting diodes (LEDs). For higher-power installations, too much energy would be wasted in a resistive ballast, so alternatives are used that depend upon

the reactance of inductors, capacitors, or both. Finally, ballasts can be as complex as the computerized, remote-controlled electronic ballasts now often used with fluorescent lamps.

Current limiting

Ballasts stabilize the current through an electrical load. These are most often used when an electrical circuit or device presents a negative (differential) resistance to the supply. If such a device were connected to a constant-voltage power supply, it would draw an increasing amount of current until it was destroyed or caused the power supply to fail. To prevent this, a ballast provides a positive resistance or reactance that limits the ultimate current to an appropriate level. In this way, the ballast provides for the proper operation of the negative-resistance device by appearing to be a legitimate, stable resistance in the circuit.

An example of a negative-resistance device is a gas-discharge lamp, where after lamp ignition, increasing arc current reduces the voltage drop.

Ballasts can also be used simply to deliberately reduce the current in an ordinary, positive-resistance circuit.

Prior to the advent of solid-state ignition, automobile ignition systems commonly included a ballast resistor to regulate the voltage applied to the ignition system.

Although LEDs are positive resistance devices, they have insufficient resistance to regulate their current consumption when operated from a voltage controlled source, so ballasts are used to control the current through the LED. Because the power dissipation is minuscule, simple resistor ballasts are normally used.

Resistors

A **ballast resistor** compensates for normal or incidental changes in the physical state of a system. It may be a fixed or variable resistor.

Fixed resistors

For simple, low-powered loads such as a neon lamp or LED, a fixed resistor is commonly used. Because the resistance of the ballast resistor is large it dominates the current in the circuit, even in the face of negative resistance introduced by the neon lamp.

The term also refers to an automobile engine component that lowers the supply voltage to the ignition system after the engine has been started. Because cranking the engine causes a very heavy load on the battery, the system voltage can drop quite low during cranking. To allow the engine to start, the ignition system must be designed to operate on this lower voltage. But once cranking is completed, the normal operating voltage is regained; this voltage would overload the ignition system. To avoid this problem, a ballast resistor is inserted in series with the supply voltage feeding the ignition system. Occasionally, this ballast resistor will fail and the classic symptom of this failure is that the engine runs while being cranked (while the resistor is

bypassed) but stalls immediately when cranking ceases (and the resistor is re-connected in the circuit).

Modern electronic ignition systems do not require a ballast resistor as they are flexible enough to operate on the low cranking voltage or the ordinary operating voltage.

In some old AC/DC receivers (universal sets), the vacuum tube heaters are connected in series. Since the voltage drop across all the filaments in series is sometimes less than the full mains voltage, it was often necessary to get rid of the excess voltage. A ballast resistor was often used for this purpose, as it was cheap and worked with both AC and DC.

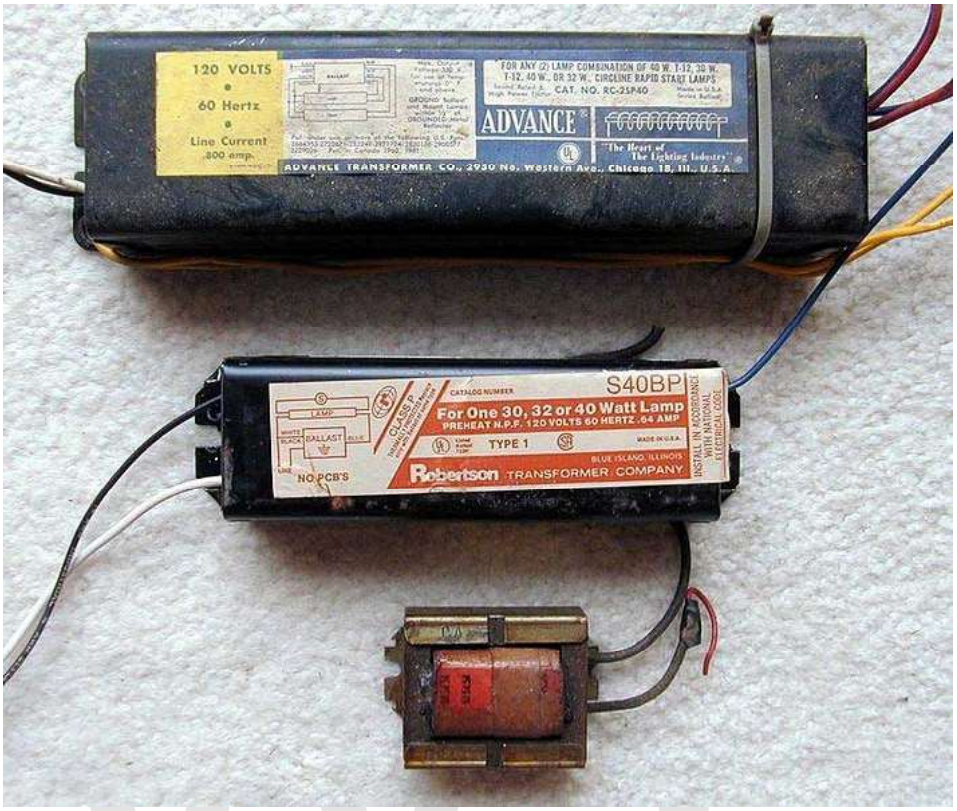
Self-variable resistors

Some ballast resistors have the property of increasing in resistance as current through them increases, and decreasing in resistance as current decreases. Physically, some such devices are often built quite like incandescent lamps. Like the tungsten filament of an ordinary incandescent lamp, if current increases, the ballast resistor gets hotter, its resistance goes up, and its voltage drop increases. If current decreases, the ballast resistor gets colder, its resistance drops, and the voltage drop decreases. Therefore the ballast resistor reduces variations in current, despite variations in applied voltage or changes in the rest of an electric circuit. These devices are sometimes termed barretters.

This property can lead to more precise current control than merely choosing an appropriate fixed resistor. The power lost in the resistive ballast is also reduced because a smaller portion of the overall power is dropped in the ballast compared to what might be required with a fixed resistor.

In times past, household clothes dryers sometimes incorporated a germicidal lamp in series with an ordinary incandescent lamp; the incandescent lamp operated as the ballast for the germicidal lamp. A commonly used light in the home in the 1960s in 220-240V countries was a circleline tube ballasted by an under-run regular mains filament lamp. Self ballasted mercury-vapor lamps incorporate ordinary tungsten filaments within the overall envelope of the lamp to act as the ballast, and it supplements the otherwise lacking red area of the light spectrum produced.

Reactive ballasts



Several typical magnetic ballasts for fluorescent lamps. The top is a high-power factor rapid start series ballast for two 30-40W lamps. The middle is a low power factor preheat ballast for a single 30-40W lamp while the bottom ballast is a simple inductor used with a 15W preheat lamp.

Because of the power that would be lost, resistors are not used as ballasts for lamps of more than about two watts. Instead, a reactance is used. Losses in the ballast due to its resistance and losses in its magnetic core may be significant, on the order of 5 to 25% of the lamp input wattage. Practical lighting design calculations must allow for ballast loss in estimating the running cost of a lighting installation.

An inductor is very common in line-frequency ballasts to provide the proper starting and operating electrical condition to power a fluorescent lamp, neon lamp, or high intensity discharge (HID) lamp. (Because of the use of the inductor, such ballasts are usually called *magnetic ballasts*.) The inductor has two benefits:

1. Its reactance limits the power available to the lamp with only minimal power losses in the inductor
2. The voltage spike produced when current through the inductor is rapidly interrupted is used in some circuits to first strike the arc in the lamp.

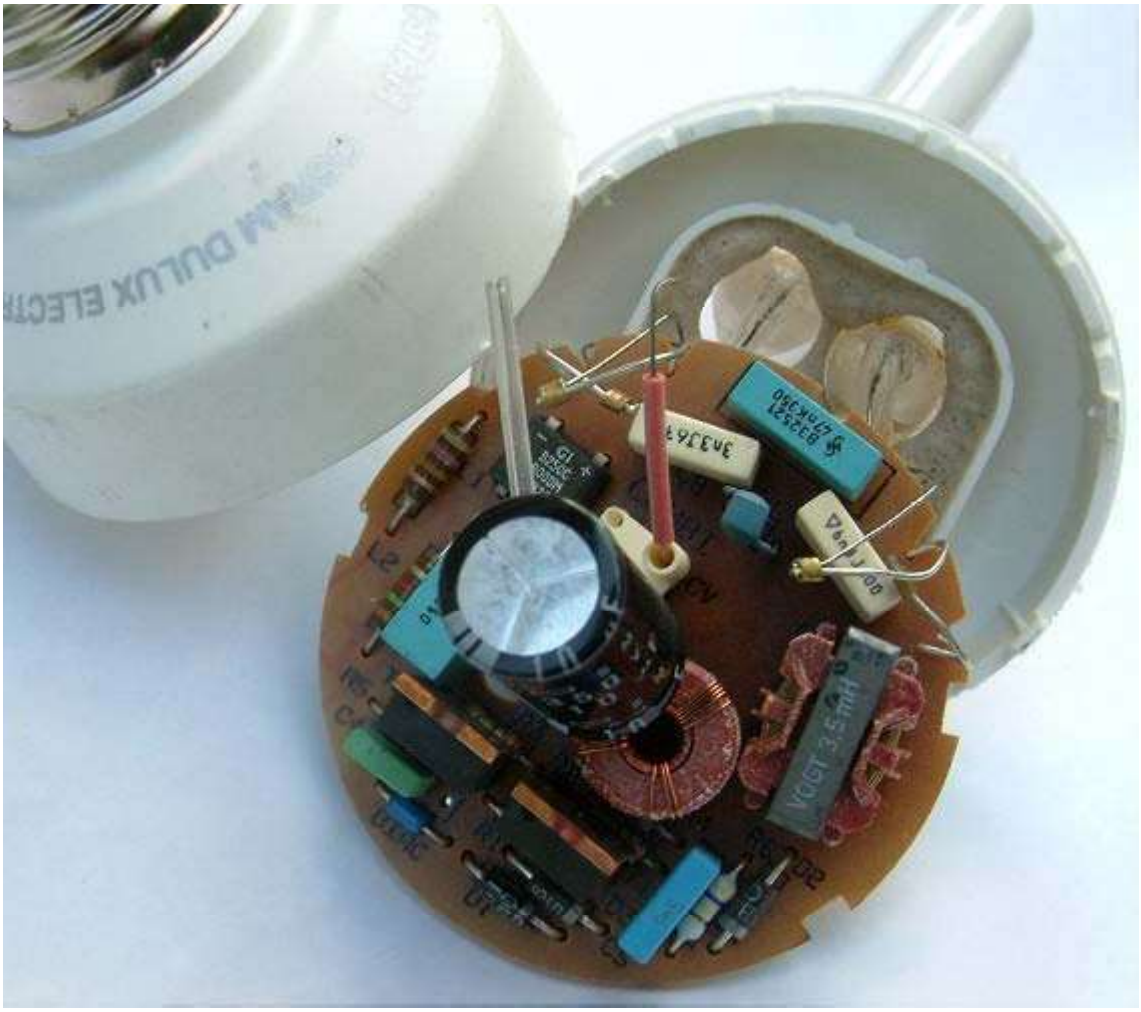
A disadvantage of the inductor is that current is shifted out of phase with the voltage, producing a poor power factor. In more expensive ballasts, a capacitor is often paired with the inductor to correct the power factor. In ballasts that control two or more lamps, line-frequency ballasts commonly use different phase relationships between the multiple lamps. This not only mitigates the flicker of the individual lamps, it also helps maintain a high power factor. These ballasts are often called *lead-lag* ballasts because the current in one lamp leads the mains phase and the current in the other lamp lags the mains phase.

For large lamps, line voltage may not be sufficient to start the lamp, so an autotransformer winding is included in the ballast to step up the voltage. The autotransformer is designed with enough leakage inductance so that the current is appropriately limited.

Because of the large inductors and capacitors that must be used, reactive ballasts operated at line frequency tend to be large and heavy. They commonly also produce acoustic noise (line-frequency hum).

Prior to 1980 in the United States, PCB-based oils were used as an insulating oil in many ballasts to provide cooling and electrical isolation.

Electronic ballasts



Electronic ballast of a compact fluorescent lamp

An **electronic lamp ballast** uses solid state electronic circuitry to provide the proper starting and operating electrical condition to power one or more fluorescent lamps and more recently HID lamps. Electronic ballasts usually change the frequency of the power from the standard mains (e.g., 60 Hz in U.S.) frequency to 20,000 Hz or higher, substantially eliminating the stroboscopic effect of flicker (a product of the line frequency) associated with fluorescent lighting. In addition, because more gas remains ionized in the arc stream, the lamps actually operate at about 9% higher efficacy above approximately 10 kHz. Lamp efficacy increases sharply at about 10 kHz and continues to improve until approximately 20 kHz. Because of the higher efficiency of the ballast itself and the improvement of lamp efficacy by operating at a higher frequency, electronic ballasts offer higher system efficacy for low pressure lamps like the fluorescent lamp. For HID lamps there is no improvement of the lamp efficacy in using higher frequency, but for these lamps the ballast losses are lower at higher frequencies and also the light depreciation is lower meaning more light after a given operating time of say 10 000 hours. Some HID lamp types like the Ceramic discharge metal halide lamp have reduced reliability when operated at high frequencies in the range of 20kHz to 200 kHz and for these lamps a square wave low frequency current drive is mostly used with frequency in the range of 100 to

400 Hz, with the same advantage of lower light depreciation. Electronic ballasts are often based on the SMPS topology, first rectifying the input power and then chopping it at a high frequency. Advanced electronic ballasts may allow dimming via pulse-width modulation or via changing the frequency to a higher value and remote control and monitoring via networks such as LonWorks, DALI, DMX-512, DSI or simple analog control using a 0-10V DC brightness control signal. Recently also systems remotely controlling the dim level via a wireless mesh network have been introduced.

Fluorescent lamp ballasts

Instant start

An instant start ballast starts lamps without heating the cathodes at all by using high voltage (around 600 V). It is the most energy efficient type, but gives the least number of starts from a lamp as emissive oxides are blasted from the cold cathode surfaces each time the lamp is started. This is the best type for installations where lamps are not turned on and off very often.

Rapid start

A rapid start ballast applies voltage and heats the cathodes simultaneously. Provides superior lamp life and more cycle life, but uses slightly more energy as the cathodes in each end of the lamp continue to consume heating power as the lamp operates. A dimming circuit can be used with a dimming ballast, which maintains the heating current while allowing lamp current to be controlled.

Programmed start

A programmed-start ballast is a more advanced version of rapid start. This ballast applies power to the filaments first, then after a short delay to allow the cathodes to preheat, applies voltage to the lamps to strike an arc. This ballast gives the best life and most starts from lamps, and so is preferred for applications with very frequent power cycling such as vision examination rooms and restrooms with a motion detector switch.

Ballast factor

For a lighting ballast, the *ballast factor* is defined as the light output (in lumens) with a test ballast, compared to the light output with a laboratory reference ballast that operates the lamp at its specified nominal power rating. The ballast factor of practical ballasts must be considered in lighting design; a low ballast factor may save energy, but will produce less light. With fluorescent lamps, an electronic ballast may produce more light than the reference test ballast, which operates the lamp with line frequency current; such electronic ballasts have a ballast factor greater than one.

Chapter 5

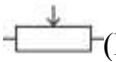
Potentiometer

Potentiometer

A typical single-turn potentiometer

Type Passive

Electronic symbol


(Europe)


(US)

A **potentiometer** (colloquially known as a "**pot**") is a three-terminal resistor with a sliding contact that forms an adjustable voltage divider. If only two terminals are used (one side and the wiper), it acts as a ***variable resistor*** or ***rheostat***. Potentiometers are commonly used to control

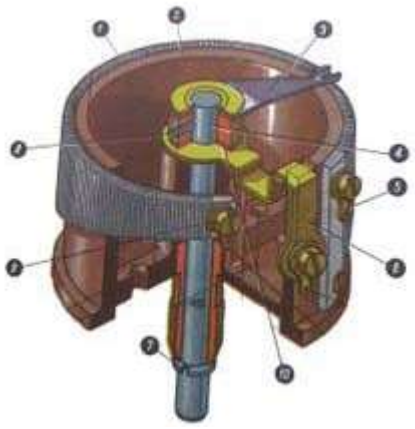
electrical devices such as volume controls on audio equipment. Potentiometers operated by a mechanism can be used as position transducers, for example, in a joystick.

Potentiometers are rarely used to directly control significant power (more than a watt), since the power dissipated in the potentiometer would be comparable to the power in the controlled load. Instead they are used to adjust the level of analog signals (e.g. volume controls on audio equipment), and as control inputs for electronic circuits. For example, a light dimmer uses a potentiometer to control the switching of a TRIAC and so indirectly control the brightness of lamps.

History

The slide-wire potentiometer was invented by Johann Christian Poggendorff (1796–1877) in 1841. Leeds and Northrup Type K model was a standard piece of apparatus in most college and university electrical measurements laboratories for the first half of the 20th century.

Potentiometer construction



Construction of a wire-wound circular potentiometer. The resistive element (1) of the shown device is trapezoidal, giving a non-linear relationship between resistance and turn angle. The wiper (3) rotates with the axis (4), providing the changeable resistance between the wiper contact (6) and the fixed contacts (5) and (9). The vertical position of the axis is fixed in the body (2) with the ring (7) (below) and the bolt (8) (above).

A potentiometer is constructed with a resistive element formed into an arc of a circle, and a sliding contact (wiper) travelling over that arc. The resistive element, with a terminal at one or both ends, is flat or angled, and is commonly made of graphite, although other materials may be used. The wiper is connected through another sliding contact to another terminal. On panel potentiometers, the wiper is usually the center terminal of three. For single-turn potentiometers, this wiper typically travels just under one revolution around the contact. "Multiturn" potentiometers also exist, where the resistor element may be helical and the wiper may move 10, 20, or more complete revolutions, though multiturn potentimeters are usually constructed of a conventional resistive element wiped via a worm gear. Besides graphite, materials used to

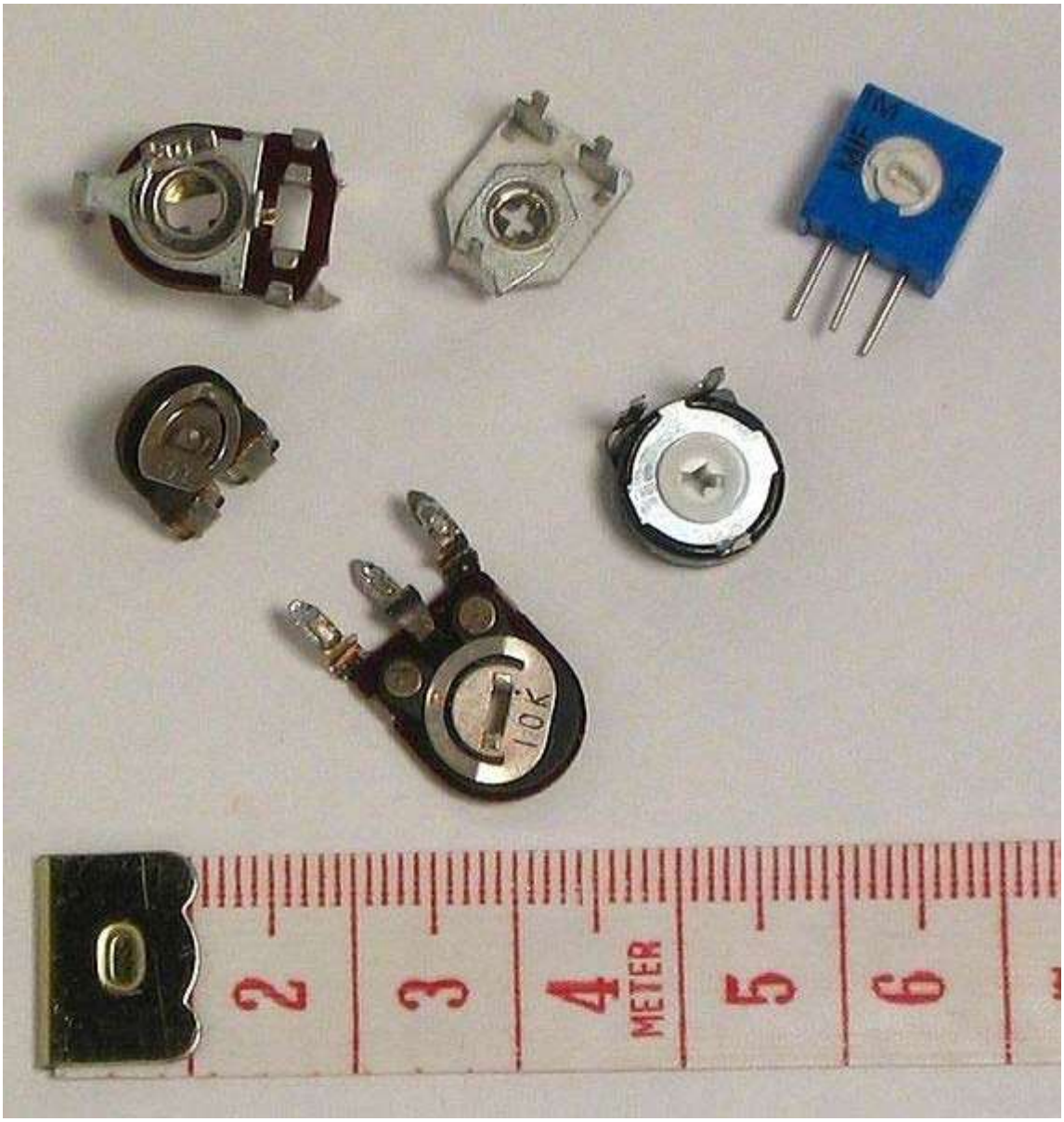
make the resistive element include resistance wire, carbon particles in plastic, and a ceramic/metal mixture called cermet.

One form of rotary potentiometer is called a String potentiometer. It is a multi-turn potentiometer operated by an attached reel of wire turning against a spring. It is used as a position transducer.

In a linear slider potentiometer, a sliding control is provided instead of a dial control. The resistive element is a rectangular strip, not semi-circular as in a rotary potentiometer. Due to the large opening slot or the wiper, this type of potentiometer has a greater potential for getting contaminated.

Potentiometers can be obtained with either linear or logarithmic relations between the slider position and the resistance (potentiometer laws or "tapers"). A letter code ("A" taper, "B" taper, etc.) may be used to identify which taper is intended, but the letter code definitions are variable over time and between manufacturers.

Manufacturers of conductive track potentiometers use conductive polymer resistor pastes that contain hard wearing resins and polymers, solvents, lubricant and carbon – the constituent that provides the conductive/resistive properties. The tracks are made by screen printing the paste onto a paper based phenolic substrate and then curing it in an oven. The curing process removes all solvents and allows the conductive polymer to polymerize and cross link. This produces a durable track with stable electrical resistance throughout its working life.



PCB mount trimmer potentiometers, or "trimpots", intended for infrequent adjustment.

Linear taper potentiometer

A *linear taper potentiometer* has a resistive element of constant cross-section, resulting in a device where the resistance between the contact (wiper) and one end terminal is proportional to the distance between them. *Linear taper* describes the electrical characteristic of the device, not the geometry of the resistive element. Linear taper potentiometers are used when an approximately proportional relation is desired between shaft rotation and the division ratio of the potentiometer; for example, controls used for adjusting the centering of (an analog) cathode-ray oscilloscope.

Logarithmic potentiometer

A *logarithmic taper potentiometer* has a resistive element that either 'tapers' in from one end to the other, or is made from a material whose resistivity varies from one end to the other. This results in a device where output voltage is a logarithmic function of the mechanical angle of the potentiometer.

Most (cheaper) "log" potentiometers are actually not logarithmic, but use two regions of different resistance (but constant resistivity) to approximate a logarithmic law. A logarithmic potentiometer can also be simulated with a linear one and an external resistor. True logarithmic potentiometers are significantly more expensive.

Logarithmic taper potentiometers are often used in connection with audio amplifiers as human perception of audio volume is logarithmic.

A high power wirewound potentiometer. Any potentiometer may be connected as a rheostat.

Rheostat

The most common way to vary the resistance in a circuit is to use a variable resistor or a **rheostat** A rheostat is a two-terminal variable resistor. Often these are designed to handle much higher voltage and current. Typically these are constructed as a resistive wire wrapped to form a

toroid coil with the wiper moving over the upper surface of the toroid, sliding from one turn of the wire to the next. Sometimes a rheostat is made from resistance wire wound on a heat-resisting cylinder with the slider made from a number of metal fingers that grip lightly onto a small portion of the turns of resistance wire. The "fingers" can be moved along the coil of resistance wire by a sliding knob thus changing the "tapping" point. They are usually used as variable resistors rather than variable potential dividers.

Any three-terminal potentiometer can be used as a two-terminal variable resistor by not connecting to the third terminal. It is common practice to connect the wiper terminal to the unused end of the resistance track to reduce the amount of resistance variation caused by dirt on the track.

Digital potentiometer

A digital potentiometer is an electronic component that mimics the functions of analog potentiometers. Through digital input signals, the resistance between two terminals can be adjusted, just as in an analog potentiometer.

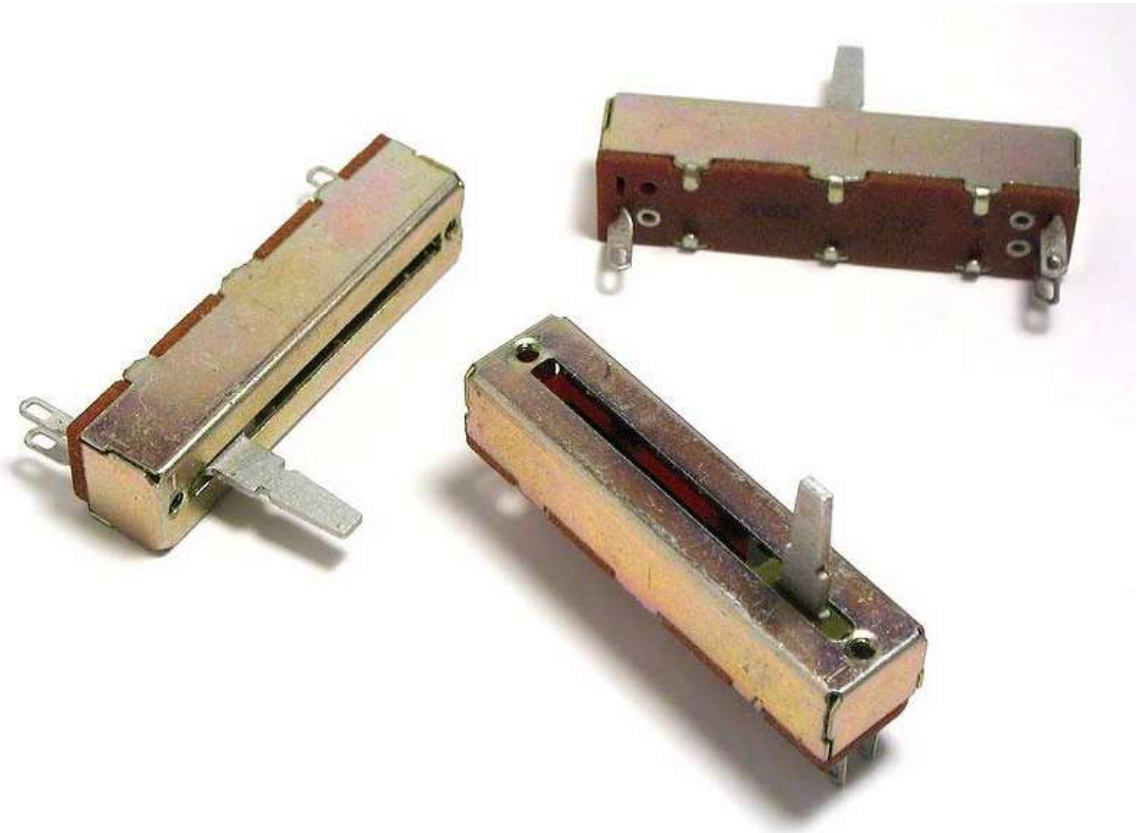
Membrane Potentiometer

A membrane potentiometer uses a conductive membrane that is deformed by a sliding element to contact a resistor voltage divider. Linearity can range from 0.5% to 5% depending on the material, design and manufacturing process. The repeat accuracy is typically between 0.1mm and 1.0mm with a theoretically infinite resolution. The service life of these types of potentiometers is typically 1 million to 20 million cycles depending on the materials used during manufacturing and the actuation method; contact and contactless (magnetic) methods are available. Many different material variations are available such as PET(foil), FR4, and Kapton. Membrane potentiometer manuafacturers offer linear, rotary, and application-specific variations. The linear versions can range from 9mm to 1000mm in length and the rotary versions range from 0° to 360°(multi-turn), with each having a height of 0.5mm. Membrane potentiometers can be used for position sensing.

Potentiometer applications

Potentiometers are widely used as user controls, and may control a very wide variety of equipment functions. The widespread use of potentiometers in consumer electronics has declined in the 1990s, with digital controls now more common. However they remain in many applications, such as volume controls and as position sensors.

Audio control



Linear potentiometers ("faders")

One of the most common uses for modern low-power potentiometers is as audio control devices. Both linear potentiometers and rotary potentiometers are regularly used to adjust loudness, frequency attenuation and other characteristics of audio signals.

The 'log pot' is used as the volume control in audio amplifiers, where it is also called an "audio taper pot", because the amplitude response of the human ear is also logarithmic. It ensures that, on a volume control marked 0 to 10, for example, a setting of 5 sounds half as loud as a setting of 10. There is also an *anti-log pot* or *reverse audio taper* which is simply the reverse of a logarithmic potentiometer. It is almost always used in a ganged configuration with a logarithmic potentiometer, for instance, in an audio balance control.

Potentiometers used in combination with filter networks act as tone controls or equalizers.

Television

Potentiometers were formerly used to control picture brightness, contrast, and color response. A potentiometer was often used to adjust "vertical hold", which affected the synchronization between the receiver's internal sweep circuit (sometimes a multivibrator) and the received picture signal.

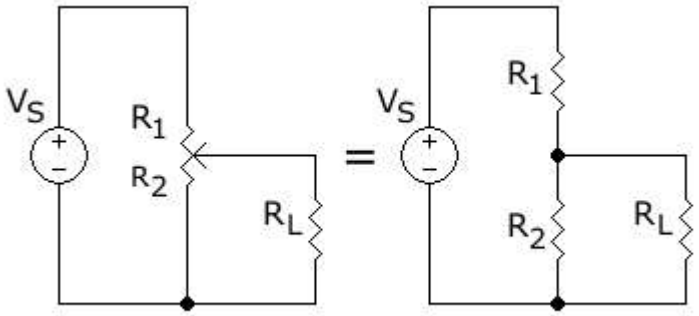
Transducers

Potentiometers are also very widely used as a part of displacement transducers because of the simplicity of construction and because they can give a large output signal.

Computation

In analog computers, high precision potentiometers are used to scale intermediate results by desired constant factors, or to set initial conditions for a calculation. A motor-driven potentiometer may be used as a function generator, using a non-linear resistance card to supply approximations to trigonometric functions. For example, the shaft rotation might represent an angle, and the voltage division ratio can be made proportional to the cosine of the angle.

Theory of operation



A potentiometer with a resistive load, showing equivalent fixed resistors for clarity.

The potentiometer can be used as a voltage divider to obtain a manually adjustable output voltage at the slider (wiper) from a fixed input voltage applied across the two ends of the potentiometer. This is the most common use of them.

The voltage across R_{L} can be calculated by:

$$V_{\mathrm{L}} = \frac{R_2 R_{\mathrm{L}}}{R_1 R_{\mathrm{L}} + R_2 R_{\mathrm{L}} + R_1 R_2} \cdot V_s.$$

If R_{L} is large compared to the other resistances (like the input to an operational amplifier), the output voltage can be approximated by the simpler equation:

$$V_{\mathrm{L}} = \frac{R_2}{R_1 + R_2} \cdot V_s.$$

As an example, assume

$$V_{\mathrm{S}} = 10\ \mathrm{V}, R_1 = 1\ \mathrm{k\Omega}, R_2 = 2\ \mathrm{k\Omega}, \text{ and } R_{\mathrm{L}} = 100\ \mathrm{k\Omega}.$$

Since the load resistance is large compared to the other resistances, the output voltage V_L will be approximately:

$$\frac{2\ \mathrm{k\Omega}}{1\ \mathrm{k\Omega} + 2\ \mathrm{k\Omega}} \cdot 10\ \mathrm{V} = \frac{2}{3} \cdot 10\ \mathrm{V} \approx 6.667\ \mathrm{V}.$$

Due to the load resistance, however, it will actually be slightly lower: ≈ 6.623 V.

One of the advantages of the potential divider compared to a variable resistor in series with the source is that, while variable resistors have a maximum resistance where some current will always flow, dividers are able to vary the output voltage from maximum (V_S) to ground (zero volts) as the wiper moves from one end of the potentiometer to the other. There is, however, always a small amount of contact resistance.

In addition, the load resistance is often not known and therefore simply placing a variable resistor in series with the load could have a negligible effect or an excessive effect, depending on the load.

Early patents

- US patent 131,334, Thomas Edison, "Coiled resistance wire rheostat", issued 1872-9-17
- Mary Hallock-Greenewalt invented a type of nonlinear rheostat for use in her visual-music instrument, the Sarabet (US patent 1,357,773)

Chapter 6

Digital Potentiometer & Bleeder Resistor

Digital Potentiometer

A **digital potentiometer** is a digitally-controlled electronic component that mimics the analog functions of a potentiometer. It is often used for trimming and scaling analog signals by microcontrollers. It is either built using an R-2R integrated circuit or a Digital-to-analog converter. A digital potentiometer is an electronic component that is often controlled by digital protocols like I²C and SPI, as well as more basic Up/Down protocols. Some typical uses of digital potentiometers are in circuits requiring gain control of amplifiers (frequently instrumentation amplifiers), small-signal audio-balancing, and offset adjustment.

Sometimes this device is also referred to as an RDAC, Resistive Digital-to-Analog Converter.

Some Digipots come with non-volatile memory, so that they retain their last programmed position after they have been power cycled. Most, though, are volatile, i.e. after they are power cycled they will default to a standard value, which is usually the mid-point.

The former can be useful, but when they are controlled by a microprocessor, or even via a Field Programmable Gate Array (FPGA), these devices can retain, in other non-volatile memory, the value to initialise the Digipot with. In these circumstances, the need for non-volatile Digipots is less obvious.

Limitations

These devices are extremely useful in the modern, digitally controlled world, but have some limitations. While quite similar to a normal potentiometer, digital potentiometers are somewhat constrained by current limits in the tens of milliamperes. Also, most, if not all digital potentiometers limit the input voltage range to the digital supply range (often 0–5 VDC), so some ingenuity is often required when attempting to replace standard resistive potentiometers with digital potentiometers. Further, instead of the seemingly continuous control that can be obtained from a multiturn resistive potentiometer, digital potentiometers have discrete steps in resistance. Eight-bit pots (256-steps) are most common, but potentiometers between 5 and 10 bits (32 to 1024 steps) are available. A fourth constraint is that special logic is often required to check for zero crossing of an analog AC signal to allow the resistance value to be changed without causing an audible click in the output for audio amplifiers.

The non-volatile Digipots also differ from their electro-mechanical cousins in that on power up, the resistance will default to (possibly) a different value after a power cycle.

Similarly, the Digipot resistance is only valid when the correct DC supply voltage(s) are present. When voltages are removed, the resistance between the two end points and the (nominal) Wiper are undefined. In an operational amplifier circuit, the Off-State impedance of a real potentiometer can help stabilise the DC operating point of the circuit during the power-up stage. This may not be the case when a Digipot is used.

Like their electro-mechanical counterparts, Digipots suffer similar weaknesses. Real potentiometers and Digipots generally have poor tolerances (typically +/- 20%), poor temperature coefficients (many hundreds of ppm per degree C), and a stop resistance that is typically about 0.5-1% of the full scale resistance. Note that Stop Resistance is the residual resistance when the terminal to wiper resistance is set to the minimum value.

Bleeder Resistor

A **bleeder resistor** is a resistor placed in parallel with a high-voltage supply for the purposes of discharging the energy stored in the power source's filter capacitors or other components that store electrical energy when the equipment is turned off.

It is a use for a standard resistor rather than a separate type of component.

Usage

DC power supplies

Power supplies, especially switchmode power supplies, use a bridge rectifier to convert mains AC power into a typical 340 volts DC for the chopper. A large filter capacitor typically stores enough energy at this high voltage to power the load during the zero crossings of the AC input. In fact, the capacitors in many supplies are large enough to support the load during AC outages lasting for a significant fraction of a second. This stored energy is clearly potentially lethal, and without a bleeder resistor it might remain long after the unit has been turned off. With a properly sized bleeder resistor, however, the voltage will quickly decay to safe levels when the supply is switched off, yet not consume too much power while the supply is on.

High voltage supply in television sets

The bleeder resistor commonly found inside a flyback transformer that supplies high voltage for a CRT is valued in the hundreds of megohms range, and can therefore not be measured with the common technician's multimeter.

Instead of a resistor inside the transformer, the focus and screen control array may be used for the same purpose, depending on the application and tolerances of the type of tube it is producing output for.

These bleeders discharge the focus supply, but not the high voltage final anode feed. The CRT itself forms a capacitor that can hold a sizable (and very dangerous) high voltage charge, so it is **always** advisable to momentarily ground a CRT's high voltage terminal before working on the unit.

Failure

The failure of a bleeder resistor prevents the discharge of the capacitors, resulting in dangerous voltages being retained for many days. This is one of several reasons for the typical warning on most equipment: "Warning - No user-serviceable parts inside". An un-suspecting user may get an electrical shock from opened equipment due to failure of a bleeder resistor, or the common practice of not fitting them.

Safe design suggests mounting a bleeder close to a dangerous capacitor, ideally directly to the capacitor terminals, and not through any connectors, so that it is difficult to disconnect the bleeder accidentally.

Despite the presence of a bleeder, it is wise to prove that any potentially dangerous capacitors are discharged, perhaps by shorting their terminals (or through a suitable low resistance for high energy capacitors), before working on any circuit.

Technical considerations

There is always a trade-off between the speed with which the bleeder operates and the amount of power wasted in the bleeder; a faster bleed-down rate wastes more power during normal, power-on operation.

The presence of a bleeder also guarantees a minimum load on the power source, which can help reduce the range of voltage change (regulation) when the normal load is changing and there is no active regulator. Use of a bleeder this way is a common design strategy for power supplies of vacuum tube power amplifiers, for instance.

Dual bleeder

Because of the speed/power tradeoff, high-powered circuits may use two separate bleeder circuits. A fast bleed circuit is switched out during normal operation so that no power is wasted; when power is switched off, the fast bleeder is connected, rapidly bleeding down the voltage. The switch controlling the fast bleeder can fail, either by connecting when it shouldn't (and overheating) or by not connecting when it should (and thereby failing to bleed off the voltage quickly). To avoid the risk of not having an operational bleeder, a secondary, slower (and less lossy) bleeder is usually permanently connected so that there is always some bleed-down capability.

Chapter 7

Attenuator (Electronics)

A 30dB 5W RF-attenuator, DC-18GHz, with N-type coaxial connectors

Coaxial Dynamics 100 Watt power attenuator

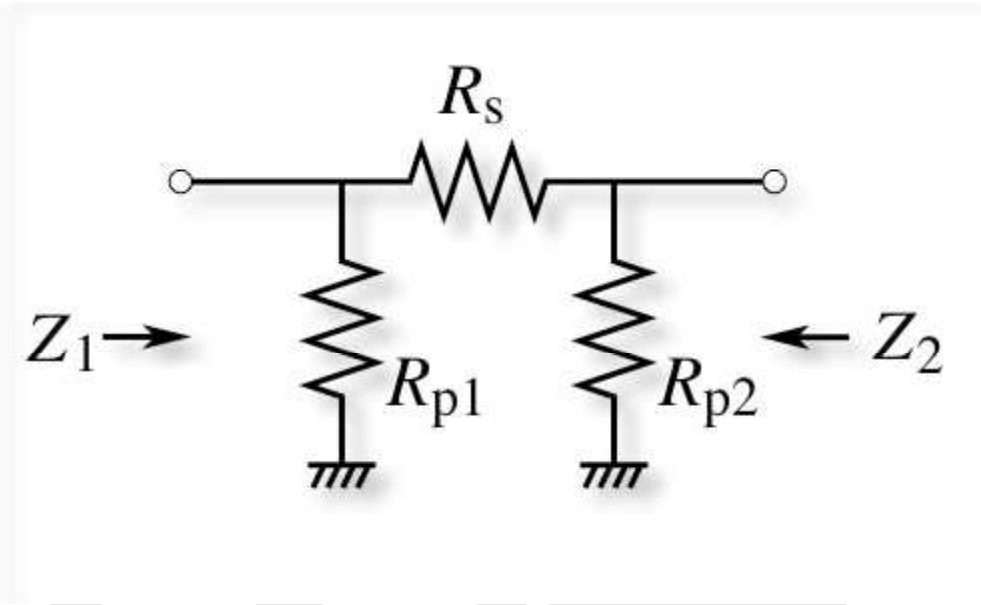
An **attenuator** is an electronic device that reduces the amplitude or power of a signal without appreciably distorting its waveform.

An attenuator is effectively the opposite of an amplifier, though the two work by different methods. While an amplifier provides gain, an attenuator provides loss, or gain less than 1.

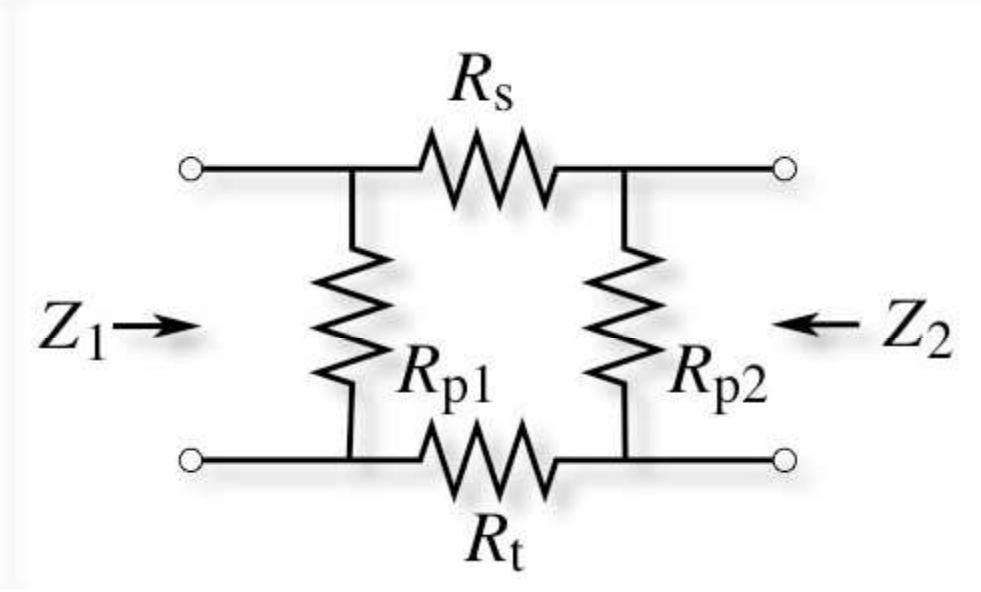
Attenuators are usually passive devices made from simple voltage divider networks. Switching between different resistances forms adjustable stepped attenuators and continuously adjustable ones using potentiometers. For higher frequencies precisely matched low VSWR resistance networks are used.

Fixed attenuators in circuits are used to lower voltage, dissipate power, and to improve impedance matching. In measuring signals, attenuator pads or adaptors are used to lower the amplitude of the signal a known amount to enable measurements, or to protect the measuring device from signal levels that might damage it. Attenuators are also used to 'match' impedances by lowering apparent SWR.

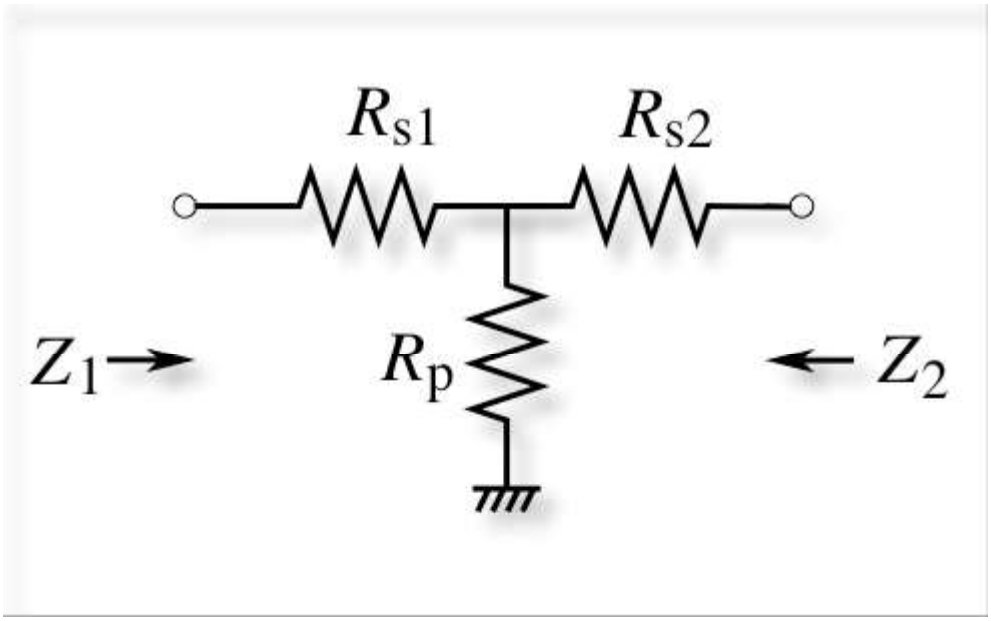
Attenuator circuits



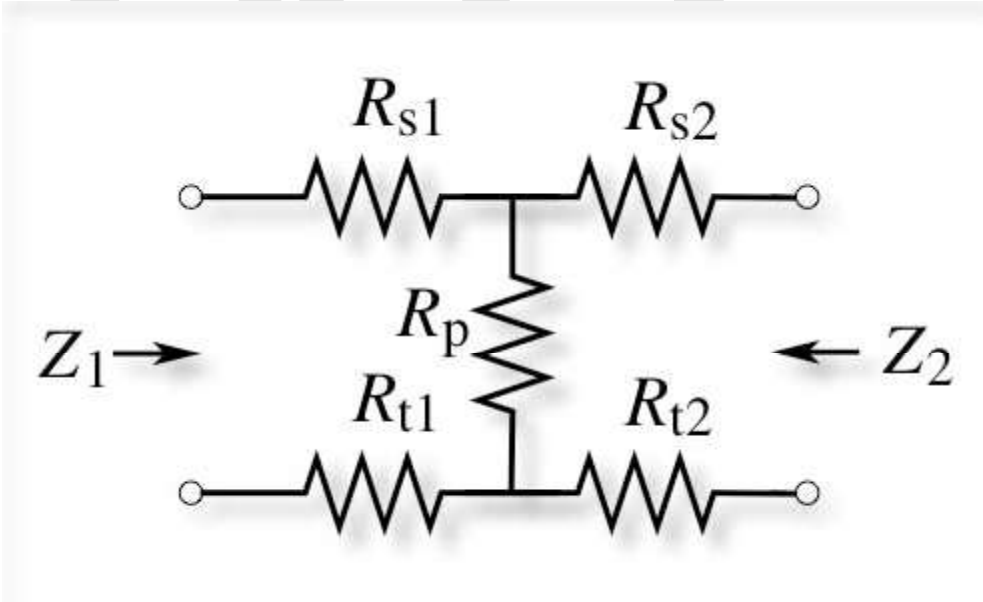
π-type unbalanced attenuator circuit



π-type balanced attenuator circuit



T-type unbalanced attenuator circuit



T-type balanced attenuator circuit

Basic circuits used in attenuators are pi pads (π-type) and T pads. These may be required to be balanced or unbalanced networks depending on whether the line geometry with which they are to be used is balanced or unbalanced. For instance, attenuators used with coaxial lines would be the unbalanced form while attenuators for use with twisted pair are required to be the balanced form.

Four fundamental attenuator circuit diagrams are given in the figures on the left. Since an attenuator circuit consists solely of passive resistor elements, it is linear and reciprocal. If the circuit is also made symmetrical (this is usually the case since it is usually required that the input and output impedances Z_1 and Z_2 are equal) then the input and output ports are not distinguished, but by convention the left and right sides of the circuits are referred to as input and output.

Attenuator characteristics

A RF Microwave Attenuator. Picture courtesy of Herley

Key specifications for attenuators are:

- **Attenuation** expressed in decibels of relative power. As a rule of thumb, a 3dB pad reduces power to one half, 6dB to one fourth, 10dB to one tenth, 20dB to one hundredth, 30dB to one thousandth and so on. For voltage you double the dBs so for example 6dB is half in voltage.
- **Frequency bandwidth**, for example DC-18 GHz
- **Power dissipation** depends on mass and surface area of resistance material as well as possible additional cooling fins.
- **SWR** is the standing wave ratio for input and output ports
- **Accuracy**
- **Repeatability**

RF attenuators

Radio frequency attenuators are typically coaxial in structure with precision connectors as ports and coaxial, microstrip or thin-film internal structure. Above SHF special waveguide structure is required.

Important characteristics are:

- accuracy,
- low SWR,

- flat frequency-response and
- repeatability.

The size and shape of the attenuator depends on its ability to dissipate power. RF attenuators are used as loads for and as known attenuations and protective dissipations of power in measuring RF signals.

Audio attenuators

A line-level attenuator in the preamp or a power attenuator after the power amplifier uses electrical resistance to reduce the amplitude of the signal that reaches the speaker, reducing the volume of the output. A line-level attenuator has lower power handling, such as a 1/2-watt potentiometer or voltage divider and controls preamp level signals, whereas a power attenuator has higher power handling capability, such as 10 watts or more, and is used between the power amplifier and the speaker.

- Power attenuator (guitar)
- Guitar amplifier

Component values for resistive pads and attenuators

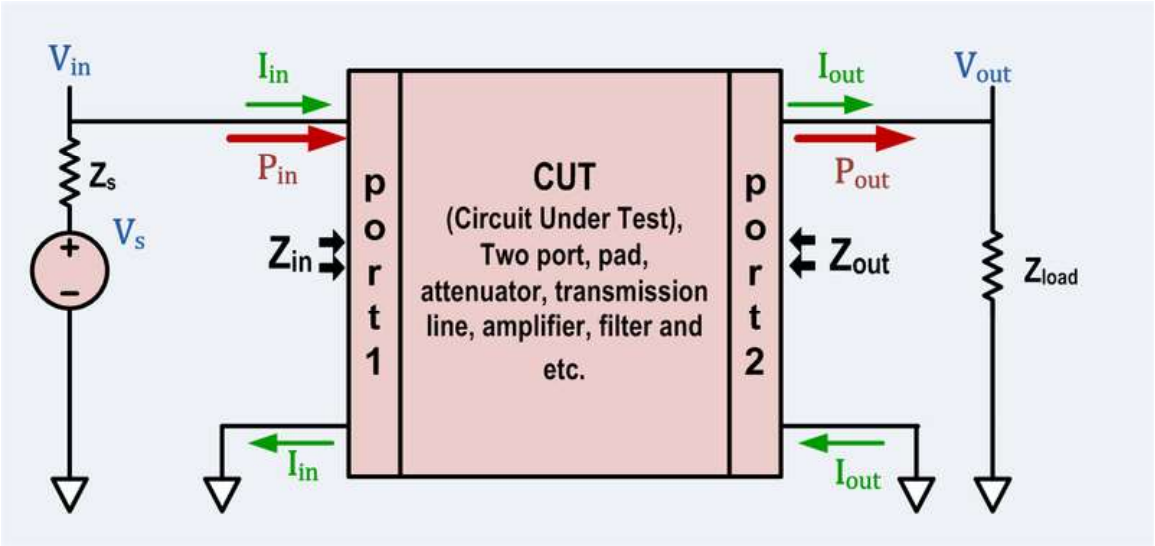
This section concerns pi-pads, T-pads and L-pads made entirely from resistors and terminated on each port with a purely real resistance.

- All impedances, currents, voltages and two-port parameters will be assumed to be purely real. For practical applications, this assumption is often close enough.

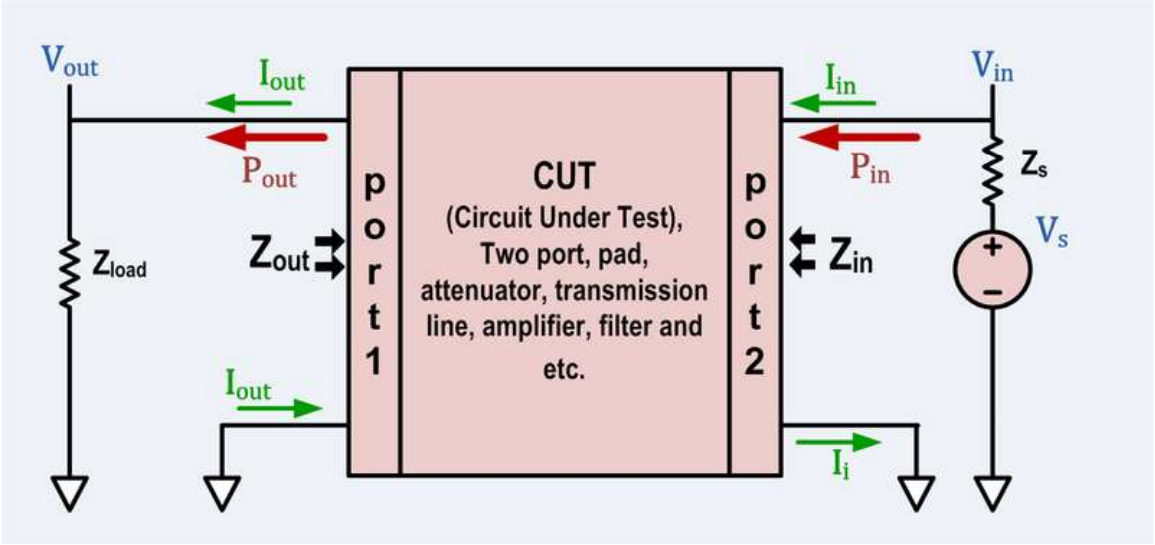
- The pad is designed for a particular load impedance, Z_{Load}, and a particular source impedance, Z_s.

    - The impedance seen looking into the input port will be Z_S if the output port is terminated by Z_{Load}.
    - The impedance seen looking into the output port will be Z_{Load} if the input port is terminated by Z_S.

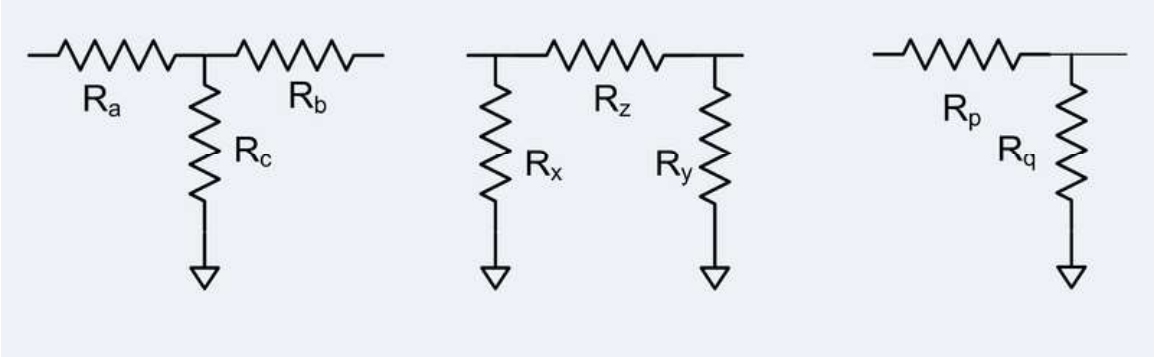
Reference figures for attenuator component calculation



This circuit is used for the general case, all T-pads, all pi-pads and L-pads when the source impedance is greater than or equal to the load impedance.



The L-pad computation assumes that port 1 has the highest impedance. If the highest impedance happens to be the output port, then use this figure.



Unique resistor designations for Tee, Pi and L pads.

The attenuator two-port is generally bidirectional. However here it will be treated as though it were one way. In general, either of the two figures above applies, but the figure on the left (which depicts the source on the left) will be tacitly assumed most of the time. In the case of the L-pad, the right figure will be used if the load impedance is greater than the source impedance.

Each resistor in each type of pad discussed is given a unique designation to decrease confusion.

The L-pad component value calculation assumes that the design impedance for port 1 (on the left) is equal or higher than the design impedance for port 2.

Terms used

- Pad will include pi-pad, T-pad, L-pad, attenuator, and two-port.
- Two-port will include pi-pad, T-pad, L-pad, attenuator, and two-port.
- Input port will mean the input port of the two-port.
- Output port will mean the output port of the two-port.
- Symmetric means a case where the source and load have equal impedance.
- Loss means the ratio of power entering the input port of the pad divided by the power absorbed by the load.
- Insertion Loss means the ratio of power that would be delivered to the load if the load were directly connected to the source divided by the power absorbed by the load when connected through the pad.

Symbols used

Passive, resistive pads and attenuators are bidirectional two-ports, but here they will be treated as unidirectional.

- Z_S = the output impedance of the source.
- Z_{Load} = the input impedance of the load.

- Z_{in} = the impedance seen looking into the input port when Z_{Load} is connected to the output port. Z_{in} is a function of the load impedance.
- Z_{out} = the impedance seen looking into the output port when Z_s is connected to the input port. Z_{out} is a function of the source impedance.
- V_s = source open circuit or unloaded voltage.
- V_{in} = voltage applied to the input port by the source.
- V_{out} = voltage applied to the load by the output port.
- I_{in} = current entering the input port from the source.
- I_{out} = current entering the load from the output port.
- $P_{in} = V_{in}\, I_{in}$ = power entering the input port from the source.
- $P_{out} = V_{out}\, I_{out}$ = power absorbed by the load from the output port.
- P_{direct} = the power that would be absorbed by the load if the load were connected directly to the source.
- $L_{pad} = 10 \log_{10} (P_{in} / P_{out})$ always. And if $Z_s = Z_{Load}$ then $L_{pad} = 20 \log_{10} (V_{in} / V_{out})$ also. Note, as defined, Loss ≥ 0 dB
- $L_{insertion} = 10 \log_{10} (P_{direct} / P_{out})$. And if $Z_s = Z_{Load}$ then $L_{insertion} = L_{pad}$.
- Loss $\equiv L_{pad}$. Loss is defined to be L_{pad}.

Symmetric T pad resistor calculation

$$A = 10^{-Loss/20} \qquad R_a = R_b = Z_S \frac{1-A}{1+A} \qquad R_c = \frac{Z_s^2 - R_b^2}{2R_b}$$

Symmetric pi pad resistor calculation

$$A = 10^{-Loss/20} \qquad R_x = R_y = Z_S \frac{1+A}{1-A} \qquad R_z = \frac{2R_x}{\left(\frac{R_x}{Z_0}\right)^2 - 1}]$$

L-Pad for impedance matching resistor calculation

If a source and load are both resistive (i.e. Z_1 and Z_2 have zero or very small imaginary part) then a resistive L-pad can be used to match them to each other. As shown, either side of the L-pad can be the source or load, but the Z_1 side must be the side with the higher impedance.

$$R_q = \frac{Z_m}{\sqrt{\rho - 1}} \qquad R_p = Z_m \sqrt{\rho - 1} \qquad Loss = 20 \log_{10} \left(\sqrt{\rho - 1} + \sqrt{\rho} \right) \quad \text{where} \quad \rho = \frac{Z_1}{Z_2} \quad Z_m = \sqrt{Z_1 Z_2}$$

Large positive numbers means loss is large. The loss is a monotonic function of the impedance ratio. Higher ratios require higher loss.

Converting T-pad to pi-pad

$$R_z = \frac{R_a R_b + R_a R_c + R_b R_c}{R_c} \qquad R_x = \frac{R_a R_b + R_a R_c + R_b R_c}{R_b} \qquad R_y = \frac{R_a R_b + R_a R_c + R_b R_c}{R_a}.$$

Converting pi-pad to T-pad

$$R_c = \frac{R_x R_y}{R_x + R_y + R_z} \qquad R_a = \frac{R_z R_x}{R_x + R_y + R_z} \qquad R_b = \frac{R_z R_y}{R_x + R_y + R_z}$$

Conversion between two-ports and pads

T-pad to impedance parameters

The impedance parameters for a passive two-port are

$$V_1 = Z_{11} I_1 + Z_{12} I_2 \qquad V_2 = Z_{21} I_1 + Z_{22} I_2 \qquad with \qquad Z_{12} = Z_{21}$$

It is always possible to represent a resistive t-pad as a two-port. The representation is particularly simple using impedance parameters as follows:

$$Z_{21} = R_c \qquad Z_{11} = R_c + R_a \qquad Z_{22} = R_c + R_b$$

Impedance parameters to T-pad

The preceding equations are trivially invertible, but if the loss is not enough, some of the t-pad components will have negative resistances.

$$R_c = Z_{21} \qquad R_a = Z_{11} - Z_{21} \qquad R_b = Z_{22} - Z_{21}$$

Impedance parameters to pi-pad

These preceding T-pad parameters can be algebraically converted to pi-pad parameters.

$$R_z = \frac{Z_{11} Z_{22} - Z_{21}^2}{Z_{21}} \qquad R_x = \frac{Z_{11} Z_{22} - Z_{21}^2}{Z_{22} - Z_{21}} \qquad R_y = \frac{Z_{11} Z_{22} - Z_{21}^2}{Z_{11} - Z_{21}}$$

Pi-pad to admittance parameters

The admittance parameters for a passive twp port are

$$I_1 = Y_{11} V_1 + Y_{12} V_2 \qquad I_2 = Y_{21} V_1 + Y_{22} V_2 \qquad with \qquad Y_{12} = Y_{21}$$

It is always possible to represent a resistive pi pad as a two-port. The representation is particularly simple using admittance parameters as follows:

$$Y_{21} = \frac{1}{R_z} \qquad Y_{11} = \frac{1}{R_x} + \frac{1}{R_z} \qquad Y_{22} = \frac{1}{R_y} + \frac{1}{R_z}$$

Admittance parameters to pi-pad

The preceding equations are trivially invertible, but if the loss is not enough, some of the pi-pad components will have negative resistances.

$$R_z = \frac{1}{Y_{21}} \qquad R_x = \frac{1}{Y_{11} - Y_{21}} \qquad R_y = \frac{1}{Y_{22} - Y_{21}}$$

General case, determining impedance parameters from requirements

Because the pad is entirely made from resistors, it must have a certain minimum loss to match source and load if they are not equal.

The minimum loss is given by

$$Loss_{min} = 20\ log_{10}\left(\sqrt{\rho - 1} + \sqrt{\rho}\ \right) \quad \text{where} \quad \rho = \frac{\max[Z_S, Z_{Load}]}{\min[Z_S, Z_{Load}]}$$

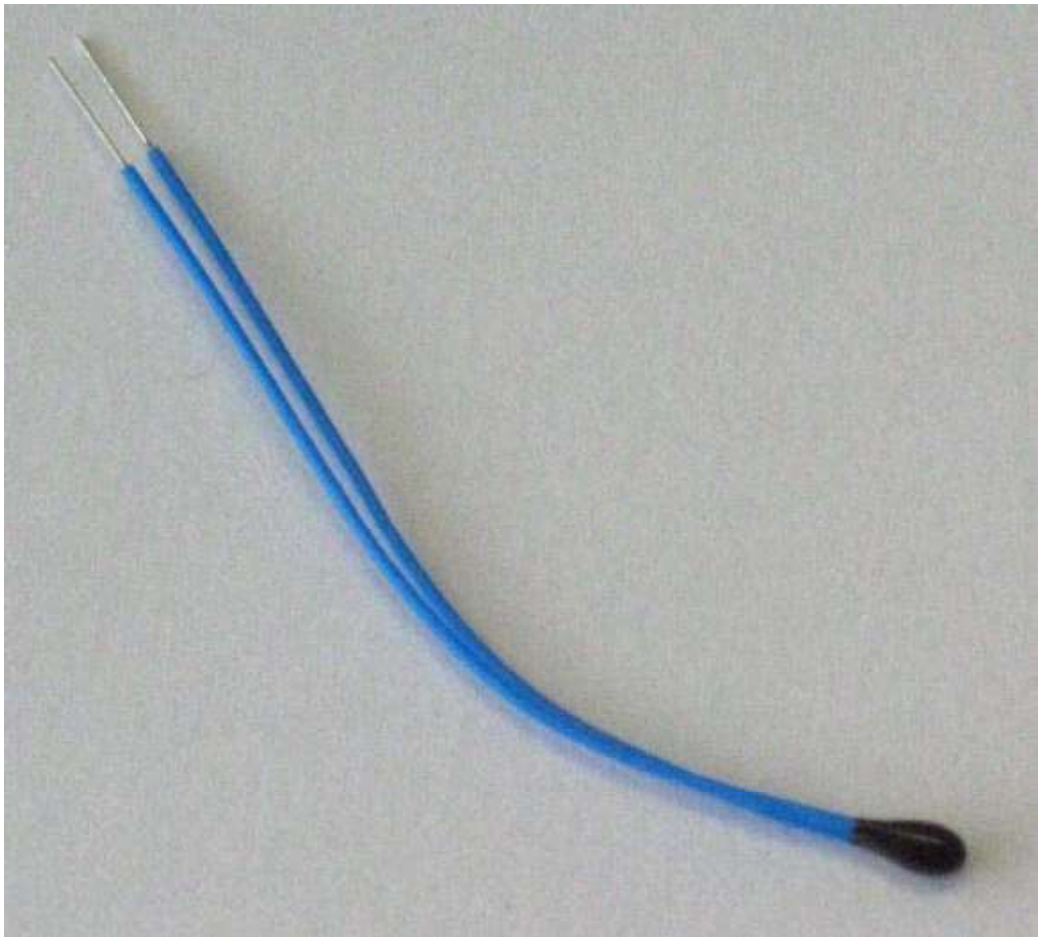
Although a passive matching two-port can have less loss, if it does it will not be convertible to a resistive attenuator pad.

$$A = 10^{-Loss/20} \qquad Z_{11} = Z_S \frac{1 + A^2}{1 - A^2} \qquad Z_{22} = Z_{Load} \frac{1 + A^2}{1 - A^2} \qquad Z_{21} = 2\frac{A\sqrt{Z_S Z_{Load}}}{1 - A^2}$$

Once these parameters have been determined, they can be implemented as a T or pi pad as discussed above.

Chapter 8

Thermistor

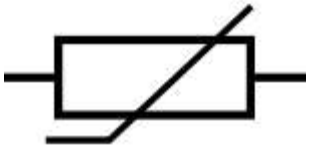


NTC thermistor, bead type, insulated wires

A **thermistor** is a type of resistor whose resistance varies significantly with temperature, more so than in standard resistors. The word is a portmanteau of *thermal* and *resistor*. Thermistors are widely used as inrush current limiters, temperature sensors, self-resetting overcurrent protectors, and self-regulating heating elements.

Thermistors differ from resistance temperature detectors (RTD) in that the material used in a thermistor is generally a ceramic or polymer, while RTDs use pure metals. The temperature

response is also different; RTDs are useful over larger temperature ranges, while thermistors typically achieve a higher precision within a limited temperature range [usually −90 °C to 130 °C].



Thermistor symbol

Assuming, as a first-order approximation, that the relationship between resistance and temperature is linear, then:

$$\Delta R = k\Delta T$$

where

ΔR = change in resistance
ΔT = change in temperature
k = first-order temperature coefficient of resistance

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

Instead of the temperature coefficient k, sometimes the *temperature coefficient of resistance* α (alpha) or α_T is used. It is defined as

$$\alpha_T = \frac{1}{R(T)}\frac{dR}{dT}.$$

For example, for the common PT100 sensor, α = 0.00385 or 0.385 %/°C. This α_T coefficient should not be confused with the α parameter below.

Steinhart-Hart equation

In practice, the linear approximation (above) works only over a small temperature range. For accurate temperature measurements, the resistance/temperature curve of the device must be

described in more detail. The Steinhart-Hart equation is a widely used third-order approximation:

$$\frac{1}{T} = a + b\,\ln(R) + c\,\ln^3(R)$$

where *a*, *b* and *c* are called the Steinhart-Hart parameters, and must be specified for each device. *T* is the temperature in kelvin and *R* is the resistance in ohms. To give resistance as a function of temperature, the above can be rearranged into:

$$R = e^{\left(x-\frac{y}{2}\right)^{\frac{1}{3}}-\left(x+\frac{y}{2}\right)^{\frac{1}{3}}}$$

where

$$y = \frac{a-\frac{1}{T}}{c} \quad \text{and} \quad x = \sqrt{\left(\frac{b}{3c}\right)^3 + \frac{y^2}{4}}$$

The error in the Steinhart-Hart equation is generally less than 0.02 °C in the measurement of temperature. As an example, typical values for a thermistor with a resistance of 3000 Ω at room temperature (25 °C = 298.15 K) are:

$$a = 1.40 \times 10^{-3}$$
$$b = 2.37 \times 10^{-4}$$
$$c = 9.90 \times 10^{-8}$$

B *parameter equation*

NTC thermistors can also be characterised with the *B* parameter equation, which is essentially the Steinhart Hart equation with $a = (1 / T_0) - (1 / B)ln(R_0)$, $b = 1 / B$ and $c = 0$,

$$\frac{1}{T} = \frac{1}{T_0} + \frac{1}{B}\ln\left(\frac{R}{R_0}\right)$$

Where the temperatures are in kelvins and R_0 is the resistance at temperature T_0 (usually 25 °C = 298.15 K). Solving for *R* yields:

$$R = R_0 e^{B(1/T - 1/T_0)}$$

or, alternatively,

$$R = r_\infty e^{B/T}$$

where $r_\infty = R_0 e^{-B/T_0}$. This can be solved for the temperature:

$$T = \frac{B}{\ln\left(R/r_\infty\right)}$$

The B-parameter equation can also be written as $\ln R = B/T + \ln r_\infty$. This can be used to convert the function of resistance vs. temperature of a thermistor into a linear function of ln*R* vs. 1 / *T*. The average slope of this function will then yield an estimate of the value of the *B* parameter.

Conduction model

Many NTC thermistors are made from a pressed disc or cast chip of a semiconductor such as a sintered metal oxide. They work because raising the temperature of a semiconductor increases the number of electrons able to move about and carry charge - it promotes them into the *conduction band*. The more charge carriers that are available, the more current a material can conduct. This is described in the formula:

$$I = n \cdot A \cdot v \cdot e$$

I = electric current (amperes)
n = density of charge carriers (count/m³)
A = cross-sectional area of the material (m²)
v = velocity of charge carriers (m/s)
e = charge of an electron ($e = 1.602 \times 10^{-19}$ coulomb)

The current is measured using an ammeter. Over large changes in temperature, calibration is necessary. Over small changes in temperature, if the right semiconductor is used, the resistance of the material is linearly proportional to the temperature. There are many different semiconducting thermistors with a range from about 0.01 kelvin to 2,000 kelvins (−273.14 °C to 1,700 °C).

Most PTC thermistors are of the "switching" type, which means that their resistance rises suddenly at a certain critical temperature. The devices are made of a doped polycrystalline ceramic containing barium titanate ($BaTiO_3$) and other compounds. The dielectric constant of this ferroelectric material varies with temperature. Below the Curie point temperature, the high dielectric constant prevents the formation of potential barriers between the crystal grains, leading to a low resistance. In this region the device has a small negative temperature coefficient. At the Curie point temperature, the dielectric constant drops sufficiently to allow the formation of potential barriers at the grain boundaries, and the resistance increases sharply. At even higher temperatures, the material reverts to NTC behaviour. The equations used for modeling this behaviour were derived by W. Heywang and G. H. Jonker in the 1960s.

Another type of PTC thermistor is the polymer PTC, which is sold under brand names such as "Polyswitch" "Semifuse", and "Multifuse". This consists of a slice of plastic with carbon grains

embedded in it. When the plastic is cool, the carbon grains are all in contact with each other, forming a conductive path through the device. When the plastic heats up, it expands, forcing the carbon grains apart, and causing the resistance of the device to rise rapidly. Like the $BaTiO_3$ thermistor, this device has a highly nonlinear resistance/temperature response and is used for switching, not for proportional temperature measurement.

Yet another type of thermistor is a **silistor**, a thermally sensitive silicon resistor. Silistors are similarly constructed and operate on the same principles as other thermistors, but employ silicon as the semiconductive component material.

Self-heating effects

When a current flows through a thermistor, it will generate heat which will raise the temperature of the thermistor above that of its environment. If the thermistor is being used to measure the temperature of the environment, this electrical heating may introduce a significant error if a correction is not made. Alternatively, this effect itself can be exploited. It can, for example, make a sensitive air-flow device employed in a sailplane rate-of-climb instrument, the electronic variometer, or serve as a timer for a relay as was formerly done in telephone exchanges.

The electrical power input to the thermistor is just:

$$P_E = IV$$

where I is current and V is the voltage drop across the thermistor. This power is converted to heat, and this heat energy is transferred to the surrounding environment. The rate of transfer is well described by Newton's law of cooling:

$$P_T = K(T(R) - T_0)$$

where $T(R)$ is the temperature of the thermistor as a function of its resistance R, T_0 is the temperature of the surroundings, and K is the **dissipation constant**, usually expressed in units of milliwatts per degree Celsius. At equilibrium, the two rates must be equal.

$$P_E = P_T$$

The current and voltage across the thermistor will depend on the particular circuit configuration. As a simple example, if the voltage across the thermistor is held fixed, then by Ohm's Law we have $I = V / R$ and the equilibrium equation can be solved for the ambient temperature as a function of the measured resistance of the thermistor:

$$T_0 = T(R) - \frac{V^2}{KR}$$

The dissipation constant is a measure of the thermal connection of the thermistor to its surroundings. It is generally given for the thermistor in still air, and in well-stirred oil. Typical values for a small glass bead thermistor are 1.5 mW/°C in still air and 6.0 mW/°C in stirred oil.

If the temperature of the environment is known beforehand, then a thermistor may be used to measure the value of the dissipation constant. For example, the thermistor may be used as a flow rate sensor, since the dissipation constant increases with the rate of flow of a fluid past the thermistor.

Applications

- PTC thermistors can be used as current-limiting devices for circuit protection, as replacements for fuses. Current through the device causes a small amount of resistive heating. If the current is large enough to generate more heat than the device can lose to its surroundings, the device heats up, causing its resistance to increase, and therefore causing even more heating. This creates a self-reinforcing effect that drives the resistance upwards, reducing the current and voltage available to the device.
- PTC thermistors are used as timers in the degaussing coil circuit of most CRT displays and televisions. When the display unit is initially switched on, current flows through the thermistor and degaussing coil. The coil and thermistor are intentionally sized so that the current flow will heat the thermistor to the point that the degaussing coil shuts off in under a second. For effective degaussing, it is necessary that the magnitude of the alternating magnetic field produced by the degaussing coil decreases smoothly and continuously, rather than sharply switching off or decreasing in steps; the PTC thermistor accomplishes this naturally as it heats up. A degaussing circuit using a PTC thermistor is simple, reliable (for its simplicity), and inexpensive.
- NTC thermistors are used as resistance thermometers in low-temperature measurements of the order of 10 K.
- NTC thermistors can be used as inrush-current limiting devices in power supply circuits. They present a higher resistance initially which prevents large currents from flowing at turn-on, and then heat up and become much lower resistance to allow higher current flow during normal operation. These thermistors are usually much larger than measuring type thermistors, and are purposely designed for this application.
- NTC thermistors are regularly used in automotive applications. For example, they monitor things like coolant temperature and/or oil temperature inside the engine and provide data to the ECU and, indirectly, to the dashboard.
- NTC thermistors can be also used to monitor the temperature of an incubator.
- Thermistors are also commonly used in modern digital thermostats and to monitor the temperature of battery packs while charging.

History

The first NTC thermistor was discovered in 1833 by Michael Faraday, who reported on the semiconducting behavior of silver sulfide. Faraday noticed that the resistance of silver sulfide decreased dramatically as temperature increased. Because early thermistors were difficult to produce and applications for the technology were limited, commercial production of thermistors did not begin until the 1930s.

Chapter 9

Capacitor

Capacitor

Modern capacitors, by a cm rule

Type Passive

Invented Ewald Georg von Kleist (October 1745)

Electronic symbol

A **capacitor** (formerly known as **condenser**) is a passive electronic component consisting of a pair of conductors separated by a dielectric (insulator). When there is a potential difference (voltage) across the conductors, a static electric field develops in the dielectric that stores energy and produces a mechanical force between the conductors. An ideal capacitor is characterized by a single constant value, capacitance, measured in farads. This is the ratio of the electric charge on each conductor to the potential difference between them.

Capacitors are widely used in electronic circuits for blocking direct current while allowing alternating current to pass, in filter networks, for smoothing the output of power supplies, in the resonant circuits that tune radios to particular frequencies and for many other purposes.

A typical electrolytic capacitor

The effect is greatest when there is a narrow separation between large areas of conductor, hence capacitor conductors are often called "plates", referring to an early means of construction. In practice the dielectric between the plates passes a small amount of leakage current and also has an electric field strength limit, resulting in a breakdown voltage, while the conductors and leads introduce an undesired inductance and resistance.

History

Battery of four Leyden jars in Museum Boerhaave, Leiden, the Netherlands.

In October 1745, Ewald Georg von Kleist of Pomerania in Germany found that charge could be stored by connecting a high voltage electrostatic generator by a wire to a volume of water in a hand-held glass jar. Von Kleist's hand and the water acted as conductors and the jar as a dielectric (although details of the mechanism were incorrectly identified at the time). Von Kleist found, after removing the generator, that touching the wire resulted in a painful spark. In a letter describing the experiment, he said "I would not take a second shock for the kingdom of France." The following year, the Dutch physicist Pieter van Musschenbroek invented a similar capacitor, which was named the Leyden jar, after the University of Leiden where he worked.

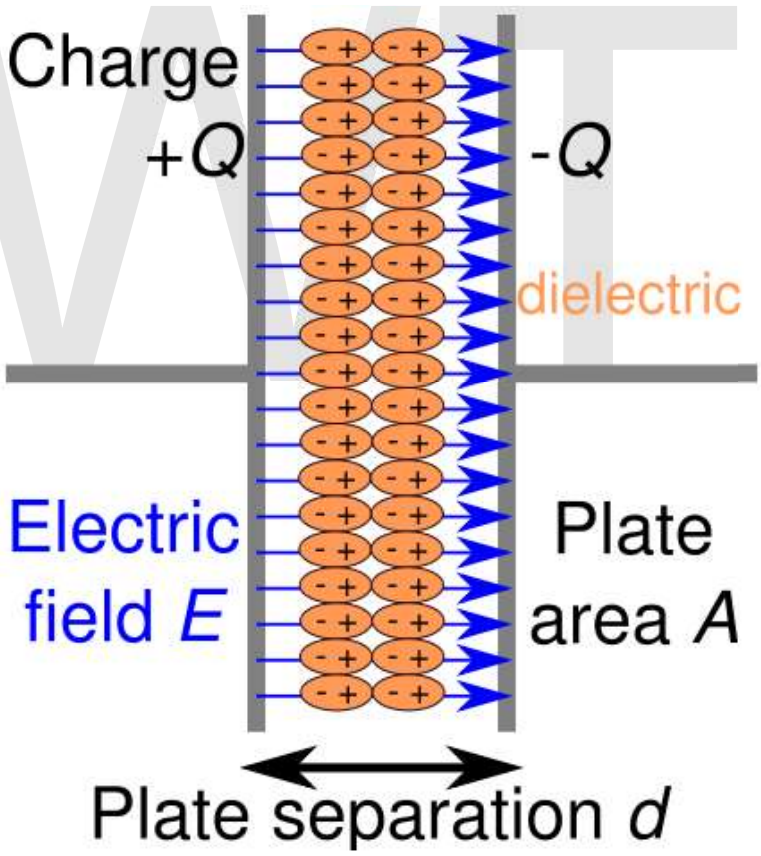
Daniel Gralath was the first to combine several jars in parallel into a "battery" to increase the charge storage capacity. Benjamin Franklin investigated the Leyden jar and "proved" that the charge was stored on the glass, not in the water as others had assumed. He also adopted the term "battery", (denoting the increasing of power with a row of similar units as in a battery of

cannon), subsequently applied to clusters of electrochemical cells. Leyden jars were later made by coating the inside and outside of jars with metal foil, leaving a space at the mouth to prevent arcing between the foils. The earliest unit of capacitance was the 'jar', equivalent to about 1 nanofarad.

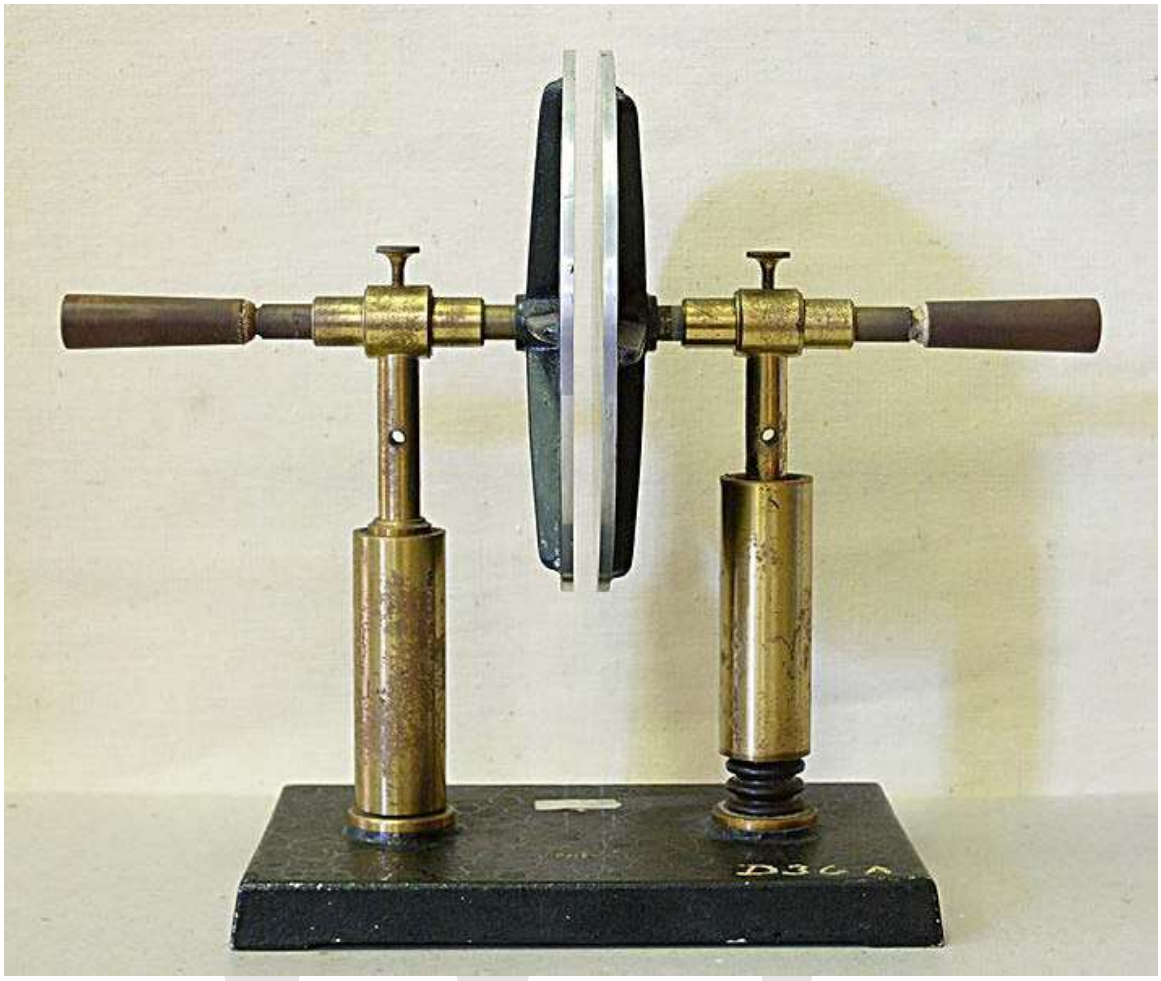
Leyden jars or more powerful devices employing flat glass plates alternating with foil conductors were used exclusively up until about 1900, when the invention of wireless (radio) created a demand for standard capacitors, and the steady move to higher frequencies required capacitors with lower inductance. A more compact construction began to be used of a flexible dielectric sheet such as oiled paper sandwiched between sheets of metal foil, rolled or folded into a small package.

Early capacitors were also known as *condensers*, a term that is still occasionally used today. The term was first used for this purpose by Alessandro Volta in 1782, with reference to the device's ability to store a higher density of electric charge than a normal isolated conductor.

Theory of operation



Charge separation in a parallel-plate capacitor causes an internal electric field. A dielectric (orange) reduces the field and increases the capacitance.



A simple demonstration of a parallel-plate capacitor

A capacitor consists of two conductors separated by a non-conductive region called the dielectric medium though it may be a vacuum or a semiconductor depletion region chemically identical to the conductors. A capacitor is assumed to be self-contained and isolated, with no net electric charge and no influence from any external electric field. The conductors thus hold equal and opposite charges on their facing surfaces, and the dielectric develops an electric field. In SI units, a capacitance of one farad means that one coulomb of charge on each conductor causes a voltage of one volt across the device.

The capacitor is a reasonably general model for electric fields within electric circuits. An ideal capacitor is wholly characterized by a constant capacitance C, defined as the ratio of charge $\pm Q$ on each conductor to the voltage V between them:

$$C = \frac{Q}{V}$$

Sometimes charge build-up affects the capacitor mechanically, causing its capacitance to vary. In this case, capacitance is defined in terms of incremental changes:

$$C = \frac{\mathrm{d}q}{\mathrm{d}v}$$

Energy storage

Work must be done by an external influence to "move" charge between the conductors in a capacitor. When the external influence is removed the charge separation persists in the electric field and energy is stored to be released when the charge is allowed to return to its equilibrium position. The work done in establishing the electric field, and hence the amount of energy stored, is given by:

$$W = \int_{q=0}^{Q} V\,\mathrm{d}q = \int_{q=0}^{Q} \frac{q}{C}\mathrm{d}q = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}CV^2 = \frac{1}{2}VQ.$$

Current-voltage relation

The current $i(t)$ through any component in an electric circuit is defined as the rate of flow of a charge $q(t)$ passing through it, but actual charges, electrons, cannot pass through the dielectric layer of a capacitor, rather an electron accumulates on the negative plate for each one that leaves the positive plate, resulting in an electron depletion and consequent positive charge on one electrode that is equal and opposite to the accumulated negative charge on the other. Thus the charge on the electrodes is equal to the integral of the current as well as proportional to the voltage as discussed above. As with any antiderivative, a constant of integration is added to represent the initial voltage v (t_0). This is the integral form of the capacitor equation,

$$v(t) = \frac{q(t)}{C} = \frac{1}{C}\int_{t_0}^{t} i(\tau)\mathrm{d}\tau + v(t_0)$$

.

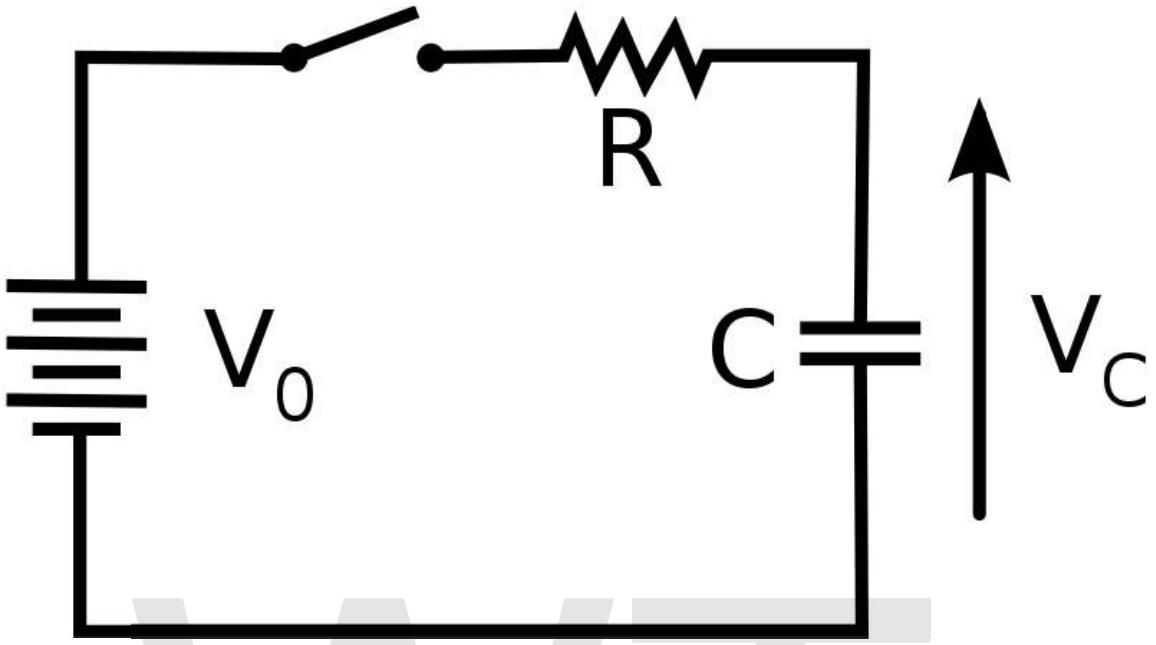
Taking the derivative of this, and multiplying by C, yields the derivative form,

$$i(t) = \frac{\mathrm{d}q(t)}{\mathrm{d}t} = C\frac{\mathrm{d}v(t)}{\mathrm{d}t}$$

.

The dual of the capacitor is the inductor, which stores energy in the magnetic field rather than the electric field. Its current-voltage relation is obtained by exchanging current and voltage in the capacitor equations and replacing C with the inductance L.

DC circuits



A simple resistor-capacitor circuit demonstrates charging of a capacitor.

A series circuit containing only a resistor, a capacitor, a switch and a constant DC source of voltage V_0 is known as a *charging circuit*. If the capacitor is initially uncharged while the switch is open, and the switch is closed at $t = 0$, it follows from Kirchhoff's voltage law that

$$V_0 = v_{\mathrm{resistor}}(t) + v_{\mathrm{capacitor}}(t) = i(t)R + \frac{1}{C}\int_0^t i(\tau)\mathrm{d}\tau.$$

Taking the derivative and multiplying by C, gives a first-order differential equation,

$$RC\frac{\mathrm{d}i(t)}{\mathrm{d}t} + i(t) = 0.$$

At $t = 0$, the voltage across the capacitor is zero and the voltage across the resistor is V_0. The initial current is then $i\,(0) = V_0\,/R$. With this assumption, the differential equation yields

$$i(t) = \frac{V_0}{R}e^{-t/\tau_0}$$
$$v(t) = V_0\left(1 - e^{-t/\tau_0}\right),$$

where $\tau_0 = RC$ is the *time constant* of the system.

As the capacitor reaches equilibrium with the source voltage, the voltage across the resistor and the current through the entire circuit decay exponentially. The case of *discharging* a charged capacitor likewise demonstrates exponential decay, but with the initial capacitor voltage replacing V_0 and the final voltage being zero.

AC circuits

Impedance, the vector sum of reactance and resistance, describes the phase difference and the ratio of amplitudes between sinusoidally varying voltage and sinusoidally varying current at a given frequency. Fourier analysis allows any signal to be constructed from a spectrum of frequencies, whence the circuit's reaction to the various frequencies may be found. The reactance and impedance of a capacitor are respectively

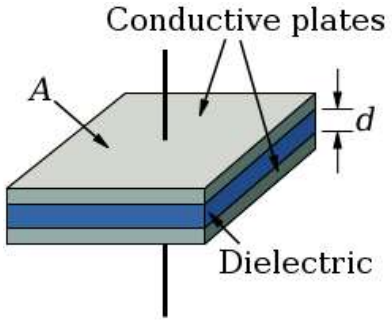
$$X = -\frac{1}{\omega C} = -\frac{1}{2\pi f C}$$
$$Z = \frac{1}{j\omega C} = -\frac{j}{\omega C} = -\frac{j}{2\pi f C}$$

where j is the imaginary unit and ω is the angular velocity of the sinusoidal signal. The $-j$ phase indicates that the AC voltage $V = Z\,I$ lags the AC current by 90°: the positive current phase corresponds to increasing voltage as the capacitor charges; zero current corresponds to instantaneous constant voltage, etc.

Note that impedance decreases with increasing capacitance and increasing frequency. This implies that a higher-frequency signal or a larger capacitor results in a lower voltage amplitude per current amplitude—an AC "short circuit" or AC coupling. Conversely, for very low frequencies, the reactance will be high, so that a capacitor is nearly an open circuit in AC analysis—those frequencies have been "filtered out".

Capacitors are different from resistors and inductors in that the impedance is *inversely* proportional to the defining characteristic, i.e. capacitance.

Parallel plate model



Dielectric is placed between two conducting plates, each of area A and with a separation of d.

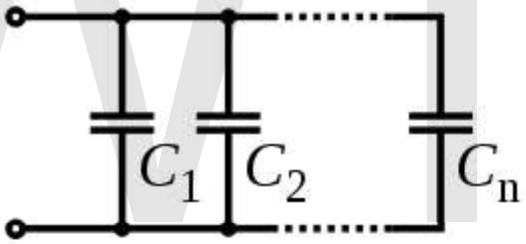
The simplest capacitor consists of two parallel conductive plates separated by a dielectric with permittivity ε (such as air). The model may also be used to make qualitative predictions for other device geometries. The plates are considered to extend uniformly over an area A and a charge density $\pm\rho = \pm Q/A$ exists on their surface. Assuming that the width of the plates is much greater than their separation d, the electric field near the centre of the device will be uniform with the magnitude $E = \rho/\varepsilon$. The voltage is defined as the line integral of the electric field between the plates

$$V = \int_0^d E\mathrm{d}z = \int_0^d \frac{\rho}{\varepsilon}\mathrm{d}z = \frac{\rho d}{\varepsilon} = \frac{Qd}{\varepsilon A}.$$

Solving this for $C = Q/V$ reveals that capacitance increases with area and decreases with separation

$$C = \frac{\varepsilon A}{d}.$$

The capacitance is therefore greatest in devices made from materials with a high permittivity.



Several capacitors in parallel.

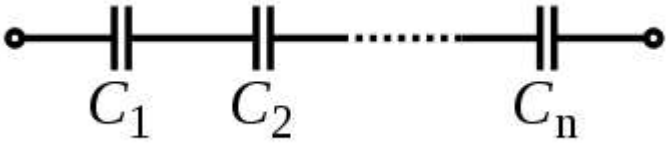
Networks

For capacitors in parallel

Capacitors in a parallel configuration each have the same applied voltage. Their capacitances add up. Charge is apportioned among them by size. Using the schematic diagram to visualize parallel plates, it is apparent that each capacitor contributes to the total surface area.

$$C_{eq} = C_1 + C_2 + \cdots + C_n$$

For capacitors in series



Several capacitors in series.

Connected in series, the schematic diagram reveals that the separation distance, not the plate area, adds up. The capacitors each store instantaneous charge build-up equal to that of every other capacitor in the series. The total voltage difference from end to end is apportioned to each capacitor according to the inverse of its capacitance. The entire series acts as a capacitor *smaller* than any of its components.

$$\frac{1}{C_{eq}} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_n}$$

Capacitors are combined in series to achieve a higher working voltage, for example for smoothing a high voltage power supply. The voltage ratings, which are based on plate separation, add up. In such an application, several series connections may in turn be connected in parallel, forming a matrix. The goal is to maximize the energy storage utility of each capacitor without overloading it.

Series connection is also used to adapt electrolytic capacitors for AC use.

Non-ideal behaviour

Capacitors deviate from the ideal capacitor equation in a number of ways. Some of these, such as leakage current and parasitic effects are linear, or can be assumed to be linear, and can be dealt with by adding virtual components to the equivalent circuit of the capacitor. The usual methods of network analysis can then be applied. In other cases, such as with breakdown voltage, the effect is non-linear and normal (i.e., linear) network analysis cannot be used, the effect must be dealt with separately. There is yet another group, which may be linear but invalidate the assumption in the analysis that capacitance is a constant. Such an example is temperature dependence.

Breakdown voltage

Above a particular electric field, known as the dielectric strength E_{ds}, the dielectric in a capacitor becomes conductive. The voltage at which this occurs is called the breakdown voltage of the device, and is given by the product of the dielectric strength and the separation between the conductors,

$$V_{\mathrm{bd}} = E_{\mathrm{ds}} d$$

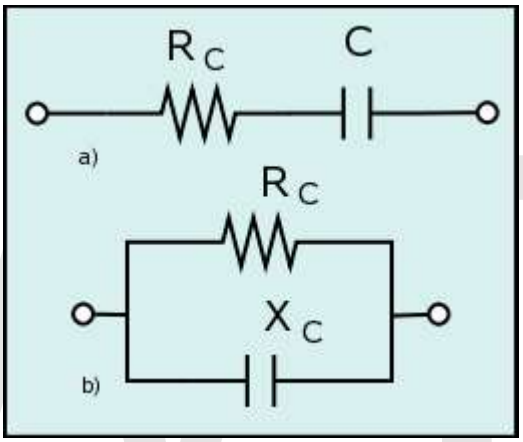
The maximum energy that can be stored safely in a capacitor is limited by the breakdown voltage. Due to the scaling of capacitance and breakdown voltage with dielectric thickness, all capacitors made with a particular dielectric have approximately equal maximum energy density, to the extent that the dielectric dominates their volume.

For air dielectric capacitors the breakdown field strength is of the order 2 to 5 MV/m; for mica the breakdown is 100 to 300 MV/m, for oil 15 to 25 MV/m, and can be much less when other materials are used for the dielectric. The dielectric is used in very thin layers and so absolute breakdown voltage of capacitors is limited. Typical ratings for capacitors used for general electronics applications range from a few volts to 100V or so. As the voltage increases, the

dielectric must be thicker, making high-voltage capacitors larger than those rated for lower voltages. The breakdown voltage is critically affected by factors such as the geometry of the capacitor conductive parts; sharp edges or points increase the electric field strength at that point and can lead to a local breakdown. Once this starts to happen, the breakdown will quickly "track" through the dielectric till it reaches the opposite plate and cause a short circuit.

The usual breakdown route is that the field strength becomes large enough to pull electrons in the dielectric from their atoms thus causing conduction. Other scenarios are possible, such as impurities in the dielectric, and, if the dielectric is of a crystalline nature, imperfections in the crystal structure can result in an avalanche breakdown as seen in semi-conductor devices. Breakdown voltage is also affected by pressure, humidity and temperature.

Equivalent circuit



Two different circuit models of a real capacitor

An ideal capacitor only stores and releases electrical energy, without dissipating any. In reality, all capacitors have imperfections within the capacitor's material that create resistance. This is specified as the *equivalent series resistance* or **ESR** of a component. This adds a real component to the impedance:

$$R_{\mathrm{C}} = Z + R_{\mathrm{ESR}} = \frac{1}{j\omega C} + R_{\mathrm{ESR}}$$

As frequency approaches infinity, the capacitive impedance (or reactance) approaches zero and the ESR becomes significant. As the reactance becomes negligible, power dissipation approaches $P_{\mathrm{RMS}} = V_{\mathrm{RMS}}^2 / R_{\mathrm{ESR}}$.

Similarly to ESR, the capacitor's leads add *equivalent series inductance* or **ESL** to the component. This is usually significant only at relatively high frequencies. As inductive reactance is positive and increases with frequency, above a certain frequency capacitance will be canceled by inductance. High-frequency engineering involves accounting for the inductance of all connections and components.

If the conductors are separated by a material with a small conductivity rather than a perfect dielectric, then a small leakage current flows directly between them. The capacitor therefore has a finite parallel resistance, and slowly discharges over time (time may vary greatly depending on the capacitor material and quality).

Ripple current

Ripple current is the AC component of an applied source (often a switched-mode power supply) whose frequency may be constant or varying. Certain types of capacitors, such as electrolytic tantalum capacitors, usually have a rating for maximum ripple current (both in frequency and magnitude). This ripple current can cause damaging heat to be generated within the capacitor due to the current flow across resistive imperfections in the materials used within the capacitor, more commonly referred to as equivalent series resistance (ESR). For example electrolytic tantalum capacitors are limited by ripple current and generally have the highest ESR ratings in the capacitor family, while ceramic capacitors generally have no ripple current limitation and have some of the lowest ESR ratings.

Capacitance instability

The capacitance of certain capacitors decreases as the component ages. In ceramic capacitors, this is caused by degradation of the dielectric. The type of dielectric and the ambient operating and storage temperatures are the most significant aging factors, while the operating voltage has a smaller effect. The aging process may be reversed by heating the component above the Curie point. Aging is fastest near the beginning of life of the component, and the device stabilizes over time. Electrolytic capacitors age as the electrolyte evaporates. In contrast with ceramic capacitors, this occurs towards the end of life of the component.

Temperature dependence of capacitance is usually expressed in parts per million (ppm) per °C. It can usually be taken as a broadly linear function but can be noticeably non-linear at the temperature extremes. The temperature coefficient can be either positive or negative, sometimes even amongst different samples of the same type. In other words, the spread in the range of temperature coefficients can encompass zero.

Capacitors, especially ceramic capacitors, and older designs such as paper capacitors, can absorb sound waves resulting in a microphonic effect. Vibration moves the plates, causing the capacitance to vary, in turn inducing AC current. Some dielectrics also generate piezoelectricity. The resulting interference is especially problematic in audio applications, potentially causing feedback or unintended recording. In the reverse microphonic effect, the varying electric field between the capacitor plates exerts a physical force, moving them as a speaker. This can generate audible sound, but drains energy and stresses the dielectric and the electrolyte, if any.

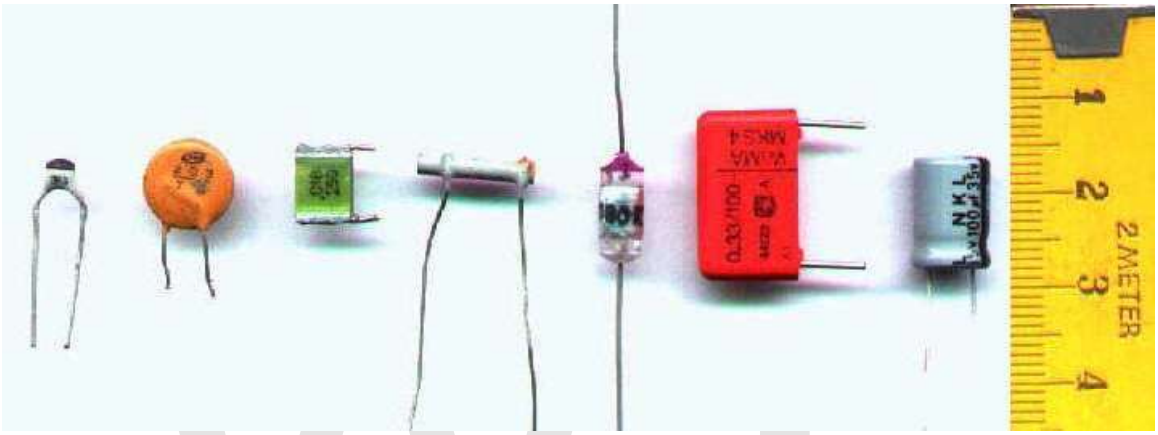
Capacitor types

Practical capacitors are available commercially in many different forms. The type of internal dielectric, the structure of the plates and the device packaging all strongly affect the characteristics of the capacitor, and its applications.

Values available range from very low (picofarad range; while arbitrarily low values are in principle possible, stray (parasitic) capacitance in any circuit is the limiting factor) to about 5 kF supercapacitors.

Above approximately 1 microfarad electrolytic capacitors are usually used because of their small size and low cost compared with other technologies, unless their relatively poor stability, life and polarised nature make them unsuitable. Very high capacity supercapacitors use a porous carbon-based electrode material.

Dielectric materials



Capacitor materials. From left: multilayer ceramic, ceramic disc, multilayer polyester film, tubular ceramic, polystyrene, metalized polyester film, aluminum electrolytic. Major scale divisions are in centimetres.

Most types of capacitor include a dielectric spacer, which increases their capacitance. These dielectrics are most often insulators. However, low capacitance devices are available with a vacuum between their plates, which allows extremely high voltage operation and low losses. Variable capacitors with their plates open to the atmosphere were commonly used in radio tuning circuits. Later designs use polymer foil dielectric between the moving and stationary plates, with no significant air space between them.

In order to maximise the charge that a capacitor can hold, the dielectric material needs to have as high a permittivity as possible, while also having as high a breakdown voltage as possible.

Several solid dielectrics are available, including paper, plastic, glass, mica and ceramic materials. Paper was used extensively in older devices and offers relatively high voltage performance. However, it is susceptible to water absorption, and has been largely replaced by plastic film capacitors. Plastics offer better stability and aging performance, which makes them useful in timer circuits, although they may be limited to low operating temperatures and frequencies. Ceramic capacitors are generally small, cheap and useful for high frequency applications, although their capacitance varies strongly with voltage and they age poorly. They are broadly categorized as class 1 dielectrics, which have predictable variation of capacitance with temperature or class 2 dielectrics, which can operate at higher voltage. Glass and mica capacitors are extremely reliable, stable and tolerant to high temperatures and voltages, but are

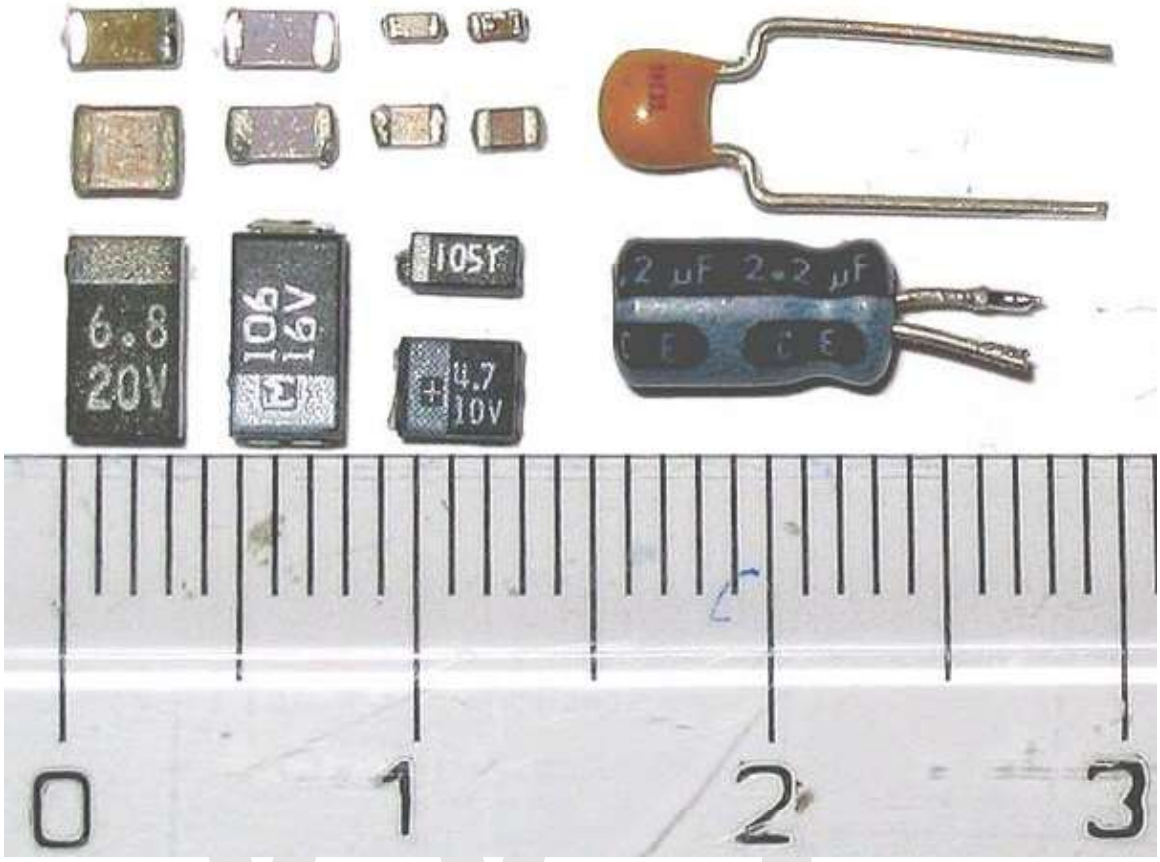
too expensive for most mainstream applications. Electrolytic capacitors and supercapacitors are used to store small and larger amounts of energy, respectively, ceramic capacitors are often used in resonators, and parasitic capacitance occurs in circuits wherever the simple conductor-insulator-conductor structure is formed unintentionally by the configuration of the circuit layout.

Electrolytic capacitors use an aluminum or tantalum plate with an oxide dielectric layer. The second electrode is a liquid electrolyte, connected to the circuit by another foil plate. Electrolytic capacitors offer very high capacitance but suffer from poor tolerances, high instability, gradual loss of capacitance especially when subjected to heat, and high leakage current. Poor quality capacitors may leak electrolyte, which is harmful to printed circuit boards. The conductivity of the electrolyte drops at low temperatures, which increases equivalent series resistance. While widely used for power-supply conditioning, poor high-frequency characteristics make them unsuitable for many applications. Electrolytic capacitors will self-degrade if unused for a period (around a year), and when full power is applied may short circuit, permanently damaging the capacitor and usually blowing a fuse or causing arcing in rectifier tubes. They can be restored before use (and damage) by gradually applying the operating voltage, often done on antique vacuum tube equipment over a period of 30 minutes by using a variable transformer to supply AC power. Unfortunately, the use of this technique may be less satisfactory for some solid state equipment, which may be damaged by operation below its normal power range, requiring that the power supply first be isolated from the consuming circuits. Such remedies may not be applicable to modern high-frequency power supplies as these produce full output voltage even with reduced input.

Tantalum capacitors offer better frequency and temperature characteristics than aluminum, but higher dielectric absorption and leakage. OS-CON (or OC-CON) capacitors are a polymerized organic semiconductor solid-electrolyte type that offer longer life at higher cost than standard electrolytic capacitors.

Several other types of capacitor are available for specialist applications. Supercapacitors store large amounts of energy. Supercapacitors made from carbon aerogel, carbon nanotubes, or highly porous electrode materials offer extremely high capacitance (up to 5 kF as of 2010) and can be used in some applications instead of rechargeable batteries. Alternating current capacitors are specifically designed to work on line (mains) voltage AC power circuits. They are commonly used in electric motor circuits and are often designed to handle large currents, so they tend to be physically large. They are usually ruggedly packaged, often in metal cases that can be easily grounded/earthed. They also are designed with direct current breakdown voltages of at least five times the maximum AC voltage.

Structure



Capacitor packages: SMD ceramic at top left; SMD tantalum at bottom left; through-hole tantalum at top right; through-hole electrolytic at bottom right. Major scale divisions are cm.

The arrangement of plates and dielectric has many variations depending on the desired ratings of the capacitor. For small values of capacitance (microfarads and less), ceramic disks use metallic coatings, with wire leads bonded to the coating. Larger values can be made by multiple stacks of plates and disks. Larger value capacitors usually use a metal foil or metal film layer deposited on the surface of a dielectric film to make the plates, and a dielectric film of impregnated paper or plastic – these are rolled up to save space. To reduce the series resistance and inductance for long plates, the plates and dielectric are staggered so that connection is made at the common edge of the rolled-up plates, not at the ends of the foil or metalized film strips that comprise the plates.

The assembly is encased to prevent moisture entering the dielectric – early radio equipment used a cardboard tube sealed with wax. Modern paper or film dielectric capacitors are dipped in a hard thermoplastic. Large capacitors for high-voltage use may have the roll form compressed to fit into a rectangular metal case, with bolted terminals and bushings for connections. The dielectric in larger capacitors is often impregnated with a liquid to improve its properties.

Capacitors may have their connecting leads arranged in many configurations, for example axially or radially. "Axial" means that the leads are on a common axis, typically the axis of the

capacitor's cylindrical body – the leads extend from opposite ends. Radial leads might more accurately be referred to as tandem; they are rarely actually aligned along radii of the body's circle, so the term is inexact, although universal. The leads (until bent) are usually in planes parallel to that of the flat body of the capacitor, and extend in the same direction; they are often parallel as manufactured.

Small, cheap discoidal ceramic capacitors have existed since the 1930s, and remain in widespread use. Since the 1980s, surface mount packages for capacitors have been widely used. These packages are extremely small and lack connecting leads, allowing them to be soldered directly onto the surface of printed circuit boards. Surface mount components avoid undesirable high-frequency effects due to the leads and simplify automated assembly, although manual handling is made difficult due to their small size.

Mechanically controlled variable capacitors allow the plate spacing to be adjusted, for example by rotating or sliding a set of movable plates into alignment with a set of stationary plates. Low cost variable capacitors squeeze together alternating layers of aluminum and plastic with a screw. Electrical control of capacitance is achievable with varactors (or varicaps), which are reverse-biased semiconductor diodes whose depletion region width varies with applied voltage. They are used in phase-locked loops, amongst other applications.

Capacitor markings

Most capacitors have numbers printed on their bodies to indicate their electrical characteristics. Larger capacitors like electrolytics usually display the actual capacitance together with the unit (for example, **220 µF**). Smaller capacitors like ceramics, however, use a shorthand consisting of three numbers and a letter, where the numbers show the capacitance in pF (calculated as XY x 10^Z for the numbers XYZ) and the letter indicates the tolerance (J, K or M for ±5%, ±10% and ±20% respectively).

Additionally, the capacitor may show its working voltage, temperature and other relevant characteristics.

Example

A capacitor with the text **473K 330V** on its body has a capacitance of 47 x 10^3 pF = 47 nF (±10%) with a working voltage of 330 V.

Applications

Capacitors have many uses in electronic and electrical systems. They are so common that it is a rare electrical product that does not include at least one for some purpose.

Energy storage

A capacitor can store electric energy when disconnected from its charging circuit, so it can be used like a temporary battery. Capacitors are commonly used in electronic devices to maintain

power supply while batteries are being changed. (This prevents loss of information in volatile memory.)

Conventional capacitors provide less than 360 joules per kilogram of energy density, while capacitors using developing technologies could provide more than 2.52 kilojoules per kilogram.

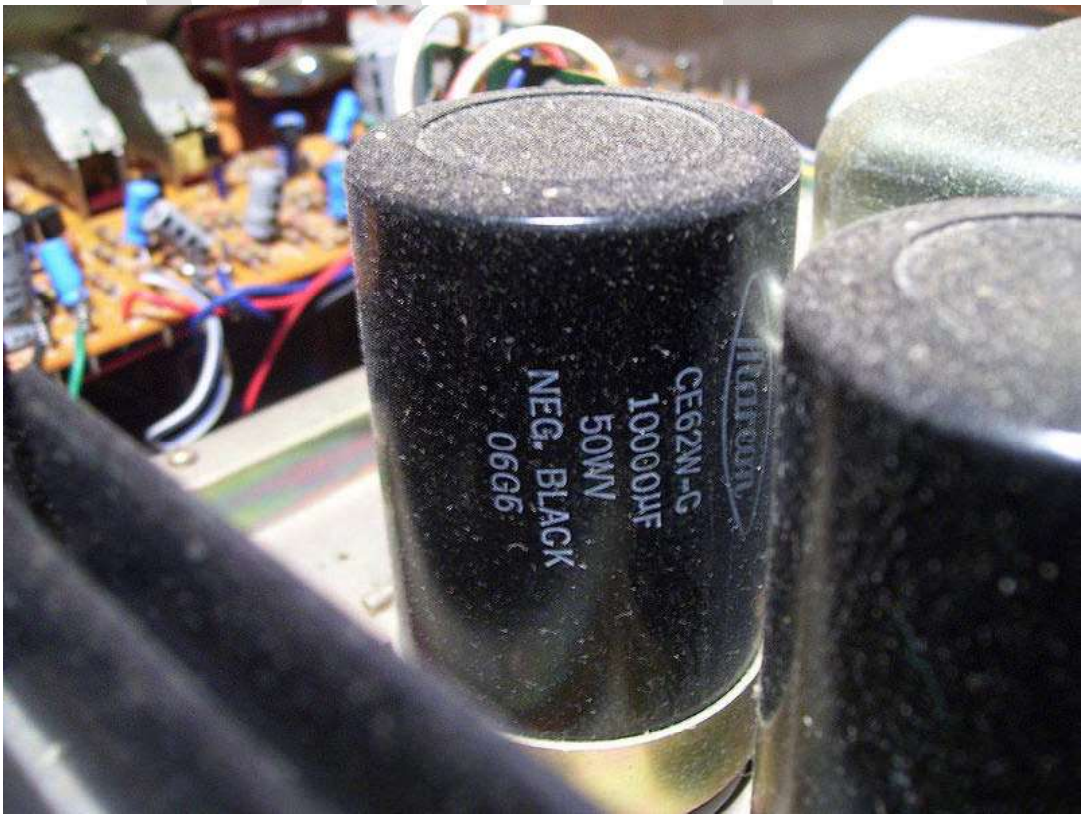
In car audio systems, large capacitors store energy for the amplifier to use on demand. Also for a flash tube a capacitor is used to hold the high voltage.

Pulsed power and weapons

Groups of large, specially constructed, low-inductance high-voltage capacitors (*capacitor banks*) are used to supply huge pulses of current for many pulsed power applications. These include electromagnetic forming, Marx generators, pulsed lasers (especially TEA lasers), pulse forming networks, radar, fusion research, and particle accelerators.

Large capacitor banks (reservoir) are used as energy sources for the exploding-bridgewire detonators or slapper detonators in nuclear weapons and other specialty weapons. Experimental work is under way using banks of capacitors as power sources for electromagnetic armour and electromagnetic railguns and coilguns.

Power conditioning



A 10,000 microfarad capacitor in a TRM-800 amplifier

Reservoir capacitors are used in power supplies where they smooth the output of a full or half wave rectifier. They can also be used in charge pump circuits as the energy storage element in the generation of higher voltages than the input voltage.

Capacitors are connected in parallel with the power circuits of most electronic devices and larger systems (such as factories) to shunt away and conceal current fluctuations from the primary power source to provide a "clean" power supply for signal or control circuits. Audio equipment, for example, uses several capacitors in this way, to shunt away power line hum before it gets into the signal circuitry. The capacitors act as a local reserve for the DC power source, and bypass AC currents from the power supply. This is used in car audio applications, when a stiffening capacitor compensates for the inductance and resistance of the leads to the lead-acid car battery.

Power factor correction

In electric power distribution, capacitors are used for power factor correction. Such capacitors often come as three capacitors connected as a three phase load. Usually, the values of these capacitors are given not in farads but rather as a reactive power in volt-amperes reactive (VAr). The purpose is to counteract inductive loading from devices like electric motors and transmission lines to make the load appear to be mostly resistive. Individual motor or lamp loads may have capacitors for power factor correction, or larger sets of capacitors (usually with automatic switching devices) may be installed at a load center within a building or in a large utility substation.

Supression and coupling

Signal coupling

Because capacitors pass AC but block DC signals (when charged up to the applied dc voltage), they are often used to separate the AC and DC components of a signal. This method is known as *AC coupling* or "capacitive coupling". Here, a large value of capacitance, whose value need not be accurately controlled, but whose reactance is small at the signal frequency, is employed.

Decoupling

A decoupling capacitor is a capacitor used to protect one part of a circuit from the effect of another, for instance to suppress noise or transients. Noise caused by other circuit elements is shunted through the capacitor, reducing the effect they have on the rest of the circuit. It is most commonly used between the power supply and ground. An alternative name is *bypass capacitor* as it is used to bypass the power supply or other high impedance component of a circuit.

Noise filters and snubbers

When an inductive circuit is opened, the current through the inductance collapses quickly, creating a large voltage across the open circuit of the switch or relay. If the inductance is large enough, the energy will generate a spark, causing the contact points to oxidize, deteriorate, or

sometimes weld together, or destroying a solid-state switch. A snubber capacitor across the newly opened circuit creates a path for this impulse to bypass the contact points, thereby preserving their life; these were commonly found in contact breaker ignition systems, for instance. Similarly, in smaller scale circuits, the spark may not be enough to damage the switch but will still radiate undesirable radio frequency interference (RFI), which a filter capacitor absorbs. Snubber capacitors are usually employed with a low-value resistor in series, to dissipate energy and minimize RFI. Such resistor-capacitor combinations are available in a single package.

Capacitors are also used in parallel to interrupt units of a high-voltage circuit breaker in order to equally distribute the voltage between these units. In this case they are called grading capacitors.

In schematic diagrams, a capacitor used primarily for DC charge storage is often drawn vertically in circuit diagrams with the lower, more negative, plate drawn as an arc. The straight plate indicates the positive terminal of the device, if it is polarized.

Motor starters

In single phase squirrel cage motors, the primary winding within the motor housing is not capable of starting a rotational motion on the rotor, but is capable of sustaining one. To start the motor, a secondary winding is used in series with a non-polarized *starting capacitor* to introduce a lag in the sinusoidal current through the starting winding. When the secondary winding is placed at an angle with respect to the primary winding, a rotating electric field is created. The force of the rotational field is not constant, but is sufficient to start the rotor spinning. When the rotor comes close to operating speed, a centrifugal switch (or current-sensitive relay in series with the main winding) disconnects the capacitor. The start capacitor is typically mounted to the side of the motor housing. These are called capacitor-start motors, that have relatively high starting torque.

There are also capacitor-run induction motors which have a permanently connected phase-shifting capacitor in series with a second winding. The motor is much like a two-phase induction motor.

Motor-starting capacitors are typically non-polarized electrolytic types, while running capacitors are conventional paper or plastic film dielectric types.

Signal processing

The energy stored in a capacitor can be used to represent information, either in binary form, as in DRAMs, or in analogue form, as in analog sampled filters and CCDs. Capacitors can be used in analog circuits as components of integrators or more complex filters and in negative feedback loop stabilization. Signal processing circuits also use capacitors to integrate a current signal.

Tuned circuits

Capacitors and inductors are applied together in tuned circuits to select information in particular frequency bands. For example, radio receivers rely on variable capacitors to tune the station frequency. Speakers use passive analog crossovers, and analog equalizers use capacitors to select different audio bands.

The resonant frequency *f* of a tuned circuit is a function of the inductance (*L*) and capacitance (*C*) in series, and is given by:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where *L* is in henries and *C* is in farads.

Sensing

Most capacitors are designed to maintain a fixed physical structure. However, various factors can change the structure of the capacitor, and the resulting change in capacitance can be used to sense those factors.

Changing the dielectric:

> The effects of varying the physical and/or electrical characteristics of the **dielectric** can be used for sensing purposes. Capacitors with an exposed and porous dielectric can be used to measure humidity in air. Capacitors are used to accurately measure the fuel level in airplanes; as the fuel covers more of a pair of plates, the circuit capacitance increases.

Changing the distance between the plates:

> Capacitors with a flexible plate can be used to measure strain or pressure. Industrial pressure transmitters used for process control use pressure-sensing diaphragms, which form a capacitor plate of an oscillator circuit. Capacitors are used as the sensor in condenser microphones, where one plate is moved by air pressure, relative to the fixed position of the other plate. Some accelerometers use MEMS capacitors etched on a chip to measure the magnitude and direction of the acceleration vector. They are used to detect changes in acceleration, e.g. as tilt sensors or to detect free fall, as sensors triggering airbag deployment, and in many other applications. Some fingerprint sensors use capacitors. Additionally, a user can adjust the pitch of a theremin musical instrument by moving his hand since this changes the effective capacitance between the user's hand and the antenna.

Changing the effective area of the plates:

> Capacitive touch switches are now used on many consumer electronic products.

Hazards and safety

Capacitors may retain a charge long after power is removed from a circuit; this charge can cause dangerous or even potentially fatal shocks or damage connected equipment. For example, even a seemingly innocuous device such as a disposable camera flash unit powered by a 1.5 volt AA battery contains a capacitor which may be charged to over 300 volts. This is easily capable of delivering a shock. Service procedures for electronic devices usually include instructions to discharge large or high-voltage capacitors. Capacitors may also have built-in discharge resistors to dissipate stored energy to a safe level within a few seconds after power is removed. High-voltage capacitors are stored with the terminals shorted, as protection from potentially dangerous voltages due to dielectric absorption.

Some old, large oil-filled capacitors contain polychlorinated biphenyls (PCBs). It is known that waste PCBs can leak into groundwater under landfills. Capacitors containing PCB were labelled as containing "Askarel" and several other trade names. PCB-filled capacitors are found in very old (pre 1975) fluorescent lamp ballasts, and other applications.

High-voltage capacitors may catastrophically fail when subjected to voltages or currents beyond their rating, or as they reach their normal end of life. Dielectric or metal interconnection failures may create arcing that vaporizes dielectric fluid, resulting in case bulging, rupture, or even an explosion. Capacitors used in RF or sustained high-current applications can overheat, especially in the center of the capacitor rolls. Capacitors used within high-energy capacitor banks can violently explode when a short in one capacitor causes sudden dumping of energy stored in the rest of the bank into the failing unit. High voltage vacuum capacitors can generate soft X-rays even during normal operation. Proper containment, fusing, and preventive maintenance can help to minimize these hazards.

Chapter 10

Transformer

Pole-mounted power distribution transformer with center-tapped secondary winding (note use of grounded conductor, right, as one leg of the primary feeder). It transforms the high voltage of the overhead distribution wires to the lower voltage used in house wiring.

A **transformer** is a static device that transfers electrical energy from one circuit to another through inductively coupled conductors—the transformer's coils. A varying current in the first or *primary* winding creates a varying magnetic flux in the transformer's core and thus a varying

magnetic field through the *secondary* winding. This varying magnetic field induces a varying electromotive force (EMF) or "voltage" in the secondary winding. This effect is called mutual induction.

If a load is connected to the secondary, an electric current will flow in the secondary winding and electrical energy will be transferred from the primary circuit through the transformer to the load. In an ideal transformer, the induced voltage in the secondary winding (V_s) is in proportion to the primary voltage (V_p), and is given by the ratio of the number of turns in the secondary (N_s) to the number of turns in the primary (N_p) as follows:

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}$$

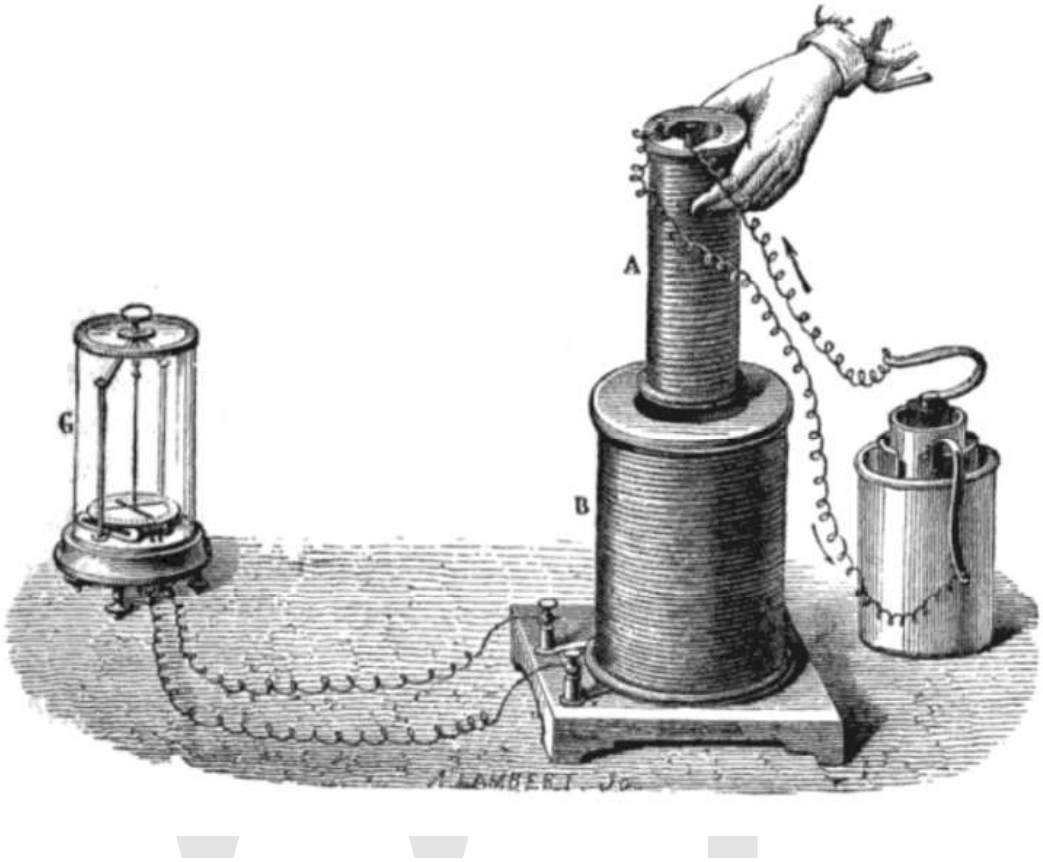
By appropriate selection of the ratio of turns, a transformer thus allows an alternating current (AC) voltage to be "stepped up" by making N_s greater than N_p, or "stepped down" by making N_s less than N_p.

In the vast majority of transformers, the windings are coils wound around a ferromagnetic core, air-core transformers being a notable exception.

Transformers range in size from a thumbnail-sized coupling transformer hidden inside a stage microphone to huge units weighing hundreds of tons used to interconnect portions of power grids. All operate with the same basic principles, although the range of designs is wide. While new technologies have eliminated the need for transformers in some electronic circuits, transformers are still found in nearly all electronic devices designed for household ("mains") voltage. Transformers are essential for high-voltage electric power transmission, which makes long-distance transmission economically practical.

History

Discovery



Faraday's experiment with induction between coils of wire

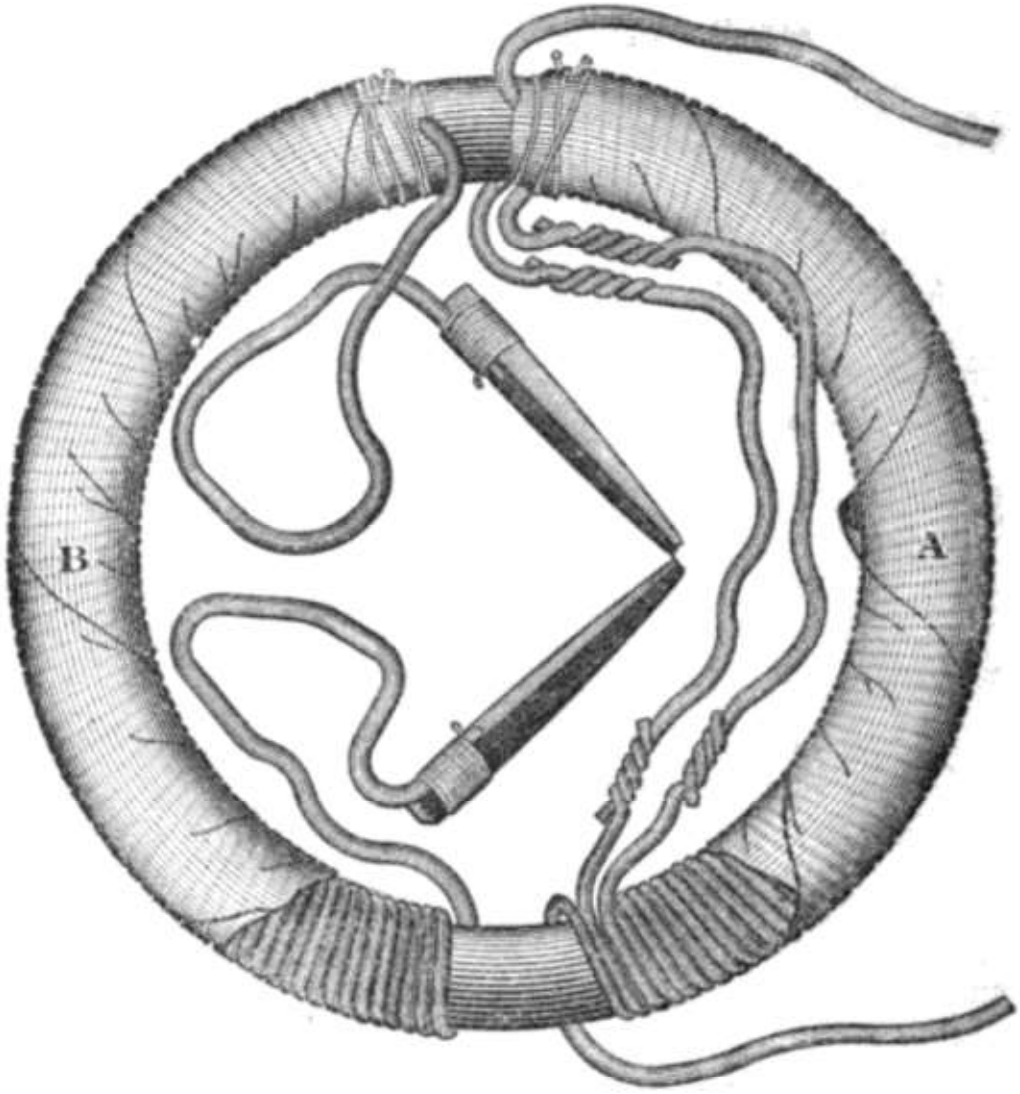
The phenomenon of electromagnetic induction was discovered independently by Michael Faraday and Joseph Henry in 1831. However, Faraday was the first to publish the results of his experiments and thus receive credit for the discovery. The relationship between electromotive force (EMF) or "voltage" and magnetic flux was formalized in an equation now referred to as "Faraday's law of induction":

$$|\mathcal{E}| = \left|\frac{d\Phi_B}{dt}\right|$$

where $|\mathcal{E}|$ is the magnitude of the EMF in volts and Φ_B is the magnetic flux through the circuit (in webers).

Faraday performed the first experiments on induction between coils of wire, including winding a pair of coils around an iron ring, thus creating the first toroidal closed-core transformer.

Induction coils



Faraday's ring transformer

The first type of transformer to see wide use was the induction coil, invented by Rev. Nicholas Callan of Maynooth College, Ireland in 1836. He was one of the first researchers to realize that the more turns the secondary winding has in relation to the primary winding, the larger is the increase in EMF. Induction coils evolved from scientists' and inventors' efforts to get higher voltages from batteries. Since batteries produce direct current (DC) rather than alternating current (AC), induction coils relied upon vibrating electrical contacts that regularly interrupted the current in the primary to create the flux changes necessary for induction. Between the 1830s and the 1870s, efforts to build better induction coils, mostly by trial and error, slowly revealed the basic principles of transformers.

In 1876, Russian engineer Pavel Yablochkov invented a lighting system based on a set of induction coils where the primary windings were connected to a source of alternating current and the secondary windings could be connected to several "electric candles" (arc lamps) of his own design. The coils Yablochkov employed functioned essentially as transformers.

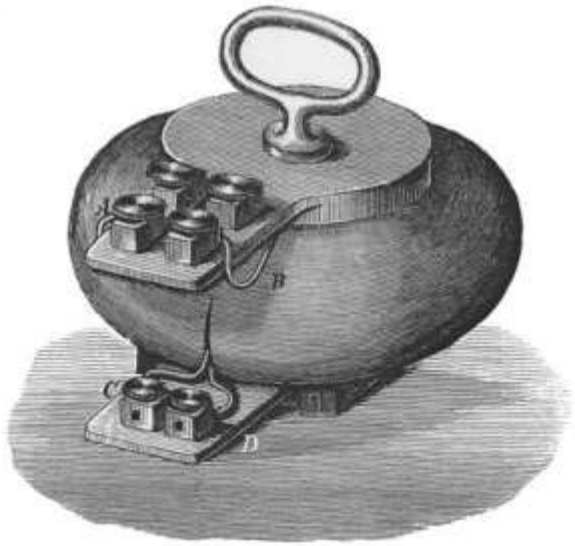
In 1878, the Ganz Company in Hungary began manufacturing equipment for electric lighting and, by 1883, had installed over fifty systems in Austria-Hungary. Their systems used alternating current exclusively and included those comprising both arc and incandescent lamps, along with generators and other equipment.

Lucien Gaulard and John Dixon Gibbs first exhibited a device with an open iron core called a "secondary generator" in London in 1882, then sold the idea to the Westinghouse company in the United States. They also exhibited the invention in Turin, Italy in 1884, where it was adopted for an electric lighting system. However, the efficiency of their open-core bipolar apparatus remained very low.

Induction coils with open magnetic circuits are inefficient for transfer of power to loads. Until about 1880, the paradigm for AC power transmission from a high voltage supply to a low voltage load was a series circuit. Open-core transformers with a ratio near 1:1 were connected with their primaries in series to allow use of a high voltage for transmission while presenting a low voltage to the lamps. The inherent flaw in this method was that turning off a single lamp affected the voltage supplied to all others on the same circuit. Many adjustable transformer designs were introduced to compensate for this problematic characteristic of the series circuit, including those employing methods of adjusting the core or bypassing the magnetic flux around part of a coil.

Efficient, practical transformer designs did not appear until the 1880s, but within a decade the transformer would be instrumental in the "War of Currents", and in seeing AC distribution systems triumph over their DC counterparts, a position in which they have remained dominant ever since.

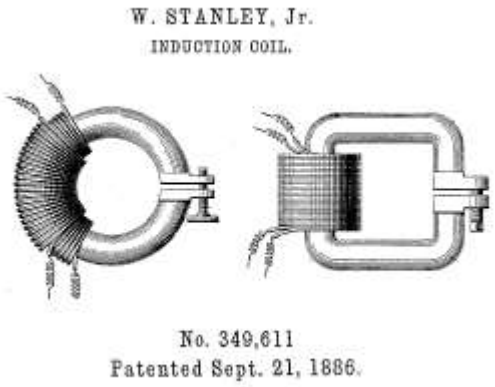
Closed-core lighting transformers



Drawing of Ganz Company's 1885 prototype. Capacity: 1400 VA, frequency: 40 Hz, voltage ratio: 120/72 V

Prototypes of the world's first high-efficiency transformers. They were built by the Z.B.D. team on 16th September 1884.

In the autumn of 1884, Ganz Company engineers Károly Zipernowsky, Ottó Bláthy and Miksa Déri had determined that open-core devices were impracticable, as they were incapable of reliably regulating voltage. In their joint patent application for the "Z.B.D." transformers, they described two designs with closed magnetic circuits: the "closed-core" and "shell-core" transformers. In the closed-core, the primary and secondary windings were wound around a closed iron ring; in the shell-core, the windings were passed *through* the iron core. In both designs, the magnetic flux linking the primary and secondary windings traveled almost entirely within the iron core, with no intentional path through air. The new Z.B.D. transformers reached 98 percent efficiency, which was 3.4 times higher than the open core bipolar devices of Gaulard and Gibs. When they employed it in parallel connected electric distribution systems, closed-core transformers finally made it technically and economically feasible to provide electric power for lighting in homes, businesses and public spaces. Bláthy had suggested the use of closed-cores, Zipernowsky the use of shunt connections, and Déri had performed the experiments; Bláthy also discovered the transformer formula, Vs/Vp = Ns/Np. The vast majority of transformers in use today rely on the basic principles discovered by the three engineers. They also reportedly popularized the word "transformer" to describe a device for altering the EMF of an electric current, although the term had already been in use by 1882. In 1886, the Ganz Company installed the world's first power station that used AC generators to power a parallel-connected common electrical network, the steam-powered Rome-Cerchi power plant.


W. STANLEY, Jr.
INDUCTION COIL.

No. 349,611
Patented Sept. 21, 1886.

Stanley's 1886 design for adjustable gap open-core induction coils

Although George Westinghouse had bought Gaulard and Gibbs' patents in 1885, the Edison Electric Light Company held an option on the U.S. rights for the Z.B.D. transformers, requiring Westinghouse to pursue alternative designs on the same principles. He assigned to William Stanley the task of developing a device for commercial use in United States. Stanley's first patented design was for induction coils with single cores of soft iron and adjustable gaps to regulate the EMF present in the secondary winding. This design was first used commercially in the U.S. in 1886. But Westinghouse soon had his team working on a design whose core comprised a stack of thin "E-shaped" iron plates, separated individually or in pairs by thin sheets of paper or other insulating material. Prewound copper coils could then be slid into place, and straight iron plates laid in to create a closed magnetic circuit. Westinghouse applied for a patent for the new design in December 1886; it was granted in July 1887.

Other early transformers

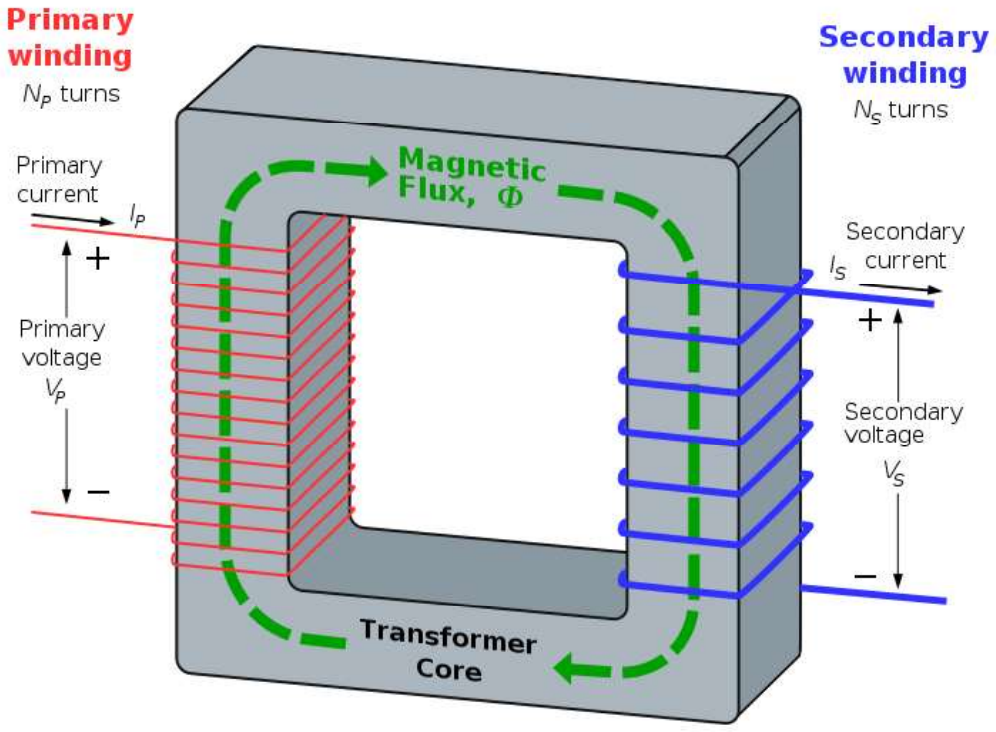
In 1889, Russian-born engineer Mikhail Dolivo-Dobrovolsky developed the first three-phase transformer at the Allgemeine Elektricitäts-Gesellschaft ("General Electricity Company") in Germany.

In 1891, Nikola Tesla invented the Tesla coil, an air-cored, dual-tuned resonant transformer for generating very high voltages at high frequency.

Audio frequency transformers ("repeating coils") were used by early experimenters in the development of the telephone.

Basic principles

The transformer is based on two principles: first, that an electric current can produce a magnetic field (electromagnetism), and, second that a changing magnetic field within a coil of wire induces a voltage across the ends of the coil (electromagnetic induction). Changing the current in the primary coil changes the magnetic flux that is developed. The changing magnetic flux induces a voltage in the secondary coil.



An ideal transformer

An ideal transformer is shown in the adjacent figure. Current passing through the primary coil creates a magnetic field. The primary and secondary coils are wrapped around a core of very

high magnetic permeability, such as iron, so that most of the magnetic flux passes through both the primary and secondary coils.

Induction law

The voltage induced across the secondary coil may be calculated from Faraday's law of induction, which states that:

$$V_\mathrm{s} = N_\mathrm{s}\frac{\mathrm{d}\Phi}{\mathrm{d}t},$$

where V_s is the instantaneous voltage, N_s is the number of turns in the secondary coil and Φ is the magnetic flux through one turn of the coil. If the turns of the coil are oriented perpendicular to the magnetic field lines, the flux is the product of the magnetic flux density B and the area A through which it cuts. The area is constant, being equal to the cross-sectional area of the transformer core, whereas the magnetic field varies with time according to the excitation of the primary. Since the same magnetic flux passes through both the primary and secondary coils in an ideal transformer, the instantaneous voltage across the primary winding equals

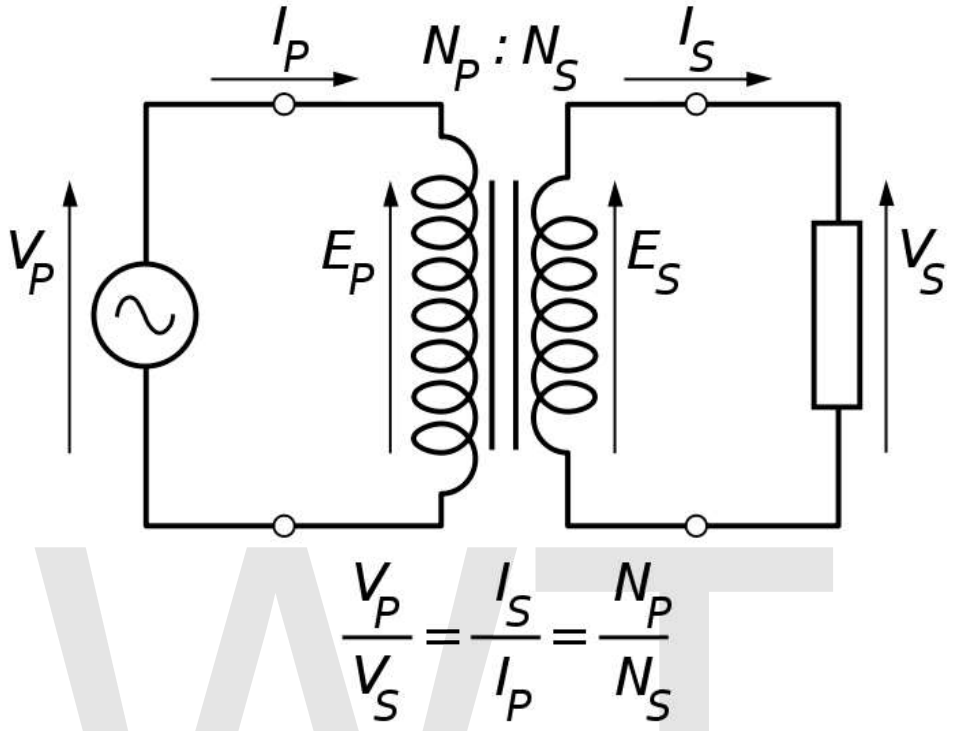
$$V_\mathrm{p} = N_\mathrm{p}\frac{\mathrm{d}\Phi}{\mathrm{d}t}.$$

Taking the ratio of the two equations for V_s and V_p gives the basic equation for stepping up or stepping down the voltage

$$\frac{V_\mathrm{s}}{V_\mathrm{p}} = \frac{N_\mathrm{s}}{N_\mathrm{p}}.$$

$N_\mathrm{p}/N_\mathrm{s}$ is known as the *turns ratio*, and is the primary functional characteristic of any transformer. In the case of step-up transformers, this may sometimes be stated as the reciprocal, $N_\mathrm{s}/N_\mathrm{p}$. *Turns ratio* is commonly expressed as an irreducible fraction or ratio: for example, a transformer with primary and secondary windings of, respectively, 100 and 150 turns is said to have a turns ratio of 2:3 rather than 0.667 or 100:150.

Ideal power equation



The ideal transformer as a circuit element

If the secondary coil is attached to a load that allows current to flow, electrical power is transmitted from the primary circuit to the secondary circuit. Ideally, the transformer is perfectly efficient; all the incoming energy is transformed from the primary circuit to the magnetic field and into the secondary circuit. If this condition is met, the incoming electric power must equal the outgoing power:

$$P_{\text{incoming}} = I_\text{p} V_\text{p} = P_{\text{outgoing}} = I_\text{s} V_\text{s},$$

giving the ideal transformer equation

$$\frac{V_\text{s}}{V_\text{p}} = \frac{N_\text{s}}{N_\text{p}} = \frac{I_\text{p}}{I_\text{s}}.$$

Transformers normally have high efficiency, so this formula is a reasonable approximation.

If the voltage is increased, then the current is decreased by the same factor. The impedance in one circuit is transformed by the *square* of the turns ratio. For example, if an impedance Z_s is attached across the terminals of the secondary coil, it appears to the primary circuit to have an

impedance of $(N_p/N_s)^2 Z_s$. This relationship is reciprocal, so that the impedance Z_p of the primary circuit appears to the secondary to be $(N_s/N_p)^2 Z_p$.

Detailed operation

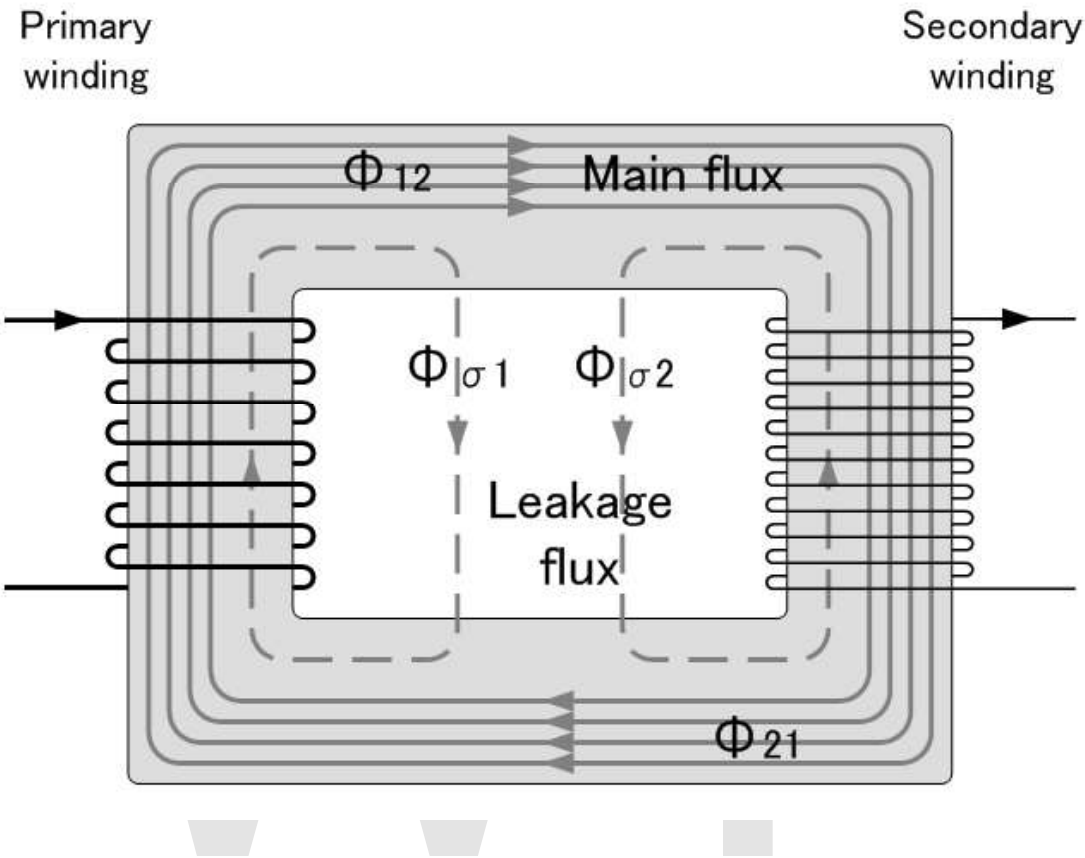
The simplified description above neglects several practical factors, in particular the primary current required to establish a magnetic field in the core, and the contribution to the field due to current in the secondary circuit.

Models of an ideal transformer typically assume a core of negligible reluctance with two windings of zero resistance. When a voltage is applied to the primary winding, a small current flows, driving flux around the magnetic circuit of the core. The current required to create the flux is termed the *magnetizing current*; since the ideal core has been assumed to have near-zero reluctance, the magnetizing current is negligible, although still required to create the magnetic field.

The changing magnetic field induces an electromotive force (EMF) across each winding. Since the ideal windings have no impedance, they have no associated voltage drop, and so the voltages V_P and V_S measured at the terminals of the transformer, are equal to the corresponding EMFs. The primary EMF, acting as it does in opposition to the primary voltage, is sometimes termed the "back EMF". This is due to Lenz's law which states that the induction of EMF would always be such that it will oppose development of any such change in magnetic field.

Practical considerations

Leakage flux



Leakage flux of a transformer

The ideal transformer model assumes that all flux generated by the primary winding links all the turns of every winding, including itself. In practice, some flux traverses paths that take it outside the windings. Such flux is termed *leakage flux*, and results in leakage inductance in series with the mutually coupled transformer windings. Leakage results in energy being alternately stored in and discharged from the magnetic fields with each cycle of the power supply. It is not directly a power loss, but results in inferior voltage regulation, causing the secondary voltage to fail to be directly proportional to the primary, particularly under heavy load. Transformers are therefore normally designed to have very low leakage inductance.

However, in some applications, leakage can be a desirable property, and long magnetic paths, air gaps, or magnetic bypass shunts may be deliberately introduced to a transformer's design to limit the short-circuit current it will supply. Leaky transformers may be used to supply loads that exhibit negative resistance, such as electric arcs, mercury vapor lamps, and neon signs; or for safely handling loads that become periodically short-circuited such as electric arc welders.

Air gaps are also used to keep a transformer from saturating, especially audio-frequency transformers in circuits that have a direct current flowing through the windings.

Leakage inductance is also helpful when transformers are operated in parallel. It can be shown that if the "per-unit" inductance of two transformers is the same (a typical value is 5%), they will automatically split power "correctly" (e.g. 500 kVA unit in parallel with 1,000 kVA unit, the larger one will carry twice the current).

Effect of frequency

Transformer universal EMF equation

If the flux in the core is purely sinusoidal, the relationship for either winding between its **rms voltage** E_{rms} of the winding , and the supply frequency f, number of turns N, core cross-sectional area a and peak magnetic flux density B is given by the universal EMF equation:

$$E_{\mathrm{rms}} = \frac{2\pi f N a B_{\mathrm{peak}}}{\sqrt{2}} \approx 4.44 f N a B$$

If the flux does not contain even harmonics the following equation can be used for **half-cycle average voltage** E_{avg} of any waveshape:

$$E_{\mathrm{avg}} = 4 f N a B_{\mathrm{peak}}$$

The time-derivative term in Faraday's Law shows that the flux in the core is the integral with respect to time of the applied voltage. Hypothetically an ideal transformer would work with direct-current excitation, with the core flux increasing linearly with time. In practice, the flux would rise to the point where magnetic saturation of the core occurs, causing a huge increase in the magnetizing current and overheating the transformer. All practical transformers must therefore operate with alternating (or pulsed) current.

The EMF of a transformer at a given flux density increases with frequency. By operating at higher frequencies, transformers can be physically more compact because a given core is able to transfer more power without reaching saturation and fewer turns are needed to achieve the same impedance. However, properties such as core loss and conductor skin effect also increase with frequency. Aircraft and military equipment employ 400 Hz power supplies which reduce core and winding weight. Conversely, frequencies used for some railway electrification systems were much lower (e.g. 16.7 Hz and 25 Hz) than normal utility frequencies (50 – 60 Hz) for historical reasons concerned mainly with the limitations of early electric traction motors. As such, the transformers used to step down the high over-head line voltages (e.g. 15 kV) are much heavier for the same power rating than those designed only for the higher frequencies.

Operation of a transformer at its designed voltage but at a higher frequency than intended will lead to reduced magnetizing current; at lower frequency, the magnetizing current will increase. Operation of a transformer at other than its design frequency may require assessment of voltages, losses, and cooling to establish if safe operation is practical. For example, transformers may need to be equipped with "volts per hertz" over-excitation relays to protect the transformer from overvoltage at higher than rated frequency.

One example of state-of-the-art design is those transformers used for electric multiple unit high speed trains, particularly those required to operate across the borders of countries using different standards of electrification. The position of such transformers is restricted to being hung below the passenger compartment. They have to function at different frequencies (down to 16.7 Hz) and voltages (up to 25 kV) whilst handling the enhanced power requirements needed for operating the trains at high speed.

Knowledge of natural frequencies of transformer windings is of importance for the determination of the transient response of the windings to impulse and switching surge voltages.

Energy losses

An ideal transformer would have no energy losses, and would be 100% efficient. In practical transformers energy is dissipated in the windings, core, and surrounding structures. Larger transformers are generally more efficient, and those rated for electricity distribution usually perform better than 98%.

Experimental transformers using superconducting windings achieve efficiencies of 99.85%. The increase in efficiency can save considerable energy, and hence money, in a large heavily-loaded transformer; the trade-off is in the additional initial and running cost of the superconducting design.

Losses in transformers (excluding associated circuitry) vary with load current, and may be expressed as "no-load" or "full-load" loss. Winding resistance dominates load losses, whereas hysteresis and eddy currents losses contribute to over 99% of the no-load loss. The no-load loss can be significant, so that even an idle transformer constitutes a drain on the electrical supply and a running cost; designing transformers for lower loss requires a larger core, good-quality silicon steel, or even amorphous steel, for the core, and thicker wire, increasing initial cost, so that there is a trade-off between initial cost and running cost.

Transformer losses are divided into losses in the windings, termed copper loss, and those in the magnetic circuit, termed iron loss. Losses in the transformer arise from:

Winding resistance
: Current flowing through the windings causes resistive heating of the conductors. At higher frequencies, skin effect and proximity effect create additional winding resistance and losses.

Hysteresis losses
: Each time the magnetic field is reversed, a small amount of energy is lost due to hysteresis within the core. For a given core material, the loss is proportional to the frequency, and is a function of the peak flux density to which it is subjected.

Eddy currents
: Ferromagnetic materials are also good conductors, and a core made from such a material also constitutes a single short-circuited turn throughout its entire length. Eddy currents therefore circulate within the core in a plane normal to the flux, and are responsible for resistive heating of the core material. The eddy current loss is a complex function of the

square of supply frequency and inverse square of the material thickness. Eddy current losses can be reduced by making the core of a stack of plates electrically insulated from each other, rather than a solid block; all transformers operating at low frequencies use laminated or similar cores.

Magnetostriction

Magnetic flux in a ferromagnetic material, such as the core, causes it to physically expand and contract slightly with each cycle of the magnetic field, an effect known as magnetostriction. This produces the buzzing sound commonly associated with transformers, and can cause losses due to frictional heating.

Mechanical losses

In addition to magnetostriction, the alternating magnetic field causes fluctuating forces between the primary and secondary windings. These incite vibrations within nearby metalwork, adding to the buzzing noise, and consuming a small amount of power.

Stray losses

Leakage inductance is by itself largely lossless, since energy supplied to its magnetic fields is returned to the supply with the next half-cycle. However, any leakage flux that intercepts nearby conductive materials such as the transformer's support structure will give rise to eddy currents and be converted to heat. There are also radiative losses due to the oscillating magnetic field, but these are usually small.

Dot convention

It is common in transformer schematic symbols for there to be a dot at the end of each coil within a transformer, particularly for transformers with multiple primary and secondary windings. The dots indicate the direction of each winding relative to the others. Voltages at the dot end of each winding are in phase; current flowing into the dot end of a primary coil will result in current flowing out of the dot end of a secondary coil.

Equivalent circuit

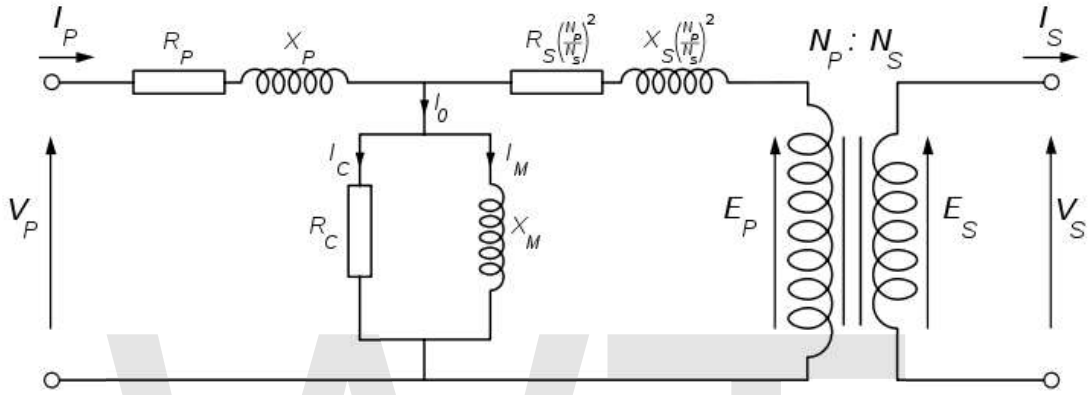
The physical limitations of the practical transformer may be brought together as an equivalent circuit model (shown below) built around an ideal lossless transformer. Power loss in the windings is current-dependent and is represented as in-series resistances R_p and R_s. Flux leakage results in a fraction of the applied voltage dropped without contributing to the mutual coupling, and thus can be modeled as reactances of each leakage inductance X_p and X_s in series with the perfectly coupled region.

Iron losses are caused mostly by hysteresis and eddy current effects in the core, and are proportional to the square of the core flux for operation at a given frequency. Since the core flux is proportional to the applied voltage, the iron loss can be represented by a resistance R_C in parallel with the ideal transformer.

A core with finite permeability requires a magnetizing current I_m to maintain the mutual flux in the core. The magnetizing current is in phase with the flux; saturation effects cause the relationship between the two to be non-linear, but for simplicity this effect tends to be ignored in most circuit equivalents. With a sinusoidal supply, the core flux lags the induced EMF by 90°

and this effect can be modeled as a magnetizing reactance (reactance of an effective inductance) X_m in parallel with the core loss component. R_c and X_m are sometimes together termed the *magnetizing branch* of the model. If the secondary winding is made open-circuit, the current I_0 taken by the magnetizing branch represents the transformer's no-load current.

The secondary impedance R_s and X_s is frequently moved (or "referred") to the primary side after multiplying the components by the impedance scaling factor $(N_p/N_s)^2$.



Transformer equivalent circuit, with secondary impedances referred to the primary side

The resulting model is sometimes termed the "exact equivalent circuit", though it retains a number of approximations, such as an assumption of linearity. Analysis may be simplified by moving the magnetizing branch to the left of the primary impedance, an implicit assumption that the magnetizing current is low, and then summing primary and referred secondary impedances, resulting in so-called equivalent impedance.

The parameters of equivalent circuit of a transformer can be calculated from the results of two transformer tests: open-circuit test and short-circuit test.

Types

A wide variety of transformer designs are used for different applications, though they share several common features. Important common transformer types include:

Autotransformer

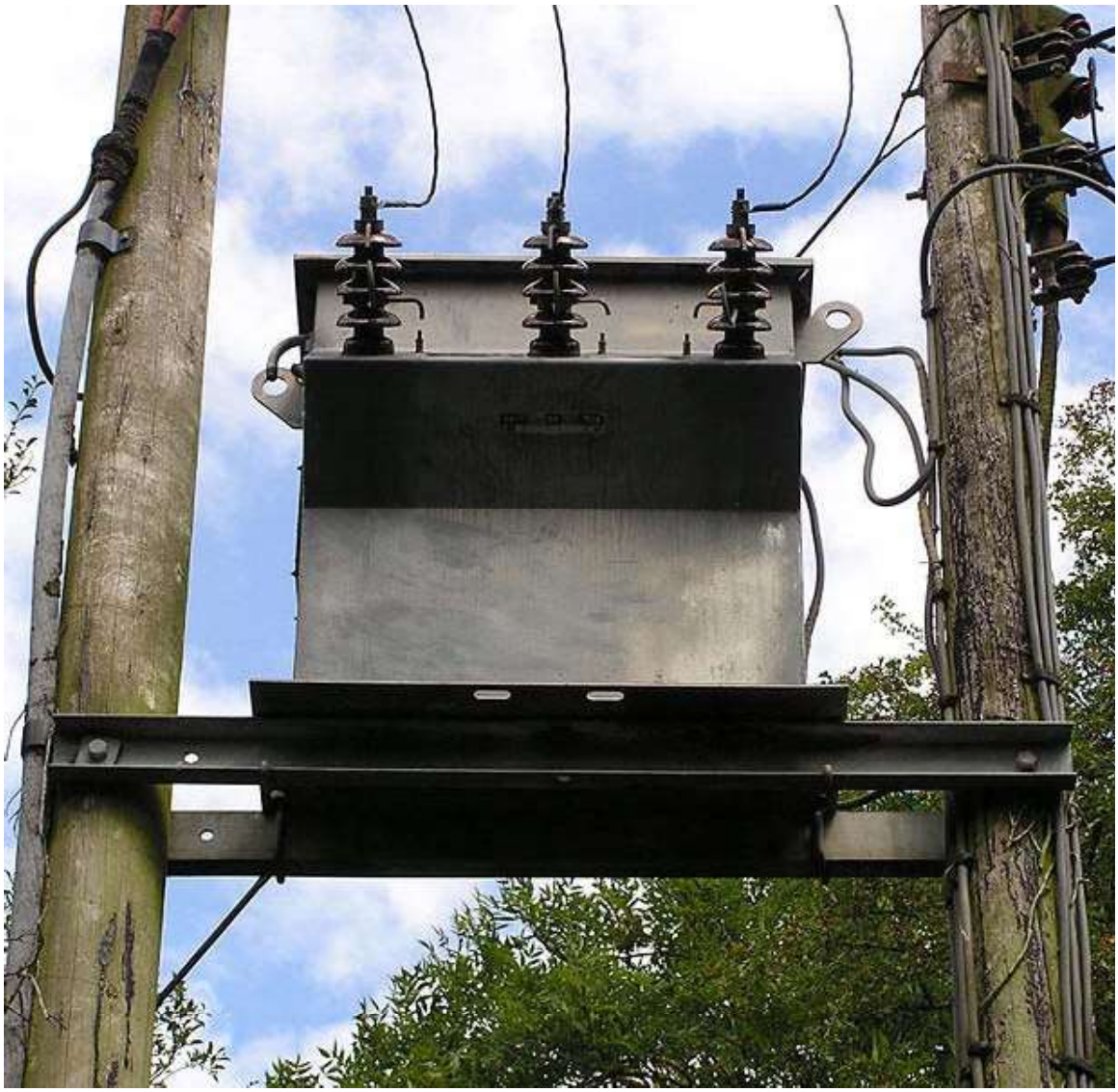
A variable autotransformer

In an autotransformer portions of the same winding act as both the primary and secondary. The winding has at least three taps where electrical connections are made. An autotransformer can be smaller, lighter and cheaper than a standard dual-winding transformer however the autotransformer does not provide electrical isolation.

Autotransformers are often used to step up or down between voltages in the 110-117-120 volt range and voltages in the 220-230-240 volt range, e.g., to output either 110 or 120V (with taps) from 230V input, allowing equipment from a 100 or 120V region to be used in a 230V region.

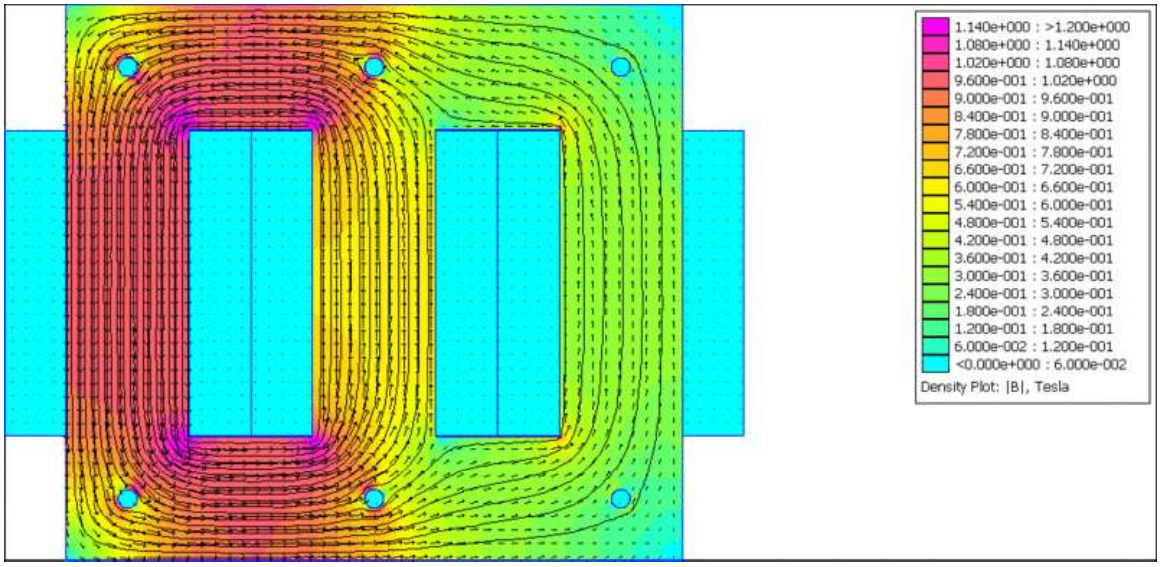
A variable autotransformer is made by exposing part of the winding coils and making the secondary connection through a sliding brush, giving a variable turns ratio. Such a device is often referred to by the trademark name *variac*.

Polyphase transformers



Three-phase step-down transformer mounted between two utility poles

For three-phase supplies, a bank of three individual single-phase transformers can be used, or all three phases can be incorporated as a single three-phase transformer. In this case, the magnetic circuits are connected together, the core thus containing a three-phase flow of flux. A number of winding configurations are possible, giving rise to different attributes and phase shifts. One particular polyphase configuration is the zigzag transformer, used for grounding and in the suppression of harmonic currents.



Screenshot of a FEM simulation of the magnetic flux inside a three-phase power transformer.

Leakage transformers

Leakage transformer

A leakage transformer, also called a stray-field transformer, has a significantly higher leakage inductance than other transformers, sometimes increased by a magnetic bypass or shunt in its core between primary and secondary, which is sometimes adjustable with a set screw. This provides a transformer with an inherent current limitation due to the loose coupling between its primary and the secondary windings. The output and input currents are low enough to prevent thermal overload under all load conditions—even if the secondary is shorted.

Leakage transformers are used for arc welding and high voltage discharge lamps (neon lights and cold cathode fluorescent lamps, which are series-connected up to 7.5 kV AC). It acts then both as a voltage transformer and as a magnetic ballast.

Other applications are short-circuit-proof extra-low voltage transformers for toys or doorbell installations.

Resonant transformers

A resonant transformer is a kind of leakage transformer. It uses the leakage inductance of its secondary windings in combination with external capacitors, to create one or more resonant circuits. Resonant transformers such as the Tesla coil can generate very high voltages, and are able to provide much higher current than electrostatic high-voltage generation machines such as the Van de Graaff generator. One of the applications of the resonant transformer is for the CCFL inverter. Another application of the resonant transformer is to couple between stages of a superheterodyne receiver, where the selectivity of the receiver is provided by tuned transformers in the intermediate-frequency amplifiers.

Audio transformers

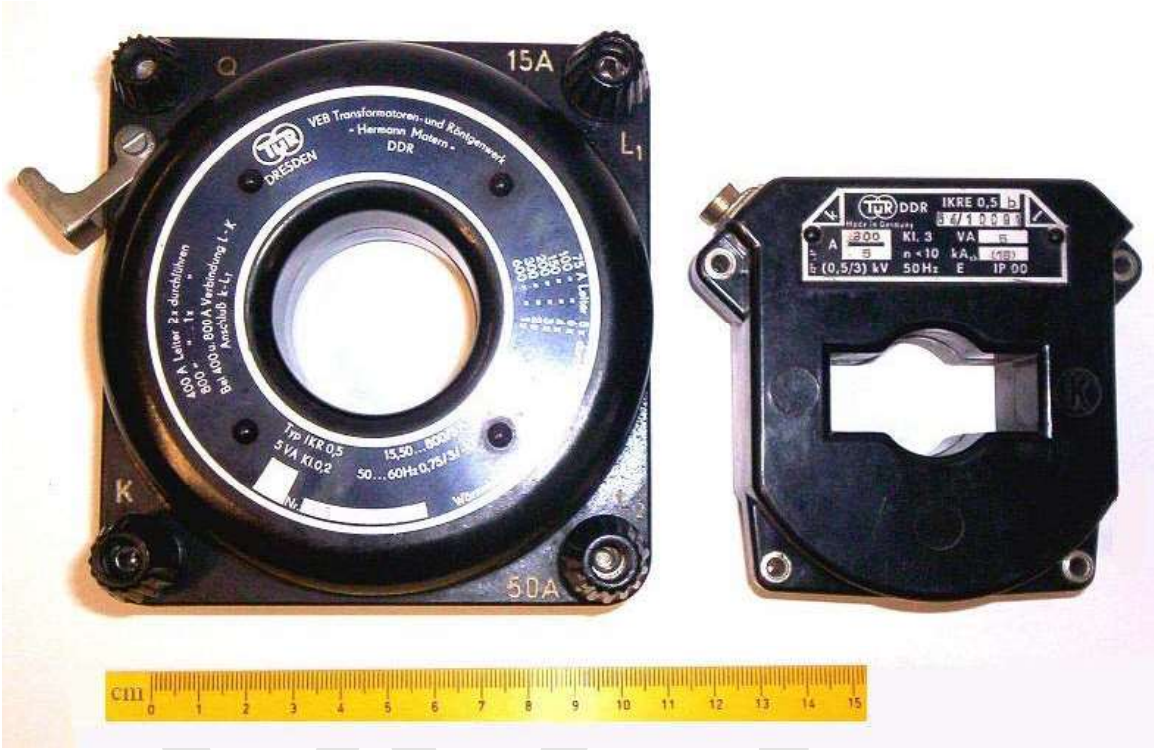
Audio transformers are those specifically designed for use in audio circuits. They can be used to block radio frequency interference or the DC component of an audio signal, to split or combine audio signals, or to provide impedance matching between high and low impedance circuits, such as between a high impedance tube (valve) amplifier output and a low impedance loudspeaker, or between a high impedance instrument output and the low impedance input of a mixing console.

Such transformers were originally designed to connect different telephone systems to one another while keeping their respective power supplies isolated, and are still commonly used to interconnect professional audio systems or system components.

Being magnetic devices, audio transformers are susceptible to external magnetic fields such as those generated by AC current-carrying conductors. "Hum" is a term commonly used to describe unwanted signals originating from the "mains" power supply (typically 50 or 60 Hz). Audio transformers used for low-level signals, such as those from microphones, often include shielding to protect against extraneous magnetically coupled signals.

Instrument transformers

Instrument transformers are used for measuring voltage and current in electrical power systems, and for power system protection and control. Where a voltage or current is too large to be conveniently used by an instrument, it can be scaled down to a standardized, low value. Instrument transformers isolate measurement, protection and control circuitry from the high currents or voltages present on the circuits being measured or controlled.



Current transformers, designed for placing around conductors

A current transformer is a transformer designed to provide a current in its secondary coil proportional to the current flowing in its primary coil.

Voltage transformers (VTs), also referred to as "potential transformers" (PTs), are designed to have an accurately known transformation ratio in both magnitude and phase, over a range of measuring circuit impedances. A voltage transformer is intended to present a negligible load to the supply being measured. The low secondary voltage allows protective relay equipment and measuring instruments to be operated at a lower voltages.

Both current and voltage instrument transformers are designed to have predictable characteristics on overloads. Proper operation of over-current protective relays requires that current transformers provide a predictable transformation ratio even during a short-circuit.

Classification

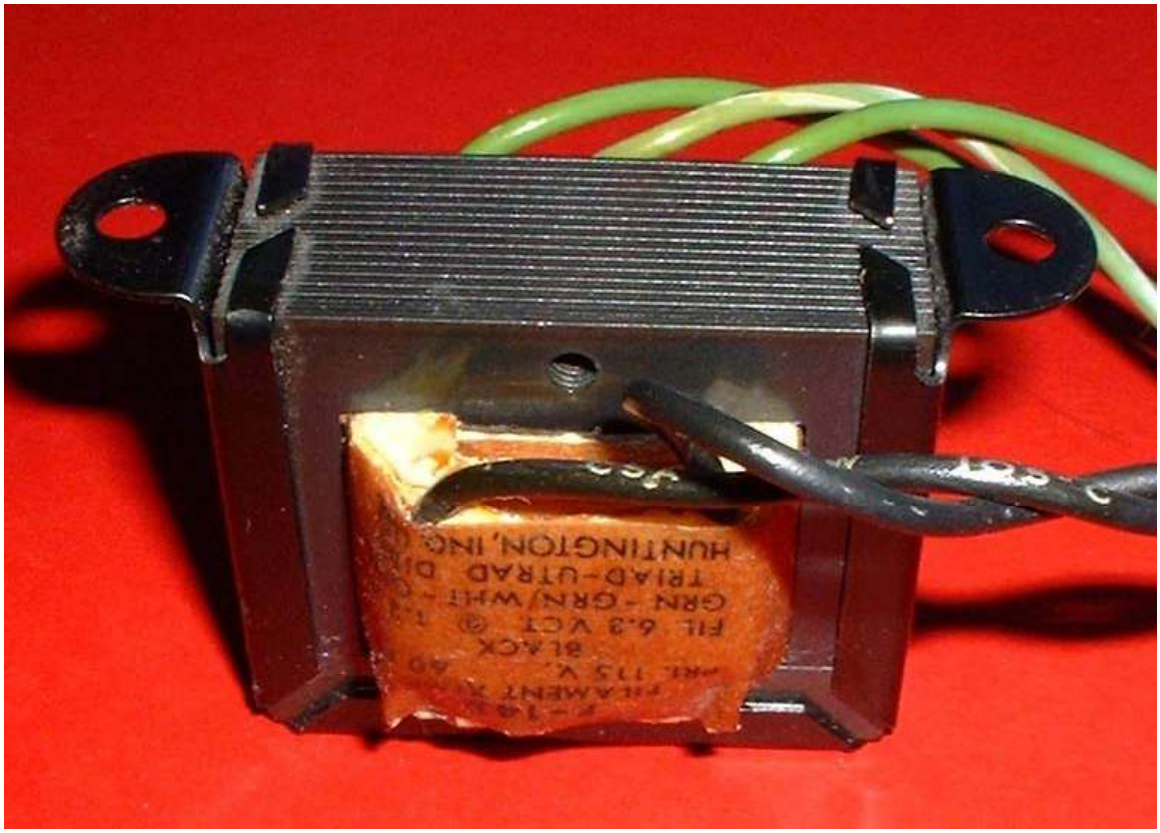
Transformers can be classified in many different ways; an incomplete list is:

- *By power capacity*: from a fraction of a volt-ampere (VA) to over a thousand MVA;
- *By frequency range*: power-, audio-, or radio frequency;
- *By voltage class*: from a few volts to hundreds of kilovolts;
- *By cooling type*: air-cooled, oil-filled, fan-cooled, or water-cooled;
- *By application*: such as power supply, impedance matching, output voltage and current stabilizer, or circuit isolation;

- *By purpose*: distribution, rectifier, arc furnace, amplifier output, etc.;
- *By winding turns ratio*: step-up, step-down, isolating with equal or near-equal ratio, variable, multiple windings.

Construction

Cores



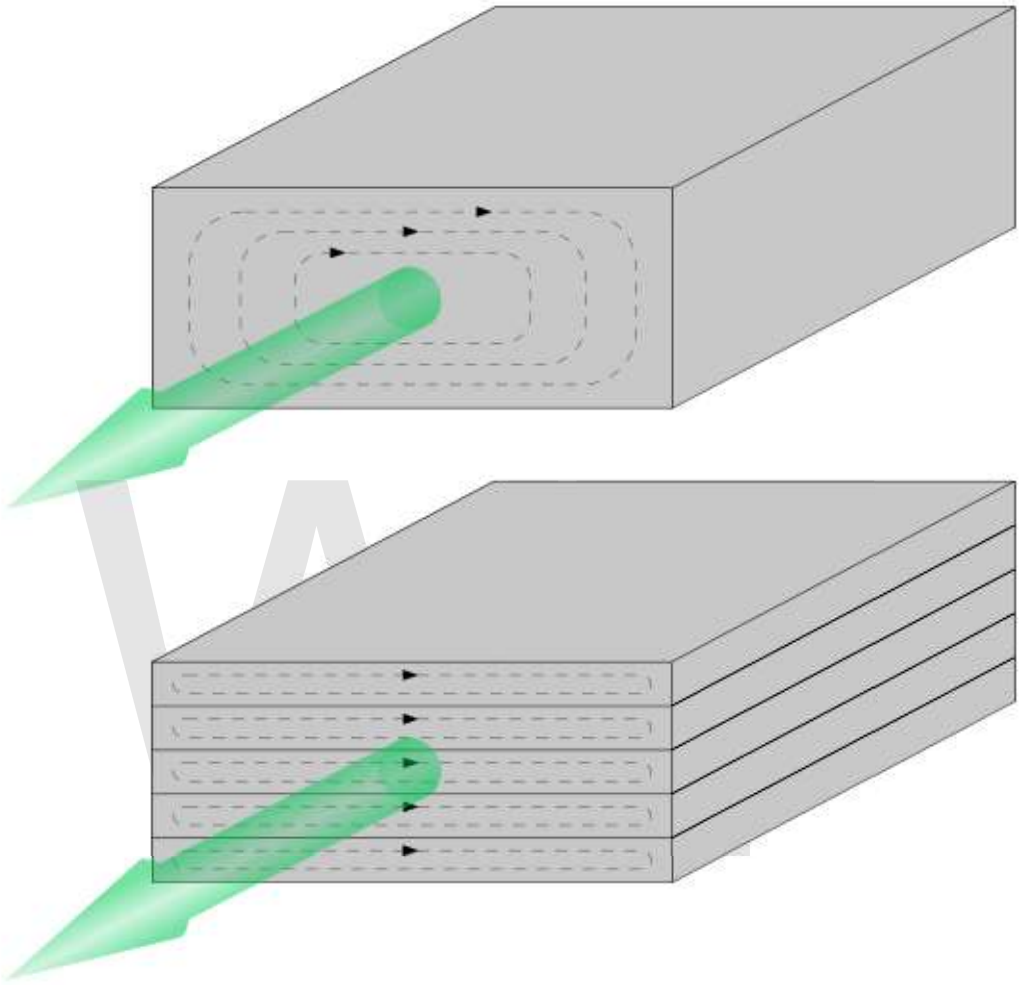
Laminated core transformer showing edge of laminations at top of photo

Laminated steel cores

Transformers for use at power or audio frequencies typically have cores made of high permeability silicon steel. The steel has a permeability many times that of free space, and the core thus serves to greatly reduce the magnetizing current, and confine the flux to a path which closely couples the windings. Early transformer developers soon realized that cores constructed from solid ion resulted in prohibitive eddy-current losses, and their designs mitigated this effect with cores consisting of bundles of insulated iron wires. Later designs constructed the core by stacking layers of thin steel laminations, a principle that has remained in use. Each lamination is insulated from its neighbors by a thin non-conducting layer of insulation. The universal transformer equation indicates a minimum cross-sectional area for the core to avoid saturation.

The effect of laminations is to confine eddy currents to highly elliptical paths that enclose little flux, and so reduce their magnitude. Thinner laminations reduce losses, but are more laborious

and expensive to construct. Thin laminations are generally used on high frequency transformers, with some types of very thin steel laminations able to operate up to 10 kHz.



Laminating the core greatly reduces eddy-current losses

One common design of laminated core is made from interleaved stacks of E-shaped steel sheets capped with I-shaped pieces, leading to its name of "E-I transformer". Such a design tends to exhibit more losses, but is very economical to manufacture. The cut-core or C-core type is made by winding a steel strip around a rectangular form and then bonding the layers together. It is then cut in two, forming two C shapes, and the core assembled by binding the two C halves together with a steel strap. They have the advantage that the flux is always oriented parallel to the metal grains, reducing reluctance.

A steel core's remanence means that it retains a static magnetic field when power is removed. When power is then reapplied, the residual field will cause a high inrush current until the effect of the remaining magnetism is reduced, usually after a few cycles of the applied alternating current. Overcurrent protection devices such as fuses must be selected to allow this harmless

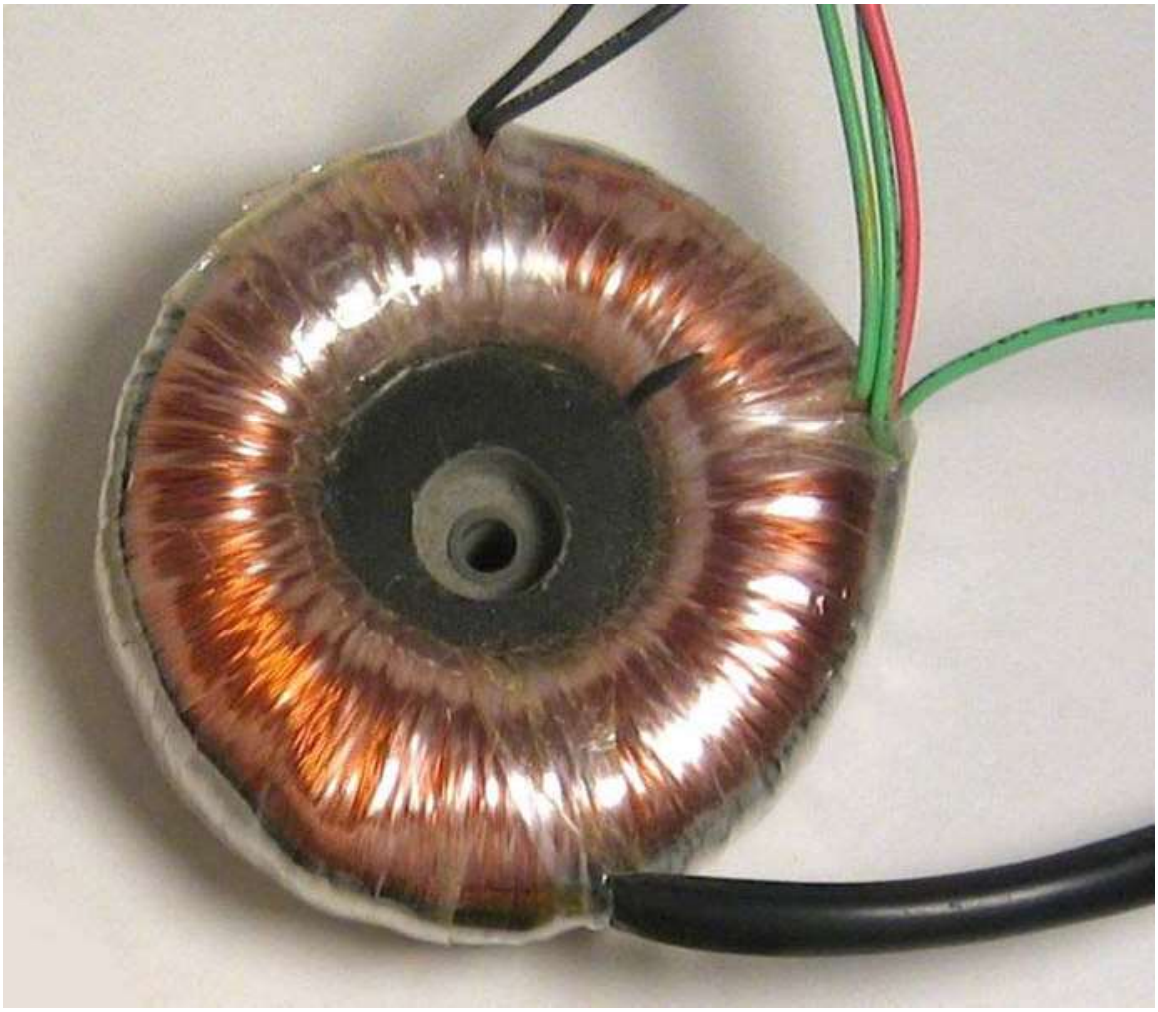
inrush to pass. On transformers connected to long, overhead power transmission lines, induced currents due to geomagnetic disturbances during solar storms can cause saturation of the core and operation of transformer protection devices.

Distribution transformers can achieve low no-load losses by using cores made with low-loss high-permeability silicon steel or amorphous (non-crystalline) metal alloy. The higher initial cost of the core material is offset over the life of the transformer by its lower losses at light load.

Solid cores

Powdered iron cores are used in circuits (such as switch-mode power supplies) that operate above main frequencies and up to a few tens of kilohertz. These materials combine high magnetic permeability with high bulk electrical resistivity. For frequencies extending beyond the VHF band, cores made from non-conductive magnetic ceramic materials called ferrites are common. Some radio-frequency transformers also have movable cores (sometimes called 'slugs') which allow adjustment of the coupling coefficient (and bandwidth) of tuned radio-frequency circuits.

Toroidal cores



Small toroidal core transformer

Toroidal transformers are built around a ring-shaped core, which, depending on operating frequency, is made from a long strip of silicon steel or permalloy wound into a coil, powdered iron, or ferrite. A strip construction ensures that the grain boundaries are optimally aligned, improving the transformer's efficiency by reducing the core's reluctance. The closed ring shape eliminates air gaps inherent in the construction of an E-I core. The cross-section of the ring is usually square or rectangular, but more expensive cores with circular cross-sections are also available. The primary and secondary coils are often wound concentrically to cover the entire surface of the core. This minimizes the length of wire needed, and also provides screening to minimize the core's magnetic field from generating electromagnetic interference.

Toroidal transformers are more efficient than the cheaper laminated E-I types for a similar power level. Other advantages compared to E-I types, include smaller size (about half), lower weight (about half), less mechanical hum (making them superior in audio amplifiers), lower exterior magnetic field (about one tenth), low off-load losses (making them more efficient in standby circuits), single-bolt mounting, and greater choice of shapes. The main disadvantages

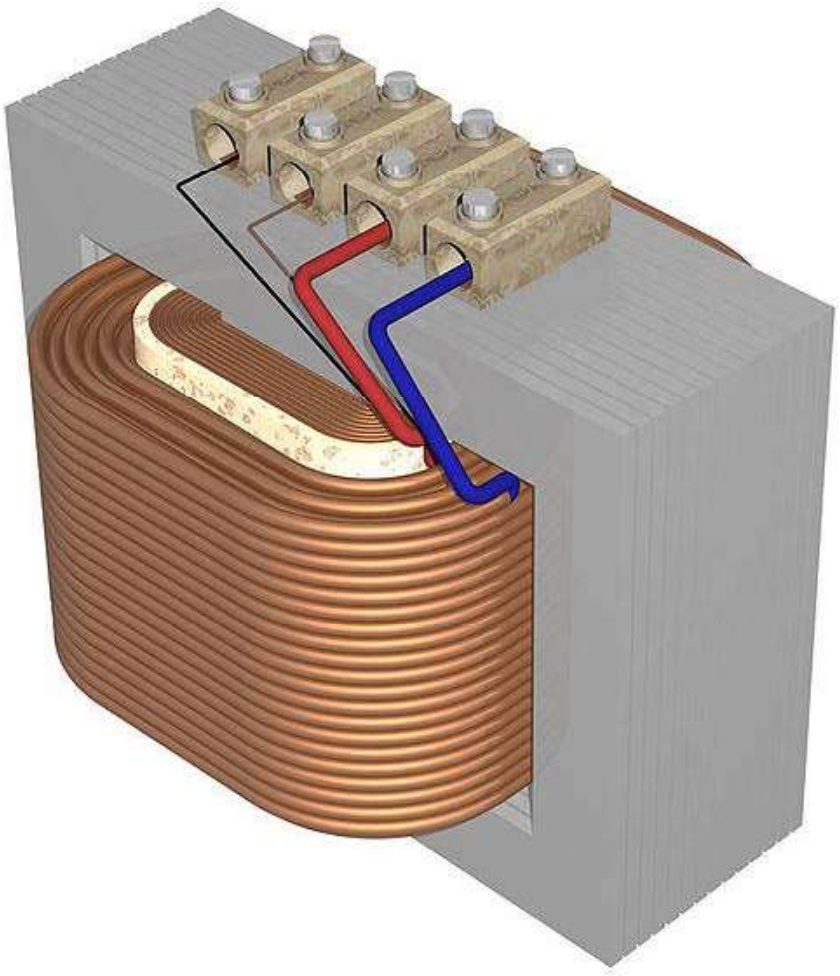
are higher cost and limited power capacity. Because of the lack of a residual gap in the magnetic path, toroidal transformers also tend to exhibit higher inrush current, compared to laminated E-I types.

Ferrite toroidal cores are used at higher frequencies, typically between a few tens of kilohertz to hundreds of megahertz, to reduce losses, physical size, and weight of switch-mode power supplies. A drawback of toroidal transformer construction is the higher labor cost of winding. This is because it is necessary to pass the entire length of a coil winding through the core aperture each time a single turn is added to the coil. As a consequence, toroidal transformers are uncommon above ratings of a few kVA. Small distribution transformers may achieve some of the benefits of a toroidal core by splitting it and forcing it open, then inserting a bobbin containing primary and secondary windings.

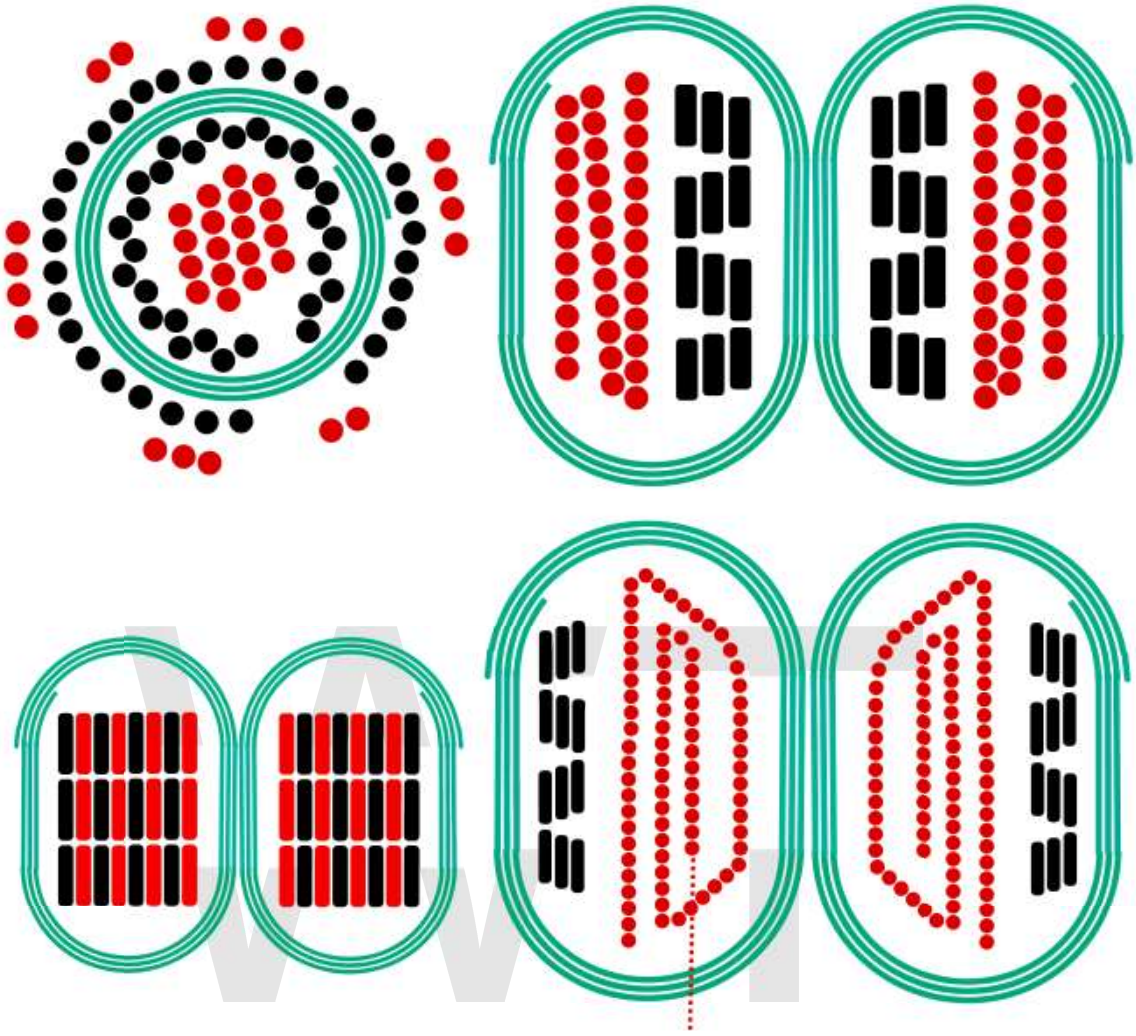
Air cores

A physical core is not an absolute requisite and a functioning transformer can be produced simply by placing the windings near each other, an arrangement termed an "air-core" transformer. The air which comprises the magnetic circuit is essentially lossless, and so an air-core transformer eliminates loss due to hysteresis in the core material. The leakage inductance is inevitably high, resulting in very poor regulation, and so such designs are unsuitable for use in power distribution. They have however very high bandwidth, and are frequently employed in radio-frequency applications, for which a satisfactory coupling coefficient is maintained by carefully overlapping the primary and secondary windings. They're also used for resonant transformers such as Tesla coils where they can achieve reasonably low loss in spite of the high leakage inductance.

Windings



Windings are usually arranged concentrically to minimize flux leakage.



Cut view through transformer windings. White: insulator. Green spiral: Grain oriented silicon steel. Black: Primary winding made of oxygen-free copper. Red: Secondary winding. Top left: Toroidal transformer. Right: C-core, but E-core would be similar. The black windings are made of film. Top: Equally low capacitance between all ends of both windings. Since most cores are at least moderately conductive they also need insulation. Bottom: Lowest capacitance for one end of the secondary winding needed for low-power high-voltage transformers. Bottom left: Reduction of leakage inductance would lead to increase of capacitance.

The conducting material used for the windings depends upon the application, but in all cases the individual turns must be electrically insulated from each other to ensure that the current travels throughout every turn. For small power and signal transformers, in which currents are low and the potential difference between adjacent turns is small, the coils are often wound from enamelled magnet wire, such as Formvar wire. Larger power transformers operating at high voltages may be wound with copper rectangular strip conductors insulated by oil-impregnated paper and blocks of pressboard.

High-frequency transformers operating in the tens to hundreds of kilohertz often have windings made of braided Litz wire to minimize the skin-effect and proximity effect losses. Large power transformers use multiple-stranded conductors as well, since even at low power frequencies non-uniform distribution of current would otherwise exist in high-current windings. Each strand is individually insulated, and the strands are arranged so that at certain points in the winding, or throughout the whole winding, each portion occupies different relative positions in the complete conductor. The transposition equalizes the current flowing in each strand of the conductor, and reduces eddy current losses in the winding itself. The stranded conductor is also more flexible than a solid conductor of similar size, aiding manufacture.

For signal transformers, the windings may be arranged in a way to minimize leakage inductance and stray capacitance to improve high-frequency response. This can be done by splitting up each coil into sections, and those sections placed in layers between the sections of the other winding. This is known as a stacked type or interleaved winding.

Both the primary and secondary windings on power transformers may have external connections, called taps, to intermediate points on the winding to allow selection of the voltage ratio. In distribution transformers the taps may be connected to an automatic on-load tap changer for voltage regulation of distribution circuits. Audio-frequency transformers, used for the distribution of audio to public address loudspeakers, have taps to allow adjustment of impedance to each speaker. A center-tapped transformer is often used in the output stage of an audio power amplifier in a push-pull circuit. Modulation transformers in AM transmitters are very similar.

Certain transformers have the windings protected by epoxy resin. By impregnating the transformer with epoxy under a vacuum, one can replace air spaces within the windings with epoxy, thus sealing the windings and helping to prevent the possible formation of corona and absorption of dirt or water. This produces transformers more suited to damp or dirty environments, but at increased manufacturing cost.

Coolant

Cut-away view of three-phase oil-cooled transformer. The oil reservoir is visible at the top. Radiative fins aid the dissipation of heat.

High temperatures will damage the winding insulation. Small transformers do not generate significant heat and are cooled by air circulation and radiation of heat. Power transformers rated up to several hundred kVA can be adequately cooled by natural convective air-cooling, sometimes assisted by fans. In larger transformers, part of the design problem is removal of heat. Some power transformers are immersed in transformer oil that both cools and insulates the windings. The oil is a highly refined mineral oil that remains stable at transformer operating temperature. Indoor liquid-filled transformers are required by building regulations in many jurisdictions to use a non-flammable liquid, or to be located in fire-resistant rooms. Air-cooled dry transformers are preferred for indoor applications even at capacity ratings where oil-cooled

construction would be more economical, because their cost is offset by the reduced building construction cost.

The oil-filled tank often has radiators through which the oil circulates by natural convection; some large transformers employ forced circulation of the oil by electric pumps, aided by external fans or water-cooled heat exchangers. Oil-filled transformers undergo prolonged drying processes to ensure that the transformer is completely free of water vapor before the cooling oil is introduced. This helps prevent electrical breakdown under load. Oil-filled transformers may be equipped with Buchholz relays, which detect gas evolved during internal arcing and rapidly de-energize the transformer to avert catastrophic failure. Oil-filled transformers may fail, rupture, and burn, causing power outages and losses. Installations of oil-filled transformers usually includes fire protection measures such as walls, oil containment, and fire-suppression sprinkler systems.

Polychlorinated biphenyls have properties that once favored their use as a coolant, though concerns over their environmental persistence led to a widespread ban on their use. Today, non-toxic, stable silicone-based oils, or fluorinated hydrocarbons may be used where the expense of a fire-resistant liquid offsets additional building cost for a transformer vault. Before 1977, even transformers that were nominally filled only with mineral oils may also have been contaminated with polychlorinated biphenyls at 10-20 ppm. Since mineral oil and PCB fluid mix, maintenance equipment used for both PCB and oil-filled transformers could carry over small amounts of PCB, contaminating oil-filled transformers.

Some "dry" transformers (containing no liquid) are enclosed in sealed, pressurized tanks and cooled by nitrogen or sulfur hexafluoride gas.

Experimental power transformers in the 2 MVA range have been built with superconducting windings which eliminates the copper losses, but not the core steel loss. These are cooled by liquid nitrogen or helium.

Insulation drying

Construction of oil-filled transformers requires that the insulation covering the windings be thoroughly dried before the oil is introduced. There are several different methods of drying. Common for all is that they are carried out in vacuum environment. The vacuum makes it difficult to transfer energy (heat) to the insulation. For this there are several different methods. The traditional drying is done by circulating hot air over the active part and cycle this with periods of vacuum (hot-air vacuum drying, HAV). More common for larger transformers is to use evaporated solvent which condenses on the colder active part. The benefit is that the entire process can be carried out at lower pressure and without influence of added oxygen. This process is commonly called vapour-phase drying (VPD).

For distribution transformers, which are smaller and have a smaller insulation weight, resistance heating can be used. This is a method where current is injected in the windings to heat the insulation. The benefit is that the heating can be controlled very well and it is energy efficient. The method is called low-frequency heating (LFH) since the current is injected at a much lower

frequency than the nominal of the grid, which is normally 50 or 60 Hz. A lower frequency reduces the effect of the inductance in the transformer, so the voltage can be reduced.

Terminals

Very small transformers will have wire leads connected directly to the ends of the coils, and brought out to the base of the unit for circuit connections. Larger transformers may have heavy bolted terminals, bus bars or high-voltage insulated bushings made of polymers or porcelain. A large bushing can be a complex structure since it must provide careful control of the electric field gradient without letting the transformer leak oil.

Applications

Image of an electrical substation in Melbourne, Australia showing 3 of 5 220kV/66kV transformers, each with a capacity of 185MVA

A major application of transformers is to increase voltage before transmitting electrical energy over long distances through wires. Wires have resistance and so dissipate electrical energy at a rate proportional to the square of the current through the wire. By transforming electrical power to a high-voltage (and therefore low-current) form for transmission and back again afterward, transformers enable economical transmission of power over long distances. Consequently, transformers have shaped the electricity supply industry, permitting generation to be located

remotely from points of demand. All but a tiny fraction of the world's electrical power has passed through a series of transformers by the time it reaches the consumer.

Transformers are also used extensively in electronic products to step down the supply voltage to a level suitable for the low voltage circuits they contain. The transformer also electrically isolates the end user from contact with the supply voltage.

Signal and audio transformers are used to couple stages of amplifiers and to match devices such as microphones and record players to the input of amplifiers. Audio transformers allowed telephone circuits to carry on a two-way conversation over a single pair of wires. A balun transformer converts a signal that is referenced to ground to a signal that has balanced voltages to ground, such as between external cables and internal circuits.

Chapter 11

Control Relay and Eddy Current Brake

Control relay

A **control relay** is an electromechanical device which activates one or more switches according to the current through a coil not connected to the switches. The thyristor is a semiconductor device which carries out similar functions.

A relay is essentially an electromagnet with two possible states arranged so that when there is sufficient current the core of the relay's coil attracts a ferromagnetic armature which mechanically operates switches; a spring holds the armature away from the core when not actuated. The spring is designed to snap the contacts between two stable mechanical states; there should not be a range of coil current which allows the contacts to be in an intermediate state.

A relay allows circuits to be switched by electrical equipment: for example, a timer circuit with a relay could switch power at a preset time. For many years relays were the standard method of controlling industrial electronic systems. A number of relays could be used together to carry out complex functions (relay logic). The principle of relay logic is based on relays which energize and de-energize associated contacts. Relay logic is the predecessor of ladder logic, which is commonly used in Programmable logic controllers.

Operation

When electric current passes through a coil, magnetic north and south poles are produced across the gap separating the coil and armature. The relay is actuated, or latched, when sufficient current through the coil makes the attractive force between the core and armature overcome the spring tension. The relay remains latched as long as at least a specific value of *holding current*, which can be lower than the actuating current, flows through the coil. When the current through the coil is reduced below this value, the magnetic attraction of the armature becomes too weak to hold it in the actuated position and the spring snaps it to the non-actuated position.

There are many considerations involved in the correct selection of a control relay for a particular application. These considerations include factors such as speed of operation, sensitivity, and hysteresis. Although typical control relays operate in the 5 ms to 20 ms range,

relays with switching speeds as fast as 100 us are available. Reed relays which are actuated by low currents and switch fast are suitable for controlling small currents.

As for any switch, the current through the relay contacts (unrelated to the current through the coil) must not exceed a certain value to avoid damage. In the particular case of high-inductance circuits such as motors other issues must be addressed. When a power source is connected to an inductance, an input surge current which may be several times larger than the steady current exists. When the circuit is broken, the current cannot change instantaneously, which creates a potentially damaging spark across the separating contacts.

Consequently for relays which may be used to control inductive loads we must specify the maximum current that may flow through the relay contacts when it actuates, the *make rating*; the continuous rating; and the *break rating*. The make rating may be several times larger than the continuous rating, which is itself larger than the break rating.

Derating factors

Type of load	**% of rated value**
Resistive	75
Inductive	35
Motor	20
Filament	10
Capacitive	75

Control relays should not be operated above rated temperature because of resulting increased degradation and fatigue. Common practice is to derate 20 degrees Celsius from the maximum rated temperature limit. Relays operating at rated load are also affected by their environment. Oil vapors may greatly decrease the contact tip life, and dust or dirt may cause the tips to burn before their normal life expectancy. Control relay life cycle varies from 50,000 to over one million cycles depending on the electrical loads of the contacts, duty cycle, application, and the extent to which the relay is derated. When a control relay is operating at its derated value, it is controlling a lower value of current than its maximum make and break ratings. This is often done to extend the operating life of the control relay. Table 1 lists the relay and switch derating factors for typical industrial control applications.

Eddy current brake

An eddy current brake of a German ICE 3 in action.

An **eddy current brake**, like a conventional friction brake, is responsible for slowing an object, such as a train or a roller coaster. Unlike friction brakes, which apply pressure on two separate objects, eddy current brakes slow an object by creating eddy currents through electromagnetic induction which create resistance, and in turn either heat or electricity.

Construction and operation

Circular eddy current brake

Circular eddy current brake on 700 Series Shinkansen

Electromagnetic brakes are similar to electrical motors; non-ferromagnetic metal discs (rotors) are connected to a rotating coil, and a magnetic field between the rotor and the coil creates a resistance used to generate electricity or heat. When electromagnets are used, control of the braking action is made possible by varying the strength of the magnetic field. A braking force is possible when electric current is passed through the electromagnets. The movement of the metal through the magnetic field of the electromagnets creates eddy currents in the discs. These eddy currents generate an opposing magnetic field, which then resists the rotation of the discs,

providing braking force. The net result is to convert the motion of the rotors into heat in the rotors.

Japanese Shinkansen trains had employed circular eddy current brake system on trailer cars since 100 Series Shinkansen. However, N700 Series Shinkansen abandoned eddy current brakes in favour of regenerative brakes since 14 of the 16 cars in the trainset used electric motors.

Linear eddy current brake

The principle of the linear eddy current brake has been described by the French physicist Foucault, hence in French the eddy current brake is called the "frein à courants de Foucault".

The linear eddy current brake consists of a magnetic yoke with electrical coils positioned along the rail, which are being magnetized alternating as south and north magnetic poles. This magnet does not touch the rail, as with the magnetic brake, but is held at a constant small distance from the rail (approximately seven millimeters). It does not move along the rail, exerting only a vertical pull on the rail.

When the magnet is moved along the rail, it generates a non-stationary magnetic field in the head of the rail, which then generates electrical tension (Faraday's induction law), and causes eddy currents. These disturb the magnetic field in such a way that the magnetic force is diverted to the opposite of the direction of the movement, thus creating a horizontal force component, which works against the movement of the magnet.

The braking energy of the vehicle is converted in eddy current losses which lead to a warming of the rail. (The regular magnetic brake, in wide use in railways, exerts its braking force by friction with the rail, which also creates heat.)

The eddy current brake does not have any mechanical contact with the rail, and thus no wear, and creates no noise or odor. The eddy current brake is unusable at low speeds, but can be used at high speeds both for emergency braking and for regular braking.

The TSI (Technical Specifications for Interoperability) of the EU for trans-European high speed rail recommends that all newly built high speed lines should make the eddy current brake possible.

Eddy current brakes at the Intamin roller coaster *Goliath* in Walibi World (Netherlands)

The first train in commercial circulation to use such a braking is the ICE 3.

Modern roller coasters use this type of braking, but utilize permanent magnets instead of electromagnets, and require no electricity. However, their braking strength cannot be adjusted as easily as with an electromagnet.

Chapter 12

Electrical Steel

Electrical steel, also called **lamination steel**, **silicon electrical steel**, **silicon steel** or **transformer steel**, is specialty steel tailored to produce certain magnetic properties, such as a small hysteresis area (small energy dissipation per cycle, or low core loss) and high permeability.

The material is usually manufactured in the form of cold-rolled strips less than 2 mm thick. These strips are called laminations when stacked together to form a core. Once assembled, they form the laminated cores of transformers or the stator and rotor parts of electric motors. Laminations may be cut to their finished shape by a punch and die, or in smaller quantities may be cut by a laser, or by wire erosion.

Metallurgy

Electrical steel is an iron alloy which may have from zero to 6.5% silicon (Si:5Fe). Silicon significantly increases the electrical resistivity of the steel, which decreases the induced eddy currents and thus reduces the core loss. Manganese and aluminum can be added up to 0.5%.

Increasing the amount of silicon inhibits eddy currents and narrows the hysteresis loop of the material, thus lowering the core losses. However, the grain structure hardens and embrittles the metal, which adversely affects the workability of the material, especially when rolling it. When alloying, the concentration levels of carbon, sulfur, oxygen and nitrogen must be kept low, as these elements indicate the presence of carbides, sulfides, oxides and nitrides. These compounds, even in particles as small as one micrometer in diameter, increase hysteresis losses while also decreasing magnetic permeability. The presence of carbon has a more detrimental effect than sulfur or oxygen. Carbon also causes magnetic aging when it slowly leaves the solid solution and precipitates as carbides, thus resulting in an increase in power loss over time. For these reasons, the carbon level is kept to 0.005% or lower. The carbon level can be reduced by annealing the steel in a decarburizing atmosphere, such as hydrogen.

Physical properties examples

Melting point: ~1,500 °C (example for ~3.1% silicon content)

Density: 7,650 kg/m^3 (example for 3% silicon content)

Resistivity: 47.2×10^{-8} (Ω·m) (example for 3% silicon content)

Grain orientation

There are two main types of electrical steel: grain-oriented and non-oriented.

Grain-oriented electrical steel usually has a silicon level of 3% (Si:11Fe). It is processed in such a way that the optimum properties are developed in the rolling direction, due to a tight control (proposed by Norman P. Goss) of the crystal orientation relative to the sheet. Due to the special orientation, the magnetic flux density is increased by 30% in the coil rolling direction, although its magnetic saturation is decreased by 5%. It is used for the cores of high-efficiency transformers, electric motor and generators. Cold Rolled Grain-oriented steel is often abbreviated to CRGO.

Non-oriented electrical steel usually has a silicon level of 2 to 3.5% and has similar magnetic properties in all directions, which makes it isotropic. It is less expensive and is used in applications where the direction of magnetic flux is changing, such as electric motors and generators. It is also used when efficiency is less important or when there is insufficient space to correctly orient components to take advantage of the anisotropic properties of grain-oriented electrical steel. Cold Rolled Non Grain-oriented steel is often abbreviated to CRNGO.

Amorphous steel

Transformers with amorphous steel cores can have core losses of one-third that of conventional steels. This material is a metallic glass prepared by pouring molten alloy steel onto a rotating cooled wheel, which cools the metal at a rate of about one megakelvin per second, so fast that crystals do not form. Amorphous steel has poorer mechanical properties and as of 2010 costs about twice as much as conventional steel, making it cost-effective only for some large distribution-type transformers.

Lamination coatings

Electrical steel is usually coated to increase electrical resistance between laminations, to provide resistance to corrosion or rust, and to act as a lubricant during die cutting. There are various coatings, organic and inorganic, and the coating used depends on the application of the steel. The type of coating selected depends on the heat treatment of the laminations, whether the finished lamination will be immersed in oil, and the working temperature of the finished apparatus. Very early practice was to insulate each lamination with a layer of paper or a varnish coating, but this reduced the stacking factor of the core and limited the maximum temperature of the core.

Magnetic properties

The magnetic properties of electrical steel are dependent on heat treatment, as increasing the average crystal size decreases the hysteresis loss. Hysteresis loss is determined by a standard test and for common grades of electrical steel may range from about 2 to 10 watts per kilogram

(1 to 5 watts per pound) at 60 Hz and 1.5 tesla magnetic field strength. Semi-processed electrical steels are delivered in a state that, after punching the final shape, a final heat treatment develops the desired 150-micrometer grain size. The fully processed steels are usually delivered with insulating coating, full heat treatment, and defined magnetic properties, for applications where the punching operation does not significantly degrade the material properties. Excessive bending, incorrect heat treatment, or even rough handling of core steel can adversely affect its magnetic properties and may also increase noise due to magnetostriction

Magnetic properties of electrical steels are tested using the internationally standardised Epstein frame method.

Practical concerns

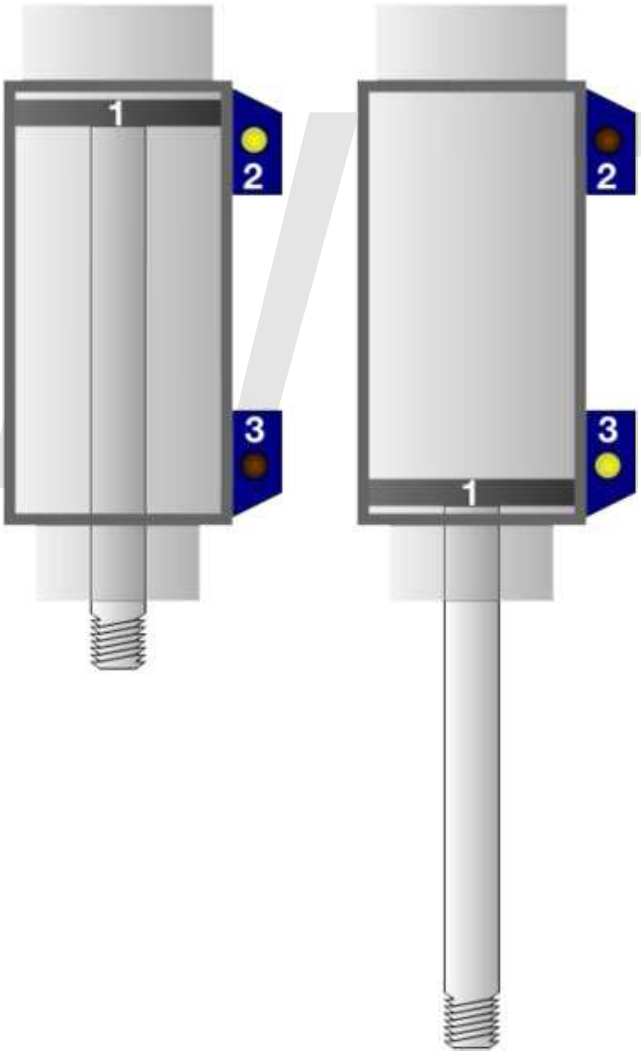
Core steel is much more costly than mild steel—in 1981 it was more than twice the cost per unit weight.

The size of magnetic domains in the sheet can be reduced by scribing the surface of the sheet with a laser, or mechanically. This greatly reduces the hysteresis losses in the assembled core.

Chapter 13

Hall Effect Sensor and Induction Loop

Hall effect sensor



The magnetic piston (1) in this pneumatic cylinder will cause the Hall effect sensors (2 and 3) mounted on its outer wall to activate when it is fully retracted or extended.

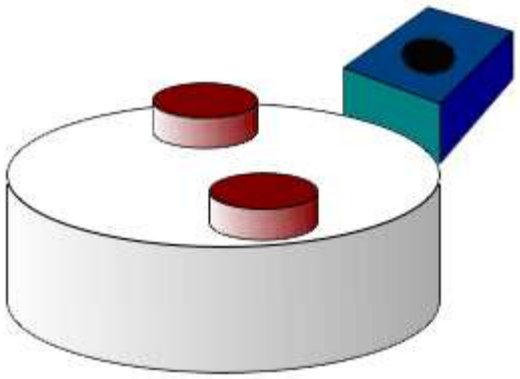
Clutch with Hall Effect sensor.

A **Hall effect sensor** is a transducer that varies its output voltage in response to changes in magnetic field. Hall sensors are used for proximity switching, positioning, speed detection, and current sensing applications.

In its simplest form, the sensor operates as an analogue transducer, directly returning a voltage. With a known magnetic field, its distance from the Hall plate can be determined. Using groups of sensors, the relative position of the magnet can be deduced.

Electricity carried through a conductor will produce a magnetic field that varies with current, and a Hall sensor can be used to measure the current without interrupting the circuit. Typically, the sensor is integrated with a wound core or permanent magnet that surrounds the conductor to be measured.

Frequently, a Hall sensor is combined with circuitry that allows the device to act in a digital (on/off) mode, and may be called a switch in this configuration. Commonly seen in industrial applications such as the pictured pneumatic cylinder, they are also used in consumer equipment; for example some computer printers use them to detect missing paper and open covers. When high reliability is required, they are used in keyboards.



Hall sensors are commonly used to time the speed of wheels and shafts, such as for internal combustion engine ignition timing, tachometers and anti-lock braking systems. They are used in brushless DC electric motors to detect the position of the permanent magnet. In the pictured wheel with two equally spaced magnets, the voltage from the sensor will peak twice for each revolution. This arrangement is commonly used to regulate the speed of disc drives.

Hall probe

A hall probe contains an indium compound crystal such as indium antimonide, mounted on an aluminum backing plate, and encapsulated in the probe head. The plane of the crystal is perpendicular to the probe handle. Connecting leads from the crystal are brought down through the handle to the circuit box.

When the Hall Probe is held so that the magnetic field lines are passing at right angles through the sensor of the probe, the meter gives a reading of the value of magnetic flux density (B). A current is passed through the crystal which, when placed in a magnetic field has a “Hall Effect” voltage developed across it. The Hall Effect is seen when a conductor is passed through a uniform magnetic field. The natural electron drift of the charge carriers causes the magnetic field to apply a Lorentz force (the force exerted on a charged particle in an electromagnetic field) to these charge carriers. The result is what is seen as a charge separation, with a build up of either positive or negative charges on the bottom or on the top of the plate. The crystal measures 5 mm square. The probe handle, being made of a non-ferrous material, has no disturbing effect on the field.

A Hall Probe is enough to measure the Earth's magnetic field. It must be held so that the Earth's field lines are passing directly through it. It is then rotated quickly so the field lines pass through the sensor in the opposite direction. The change in the flux density reading is double the Earth's magnetic flux density. A hall probe must first be calibrated against a known value of magnetic field strength. For a solenoid the hall probe is placed in the center.

Hall Effect Sensor Interface

Hall effect sensor may require analog circuitry to be interfaced to microprocessors. These interfaces may include inpu diagnostics, fault protection for transient conditions and short/open

circuit detection. It may also provide and monitor the current to the hall effect sensor itself. There are precision IC products available to handle these features.

Induction loop

Induction loop is a term used to describe an electromagnetic communication- and detection system, relying on the fact that a moving magnet will induce an electrical current in a nearby conducting wire. Induction loops are used for transmission and reception of communication signals, or for detection of metal objects in metal detectors or vehicle presence indicators. A common modern use for induction loops is to provide hearing assistance to hearing-aid users.

Implementation

An example of the Inductance loop installed in the road for cars and bikes

The "aerial" system of an induction loop installation can consist of one or more loops of a conductive element.

In industrial applications this might be a large single- or multi-turn, loop, or a complex multi-lobed, phase coincident sub-loop design, most effectively mounted above the required reception area in industrial applications.

An audio induction loop might have one or more loops sometimes with a phase shift between them, and either near to or around the area in which a hearing aid user would be present. Many different configurations can be used depending on the application.

Such an induction loop receiver is classically a very small iron-cored inductor (telecoil), although Rediffusion demonstrated a prototype Hall-Effect system in its PLL FM system.

The system commonly uses an analogue power amplifier matched to the low impedance of the transmission loop. The transmission is normally direct rather than superimposed or modulated upon a carrier, though multi-channel systems have been implemented using modulation.

Vehicle detection (inductive) loops are used to count vehicles passing or arriving at a certain point, for instance approaching a traffic light, and in motorway traffic management. An insulated, electrically conducting loop is installed under the road. An electrical voltage is generated when a ferrous (containing iron or steel) body passes close to the wire/loop

Other definitions

A different sort of "induction loop" is applied to metal detectors, where a large coil, which forms part of a resonant circuit, is effectively "detuned" by the coil's proximity to a conductive object. The detected object may be metallic, (metal and cable detection) or conductive/capacitive (stud/cavity detection) Other configurations of this equipment use two or more receiving coils, and the detected object modifies the inductive coupling or alters the phase angle of the voltage induced in the receiving coils relative to the oscillator coil.

An increasingly common application is for providing hearing aid-compatible "assistive listening" telecoil. In this application a loop or series of loops is used to provide an audio frequency oscillating magnetic field in an area where a hearing aid user may be present. Many hearing aids contain a telecoil which allows the user to receive and hear the magnetic field and remove the normal audio signal provided from the hearing aid microphone site. These loops are often referred to as a hearing loop or Audio induction loop.

Modern day applications

- vehicle detection (e.g. at traffic lights)
- car park Parking Guidance and Information systems
- metal detectors
- audio induction loops

Historical applications

- An Anti-submarine indicator loop was a device used to detect submarines and surface vessels using specially designed submerged cables connected to a galvanometer.

Chapter 14

Magnetic Amplifier

The **magnetic amplifier** (colloquially known as a "mag amp") is an electromagnetic device for amplifying electrical signals. The magnetic amplifier was invented early in the 20th century, and was used as an alternative to vacuum tube amplifiers where robustness and high current capacity were required. World War II Germany perfected this type of amplifier, and it was used for instance in the V-2 rocket. The magnetic amplifier was most prominent in power control and low-frequency signal applications from 1947 to about 1957, when the transistor began to supplant it. The magnetic amplifier has now been largely superseded by the transistor-based amplifier, except in a few safety critical, high reliability or extremely demanding applications. Combinations of transistor and mag-amp techniques are still used.

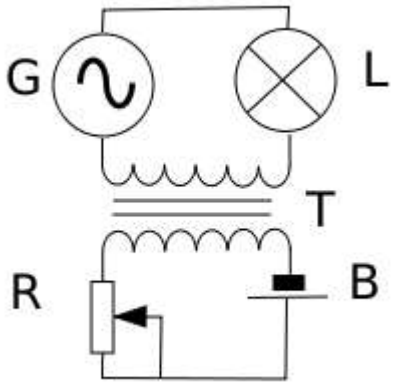
Strengths

The magnetic amplifier is a static device with no moving parts. It has no wear-out mechanism and has a good tolerance to mechanical shock and vibration. It requires no warm-up time. Multiple isolated signals may be summed by additional control windings on the magnetic cores. The windings of a magnetic amplifier have a higher tolerance to momentary overloads than comparable solid-state devices. The magnetic amplifier is also used as a transducer in applications such as current measurement and the flux gate compass

Limitations

The gain available from a single stage is limited and low compared to electronic amplifiers. Frequency response of a high gain amplifier is limited to about one-tenth the excitation frequency, although this is often mitigated by exciting magnetic amplifiers with currents at higher than utility frequency. Solid-state amplifiers can be more compact and efficient than magnetic amplifiers. The bias and feedback windings are not unilateral, and may couple energy back from the controlled circuit into the control circuit. This complicates the design of multistage amplifiers when compared with electronic devices.

Principle of operation



A saturable reactor, illustrating the principle of a magnetic amplifier

Visually a mag amp device may resemble a transformer but the operating principle is quite different from a transformer - essentially the mag amp is a saturable reactor. It makes use of magnetic saturation of the core, a non-linear property of a certain class of transformer cores. For controlled saturation characteristics the magnetic amplifier employs core materials that have been designed to have a specific B-H curve shape that is highly rectangular, in contrast to the slowly-tapering B-H curve of softly saturating core materials that are often used in normal transformers.

The typical magnetic amplifier consists of two physically separate but similar transformer magnetic cores, each of which has two windings - a control winding and an AC winding. A small DC current from a low impedance source is fed into the series-connected control windings. The AC windings may be connected either in series or in parallel, the configurations resulting in different types of mag amps. The amount of control current fed into the control winding sets the point in the AC winding waveform at which either core will saturate. In saturation, the AC winding on the saturated core will go from a high impedance state ("off") into a very low impedance state ("on") - that is, the control current controls at which voltage the mag amp switches "on".

A relatively small DC current on the control winding is able to control or switch large AC currents on the AC windings. This results in current amplification.

Applications

Magnetic amplifiers were important as modulation and control amplifiers in the early development of voice transmission by radio. A magnetic amplifier was used as voice modulator for a 2 kilowatt Alexanderson alternator, and magnetic amplifiers were used in the keying circuits of large high-frequency alternators used for radio communications. Magnetic amplifiers were also used to regulate the speed of Alexanderson alternators to maintain the accuracy of the transmitted radio frequency.

The ability to control large currents with small control power made magnetic amplifiers useful for control of lighting circuits, for stage lighting and for advertising signs. Saturable reactor amplifiers were used for control of power to industrial furnaces. Small magnetic amplifiers were used for radio tuning indicators, control of small motor and cooling fan speed, control of battery chargers.

Magnetic amplifiers were used extensively as the switching element in early switched-mode (SMPS) power supplies, as well as in lighting control. Semiconductor based solid-state switches have largely superseded them, though recently there has been some regained interest in using mag amps in compact and reliable switching power supplies. PC ATX power supplies often use mag amps for secondary side voltage regulation.

Magnetic amplifiers are still used in some arc welders.

Magnetic amplifier transformer cores designed specifically for switch mode power supplies are currently manufactured by several large electromagnetics companies, including Metglas and Mag-Inc.

Magnetic amplifiers can be used for measuring high DC-voltages without direct connection to the high voltage and are therefore still used in the HVDC-technique.

Magnetic amplifiers were used by locomotives to detect wheel slip, until replaced by Hall Effect current transducers. The cables from two traction motors passed through the core of the device. During normal operation the resultant flux was zero as both currents were the same and in opposite directions. However, the currents would differ during wheel slip, producing a resultant flux that acted as the Control winding, developing a voltage across a resistor in series with the AC winding which was sent to the wheel slip correction circuits.

History

Early development

A voltage source and a series connected variable resistor may be regarded as a direct current signal source for a low resistance load such as the control coil of a saturable reactor which amplifies the signal. Thus, in principle, a saturable reactor is already an amplifier, although before 20th century they were used for simple tasks, such as controlling lighting and electrical machinery as early as 1885.

In the early 20th Century, the General Electric Company, under the direction of engineer E. F. W. Alexanderson, developed a system of transoceanic radio communications, using continuous wave transmission over great distances. Alexanderson drew upon the work of Nikola Tesla and Reginald Fessenden as the inspiration for his system.

The result of this work was the 2 kW Alexanderson alternator, which produced radio frequencies from 50 to 100 kHz and which critics had previously denounced as impractical.

Later, Guglielmo Marconi took a vested interest in the project and, in 1915, witnessed a demonstration of a new, 50 kW, 50 kHz alternator.

The experimental telegraphy and telephony demonstrations made during 1917 attracted the attention of the US Government, especially in light of partial failures in the transoceanic cable that snaked across the bottom of the Atlantic Ocean. The 50 kW alternator was commandeered by the US Navy and put into service in January 1918 and was used until 1920, when a 200 kW generator-alternator set was built and installed.

Usage in radio

Magnetic amplifiers were used early on to control large, high-power alternators by turning them on and off for telegraphy or to vary the signal for voice modulation. However, the alternator's frequency limits were rather low to where a frequency multiplier had to be utilized to generate higher radio frequencies than the alternator was capable of producing. Even so, early magnetic amplifiers incorporating powdered-iron cores were incapable of producing radio frequencies above approximately 200 kHz. Other core materials, such as ferrite cores and oil-filled transformers, would have to be developed to allow the amplifier to produce higher frequencies.

Usage in aircraft

Magnetic amplifiers were used in aircraft systems (avionics) before the advent of high reliability semiconductors. They were important in implementing early autoland systems and Concorde made use of the technology for the control of its engine air intakes before subsequent development of a replacement system using digital electronics.

Usage in computing

Magnetic amplifiers were widely studied during the 1950s as a potential switching element for mainframe computers. Mag amps could be used to sum several inputs in a single core, which was very useful in the arithmetic logic unit (ALU). Custom tubes could do the same, but transistors could not, so the mag amp was able to combine the advantages of tubes and transistors in an era when the latter were expensive and unreliable.

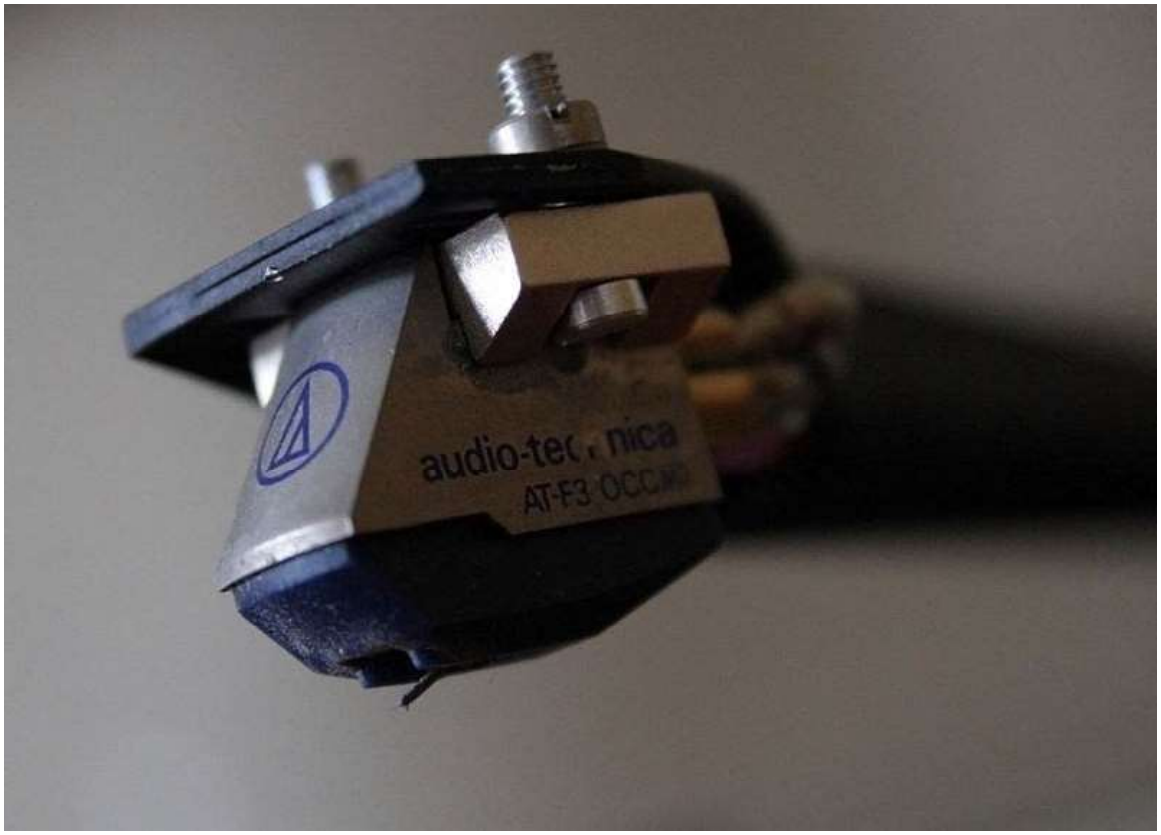
However, that era was very short, lasting from the mid 1950s to about 1960, at which point new fabrication techniques were producing great improvements in transistors and dramatically lowering their price points. Only one large-scale mag amp machine was put into production, the UNIVAC Solid State, but a number of contemporary late-1950's/early-1960s computers made some use of the technology, like the Ferranti Orion.

Misnomer uses

In the 1970s, Robert Carver designed and produced several high quality high-powered audio amplifiers, calling them magnetic amplifiers. In fact, they were in most respects conventional audio amplifier designs with an unusual power supply circuit.

Chapter 15

Magnetic Cartridge



An Audio Technica AT-F3 MC cartridge

A **magnetic cartridge** is a transducer used for the playback of gramophone records on a turntable or phonograph. It converts mechanical vibrational energy from a stylus riding in a spiral record groove into an electrical signal that is subsequently amplified and then converted back to sound by a loudspeaker system.

History

The first electric pick-ups were developed in about 1925. They used a piezo-electric crystal of quartz, stimulated by a stylus made of sapphire or diamond. The magnetic cartridge is presently

the most common form of sound pickup used and came into use in the 1950s, following the introduction of magnetic cutter heads around 1945 for mastering records.

Types

In high-fidelity systems, crystal and ceramic pickups have been replaced by the magnetic cartridge, using either a **moving magnet** or a **moving coil**.

Compared to the crystal and ceramic pickups, the magnetic cartridge gives improved playback fidelity and reduces record wear by tracking the groove with lighter pressure. Magnetic cartridges use much lower tracking forces and thus damage the record grooves less. They also have a lower output voltage than a crystal or ceramic pickup, in the range of only a few millivolts, thus requiring greater amplification.

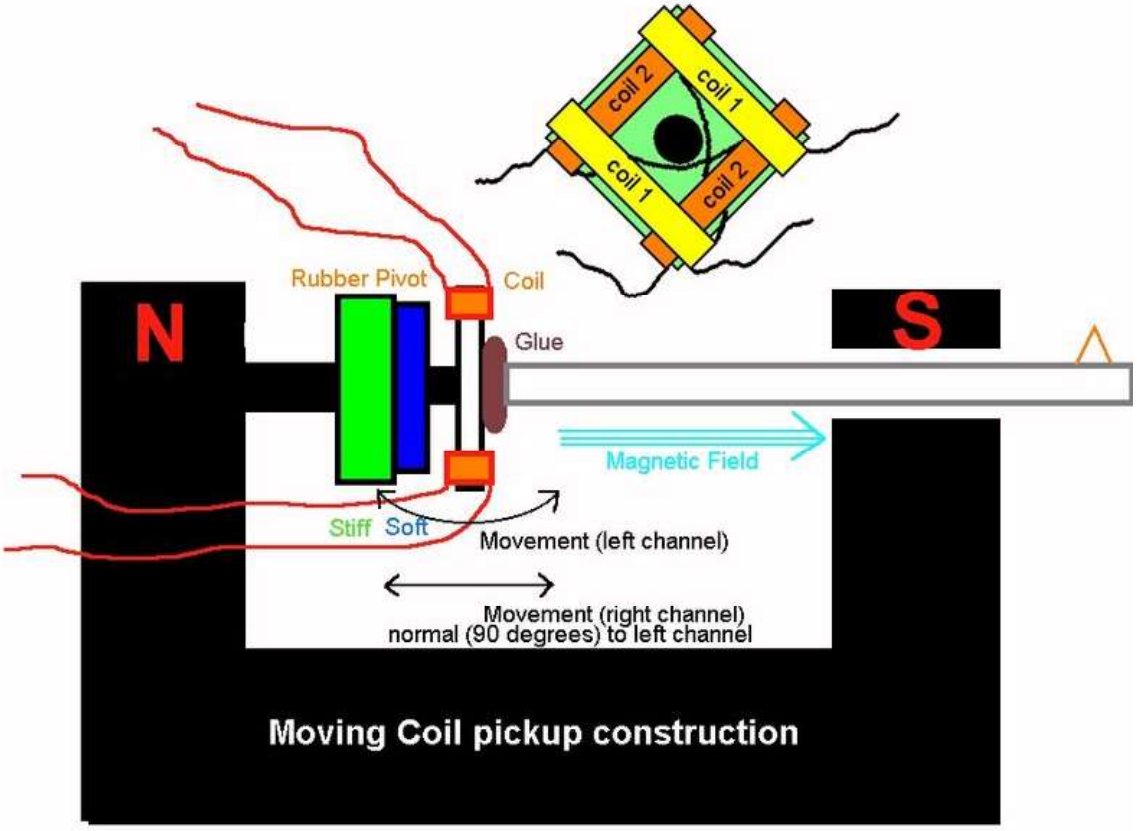
Moving Magnet (MM) cartridges

In a moving magnet cartridge, the stylus cantilever carries a tiny permanent magnet, which is positioned between two sets of fixed coils (in a stereophonic cartridge), forming a tiny electromagnetic generator. As the magnet vibrates in response to the stylus following the record groove, it induces a tiny current in the coils.

Because the magnet is small and has little mass, and is not coupled mechanically to the generator (as in a ceramic cartridge), a properly adjusted stylus follows the groove more faithfully while requiring less tracking force (the downward pressure on the stylus).

There is a sub-category. Moving iron and induced magnet types (ADC being a well known example) which have the magnet fixed and move a piece of iron or other ferrous alloy in the field of the magnet to produce the signal within the fixed coils.

Moving Coil (MC) cartridges



The MC design is again a tiny electromagnetic generator, but (unlike an MM design) with the magnet and coils reversed: the coils are attached to the stylus, and move within the field of a permanent magnet. The coils are tiny and made from very fine wire, so are even lighter than the small magnet used in an MM cartridge, thus improving the tracking ability of the cartridge. This can give extended frequency response as well as greater fidelity.

A disadvantage however is that moving-coil cartridges generate an even lower voltage signal than a moving-magnet type cartridge. This is because the moving coil cannot be large enough (it would be too heavy) to generate equivalent voltage levels. The resulting signal is only a few hundred microvolts, and thus more easily swamped by noise, induced hum, etc. Thus it is more challenging to design a preamplifier with the extremely low noise inputs needed for moving-coil cartridge, therefore a "step up transformer" is sometimes used instead.

Moving coil cartridges are extremely small precision instruments and are therefore generally expensive, but are frequently preferred by audiophiles due to their better performance.

Moving Micro Cross (MMC) cartridges

The MMC design was invented and patented by Bang & Olufsen. Since it often uses a special mounting, it can be mostly found in Bang & Olufsen turntables (which also cannot use another type of cartridge). Apart from being produced for the Bang & Olufsen mounting system, the SP12 and SP14 were also available in standard 1/2" mount.

The MMC cartridge is a *moving iron* design. Magnets and coils are stationary while the *micro cross* moves with the stylus, thereby varying the distances between the arms of the cross and the magnets. The design obviously offers more freedom concerning magnet and coil mass (compared to MC and MM cartridges). For example, the MMC20 (presented in 1978) uses four coils wound around the magnetic cores with 1200 turns each. Minimizing the moving mass also reduces the unavoidable wear on the records.

It is also claimed that the MMC design allows for superior channel separation, since each channel's movements appear on a separate axis.

Moving Magnet vs. Moving Coil debate

Moving magnet cartridges are more commonly found at the 'lower-end' of the market, while the 'higher-end' tends to be dominated by moving coil designs. The debate as to whether MM or MC designs can ultimately produce the better sound is often heated and subjective. The distinction between the two is often blurred by cost and design considerations - i.e. can an MC cartridge requiring another step-up amplification outperform well made MM cartridges that need simpler front-end stages? Every now and then a design comes along to re-open this debate. A good example being the Linn K9 (now discontinued) - regarded by some as one of the better MM designs and competitive with MC alternatives costing more. Amongst others, Grace, ADC and Grado also manufactured notably good designs.

"Decca" Cartridges (aka "Moving Iron")

The Decca phono cartridges were a unique design, with fixed magnets and coils. The stylus shaft was composed of the diamond tip, a short piece of soft iron, and an L-shaped cantilever made of non-magnetic steel. Since the iron was placed very close to the tip (within 1 mm), the motions of the tip could be tracked very accurately. Decca engineers called this "positive scanning". Vertical and lateral compliance was controlled by the shape and thickness of the cantilever. Decca cartridges had a reputation for being very musical; however early versions required more tracking force than competitive designs - making record wear a concern.

Stereo reproduction

One reason that magnetic cartridges superseded the crystal pick-up was the relative ease with which it could be made to reproduce stereo recordings, which were introduced in 1958. In a stereo recording, the two channels are arranged to drive the record cutter head at an angle of 45° to the vertical, effectively encoding each channel in the left and right V-shaped walls of the record groove. This system worked well, since it provided full compatibility with a monaural

pick-up, so stereo records could be played on older mono equipment. To reproduce the stereo signal, the cartridge simply arranges pairs of coils at 45° to complement the cutting process. With careful design, the coils can be shielded from each other electrically and mechanically such that stereo separation is maximised.

Comparison with crystal technology

Piezoelectric crystal or ceramic pickups had a few clear advantages. They were much easier and cheaper to make and were more robust than the delicate magnetic pickup. In addition, the output voltage from a crystal pickup is relatively large, requiring less amplification, which helped improve the signal-to-noise ratio. However, the signal from a crystal is not an accurate reproduction of the recording, as there is a lot of distortion introduced. The stylus is coupled to the crystal in a fairly rigid manner, which is also not as good at following rapid changes in the record grooves, so frequency response also suffers. This requires a greater tracking force, which in turn wears the records out faster. (This has earned cheap portable record players, the nickname "portable grinding wheel", in some circles.)

By contrast, since the magnetic cartridge is not mechanically coupled, the stylus and lever arm weight can be made exceedingly small. This gives extended frequency response and low distortion. The distortion is further minimised by the fact that there is more inherent linearity in the induction principle than there is in the piezo-electric one. Since the lighter stylus requires very low tracking forces, it requires a more sophisticated counterweighted arm, but reduces record wear.

The output from the magnetic cartridge is only a few millivolts compared to several tenths of a volt from a crystal or ceramic pickup. This requires an additional preamplifier stage. Careful design and shielding in the signal cables and amplifier is needed to prevent unwanted noise (shot noise or EMI). The magnetic induction principle also naturally leads to a linearly rising response with increasing frequency, and this needs to be compensated for to give correct (flat) frequency reproduction. Conversely, the very low bass frequencies are not efficiently picked up, so a strong bass boost is needed. This can amplify unwanted low frequency noise such as that from the turntable motor and drive mechanism itself (rumble). Crystal pickups do not suffer from these drawbacks and give a much better bass performance, so they may be preferred in some applications (such as DJ-ing), where robustness and good bass are favoured over highest fidelity reproduction. The moving coil pickups tend to have an even lower level of output, and so usually require a step-up transformer or very special preamplifiers to bring these signals up to the level of input that a standard amplifier requires. These special preamplifiers must have very low noise indeed (some have even been cryogenically cooled), and hence tend to be expensive. Many audiophiles claim that the benefits are worthwhile. There are higher output moving coil pickup cartridges that can be used directly into a normal moving magnet input, but these are in the minority.

The bass boost and high frequency rolloff can be conveniently incorporated into the preamplifier which implements the RIAA equalization curve, a noise reduction technique used on all modern records.

The magnetic nature of the modern pickup means it must be shielded from external fields, especially from the loudspeakers of the same system, or unpleasant and possibly damaging feedback can occur. For this reason, the cartridge itself has a shield, usually of mu-metal, to help screen out unwanted fields.

The stylus

The stylus, or "needle", is a crucial part of the record player (or 'phonograph' or 'gramophone', both terms now archaic), as it is the one part of the system that actually contacts the recorded disc and transfers its vibrations to the rest of the system. It is the part which also suffers the greatest wear. There are two desired qualities in a stylus: first, that it faithfully follows the contours of the recorded groove and transfers the vibration to the system, and second, that it does not damage the recorded disc.

History of materials

Early phonograph styli in mechanical players were steel, or even fibre, needles, usually with a shank about 1/8" (3 mm) in diameter, ground to a sharp point. These were easily replaceable by the user, as they had a very limited life and wore out fairly rapidly with use. Extensive play tended to wear records out as well as needles.

When the electronic phonograph was introduced, styli were included as part of the pickup cartridge.

In early times, wear of cylinders and glass styli was problematic. By 1908, sapphire styli were being attached to the rice paper diaphragms, and in 1912 Edison began to use the diamond stylus. Typical low cost crystal cartridges of the 1950s tracked at 5-8 grams of force and used replaceable osmium tipped steel styli called *needles*, which styli might last only five to ten plays before needing replacement. To help make the stylus last longer, sapphire styli for 78 rpm records or diamond styli for 33 1/3 rpm LPs were re-introduced. A 78 rpm stylus typically has a 3 mil (0.08 mm) diameter while a stylus for 33 1/3 rpm *microgroove* LP discs has a 1 mil (0.03 mm) diameter tip to fit the narrower groove. Sapphire might be good for 40 or 50 hours while a diamond would last at least ten times that long. Typically, these early cartridges were of the "flip-over" type; the cartridge had a stylus on either side, one for 78 rpm discs, the other for 33 and 45 rpm ("microgroove") records. The entire cartridge could be rotated 180° by means of a knob or lever at the end of the tonearm to use the desired stylus.

Later, starting in the 1960s, most manufacturers settled on diamond-tipped styli for all cartridges. Magnetic cartridges lowered tracking forces to 1-2 grams, and with the obsolescence of 78's, diamond became the standard stylus material. Moving magnet cartridges often have replaceable styli; most moving coil cartridges do not offer user-replaceable styli, although some manufacturers offer a trade-in or a re-tipping service.

Magnetic cartridge manufacturers usually provide a specialized range of styli for DJ use. More rugged conical styli are required due to the frequent reversals of direction involved in scratching and back-cueing.

Stylus shape

The physical shape of the stylus has a bearing on its performance. The most obvious shape is the spherical stylus (also known as "conical"), where the tip of the stylus is ground to a hemisphere for playing monophonic recordings or for rugged use. However, this shape is unable to faithfully track all possible variations in a record groove. Better quality LP styli use an elliptical or "line contact" shape, arranged with its 0.02 mm (0.0007 inch) long axis across the record groove. The short axis may be from 0.005 to 0.01 mm (.0002 to .0004 inch) depending on the particular design. This shape followed the undulations of the groove better than spherical styli because they more closely resembled the triangular cutter used to create the groove.

In later years, bi-radial styli appeared where, in the perpendicular axis, the tip of the stylus was ground to a smaller radius than the main body. The result was a stylus that rode slightly lower in the groove and even more closely matched the shape of the triangular cutter than the elliptical design. There were several variations on the basic bi-radial design depending on the manufacturer who produced them.

Cantilevers

The cantilever is the arm that connects the stylus to the magnet or pickup coils. Most cartridges have cantilevers made from aluminium or boron; some very expensive models have ruby, diamond, beryllium or carbon fiber cantilevers chosen for their exceptional stiffness.

Chapter 16

Magnetic Core

A **magnetic core** is a piece of magnetic material with a high permeability used to confine and guide magnetic fields in electrical, electromechanical and magnetic devices such as electromagnets, transformers, electric motors, inductors and magnetic assemblies. It is made of ferromagnetic metal such as iron, or ferrimagnetic compounds such as ferrites. The high permeability, relative to the surrounding air, causes the magnetic field lines to be concentrated in the core material. The magnetic field is often created by a coil of wire around the core that carries a current. The presence of the core can increase the magnetic field of a coil by a factor of several thousand over what it would be without the core.

The use of a magnetic core can enormously concentrate the strength and increase the effect of magnetic fields produced by electric currents and permanent magnets. The properties of a device will depend crucially on the following factors:

- the geometry of the magnetic core.
- the amount of air gap in the magnetic circuit.
- the properties of the core material (especially permeability and hysteresis).
- the operating temperature of the core.
- whether the core is laminated to reduce eddy currents.

In many applications it is undesirable that the core itself retain magnetization and become magnetized by the external field. This property, called *hysteresis* can cause energy losses in applications such as transformers. Therefore 'soft' magnetic materials with low hysteresis, such as silicon steel, rather than the 'hard' magnetic materials used for magnets, are usually used in cores.

Commonly used structures

Air core

A coil not containing a magnetic core is called an air core coil. This includes coils wound on a plastic or ceramic form in addition to those made of stiff wire that are self-supporting and have air inside them. Air core coils have lower inductance than similarly sized ferromagnetic core coils, but are used in radio frequency circuits to prevent energy losses called core losses that occur in magnetic cores. The absence of core losses permits a higher Q factor, so air core coils are used in high frequency resonant circuits.

Straight cylindrical rod

On the left, a non-adjustable ferrite rod with connection wires glued to the ends. On the right, a molded ferrite rod with holes, with a single wire threaded through the holes.

Most commonly made of ferrite or a similar material, and used in radios especially for tuning an inductor. The rod sits in the middle of the coil and small adjustments of the rod's position will fine tune the inductance. Often the rod is threaded to allow adjustment with a screwdriver. In radio circuits, a blob of wax or resin is used once the inductor has been tuned to prevent the core from moving.

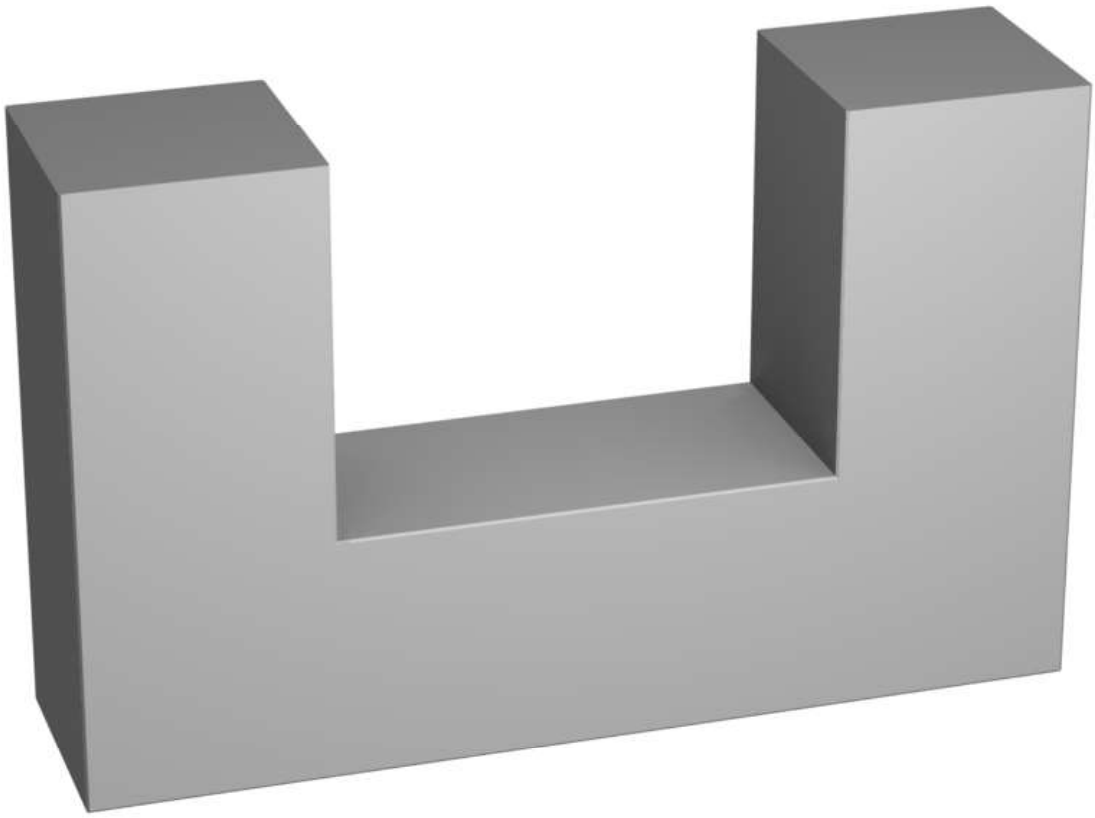
The presence of the high permeability core increases the inductance but the field must still spread into the air at the ends of the rod. The path through the air ensures that the inductor remains linear. In this type of inductor radiation occurs at the end of the rod and electromagnetic interference may be a problem in some circumstances.

Single "I" core

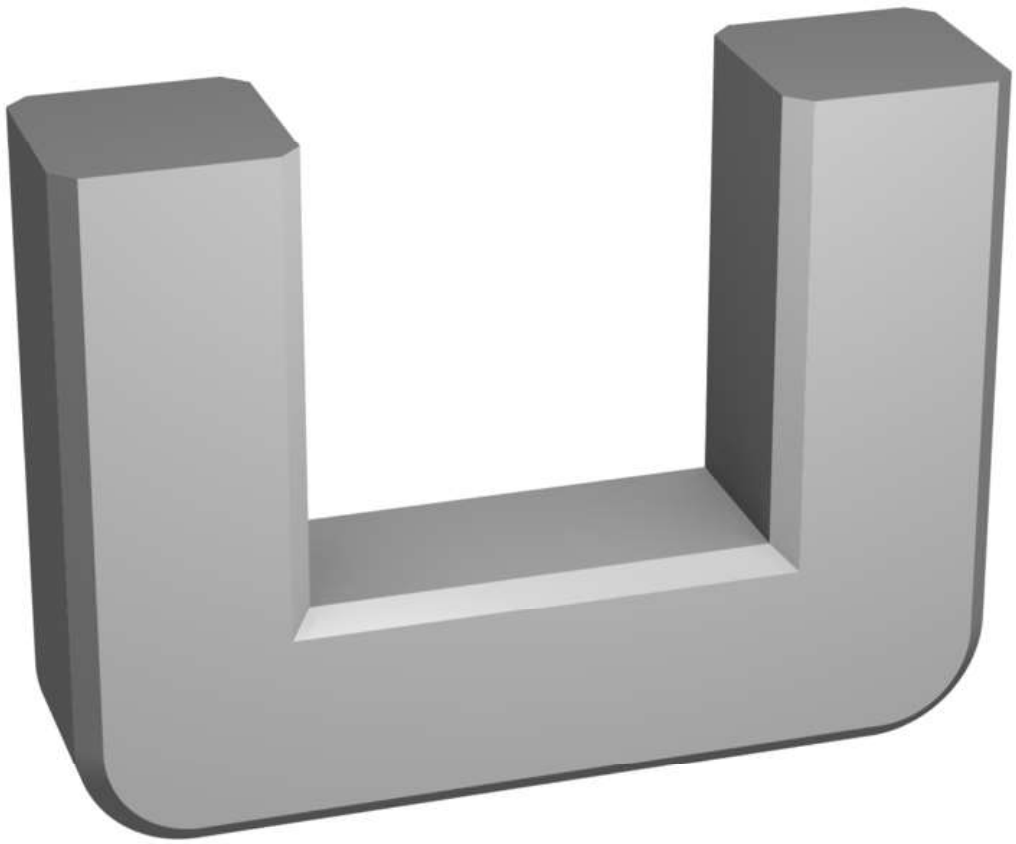
Like a cylindrical rod but square, rarely used on its own. This type of core is most likely to be found in car ignition coils.

"C" or "U" core

U and *C*-shaped cores are used with *I* or another *C* or *U' core to make a square closed core, the simplest closed core shape. Windings may be put on one or both legs of the core.*



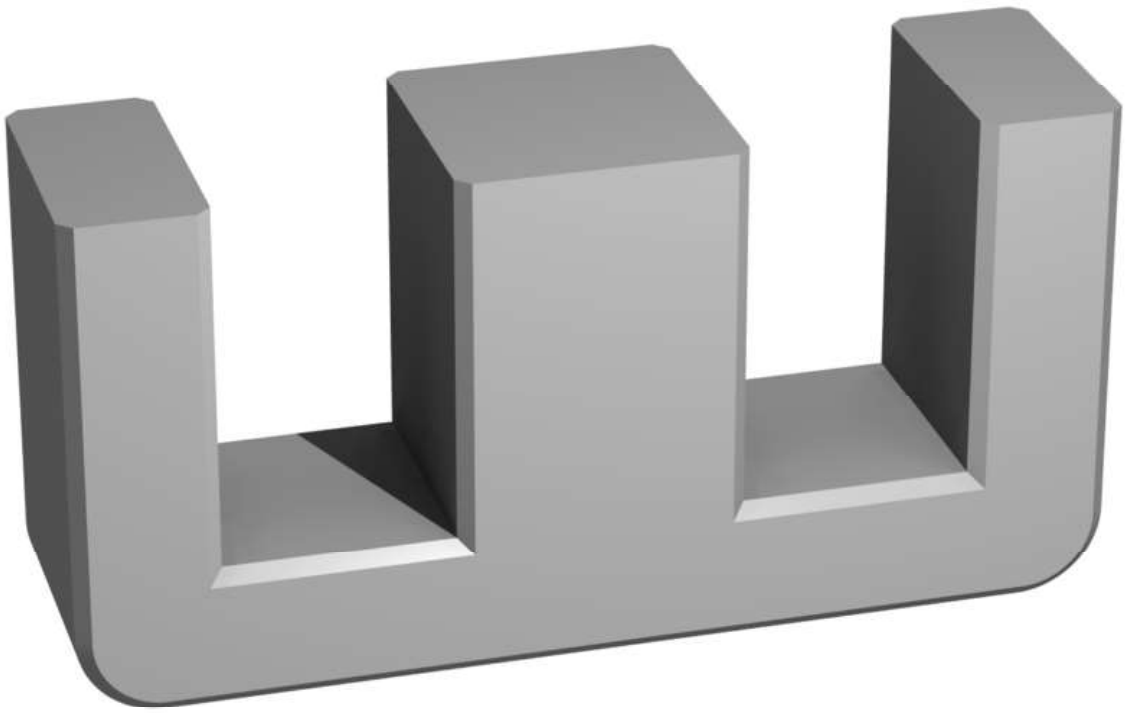
a U-shaped core, with sharp corners



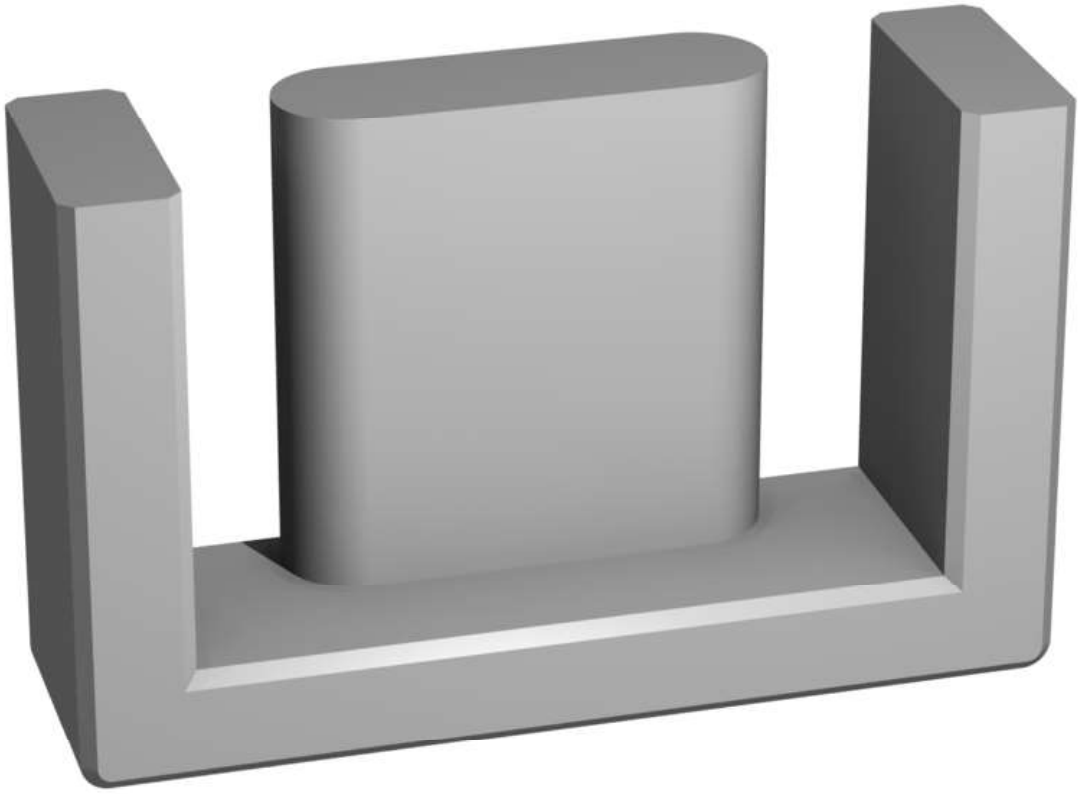
the C-shaped core, with rounded corners

"E" core

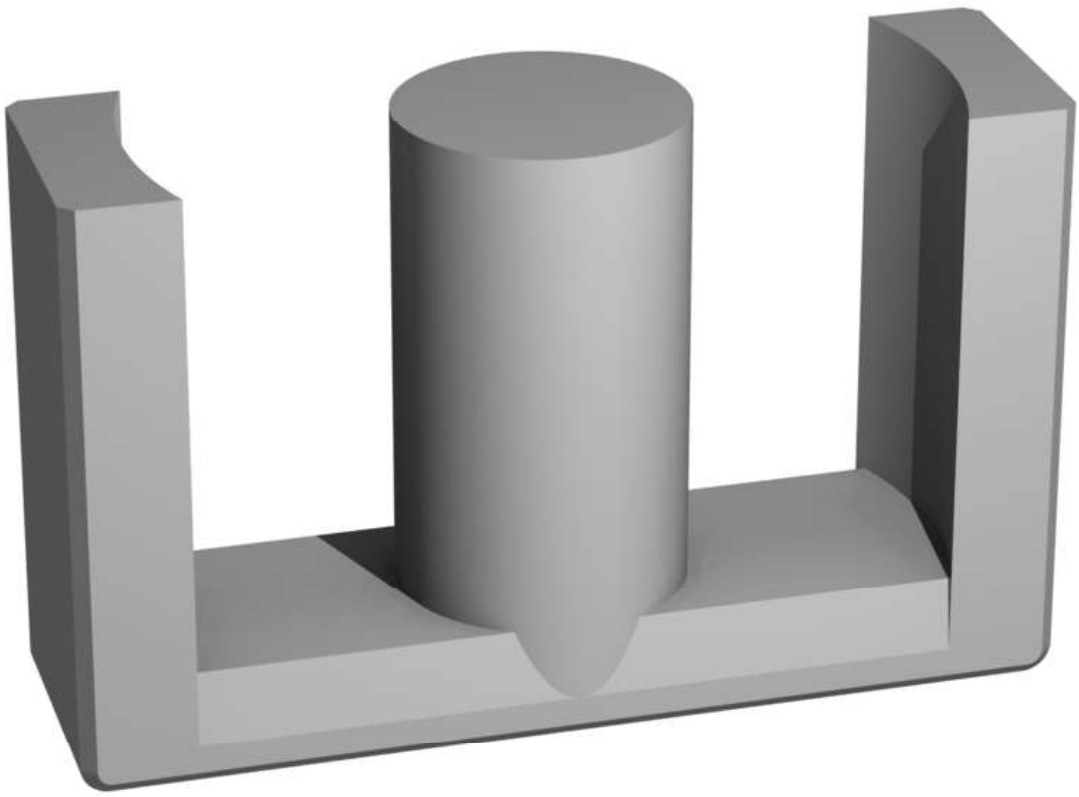
E-shaped core are more symmetric solutions to form a closed magnetic system. Most of the time, the electric circuit is wound around the center leg, whose section area is twice that of each individual outer leg.



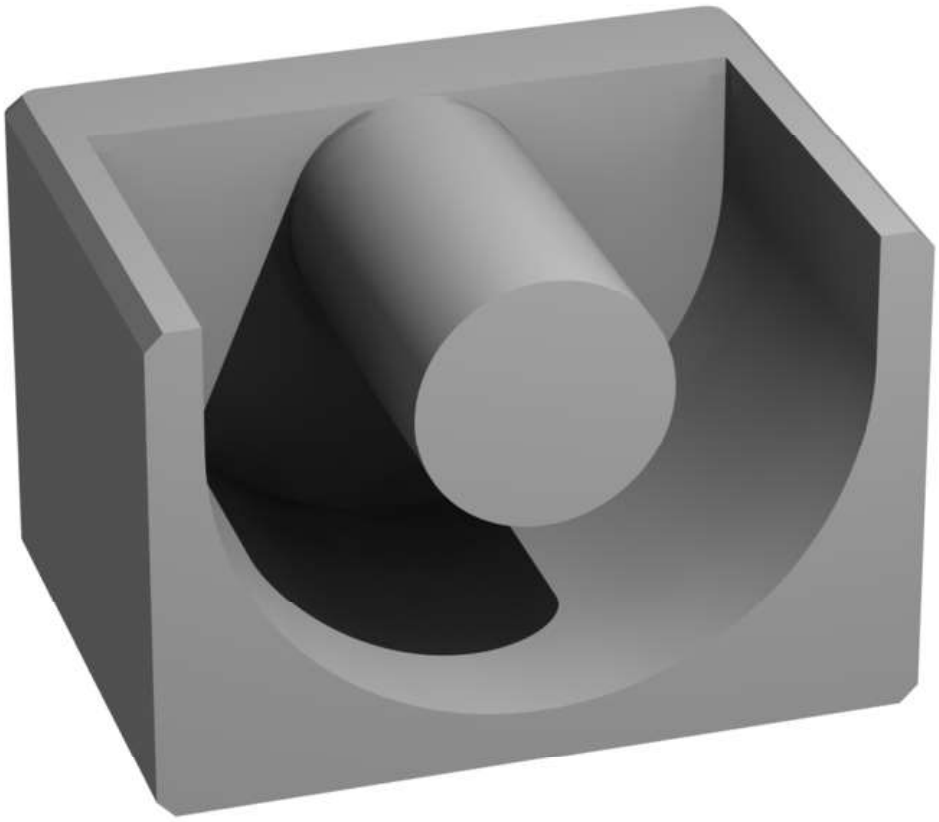
Classical *E* core



The *EFD'* core allows for construction of inductors or transformers with a lower profile



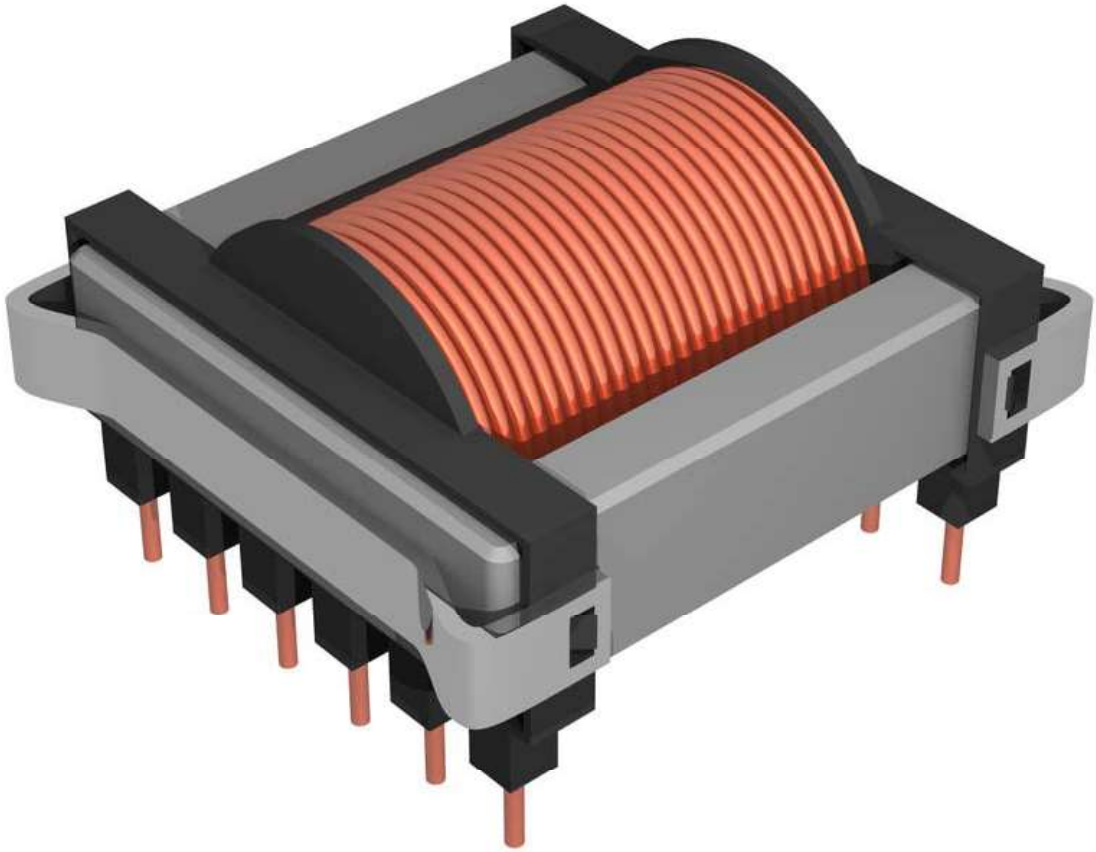
The *ER* core has a cylindrical central leg.



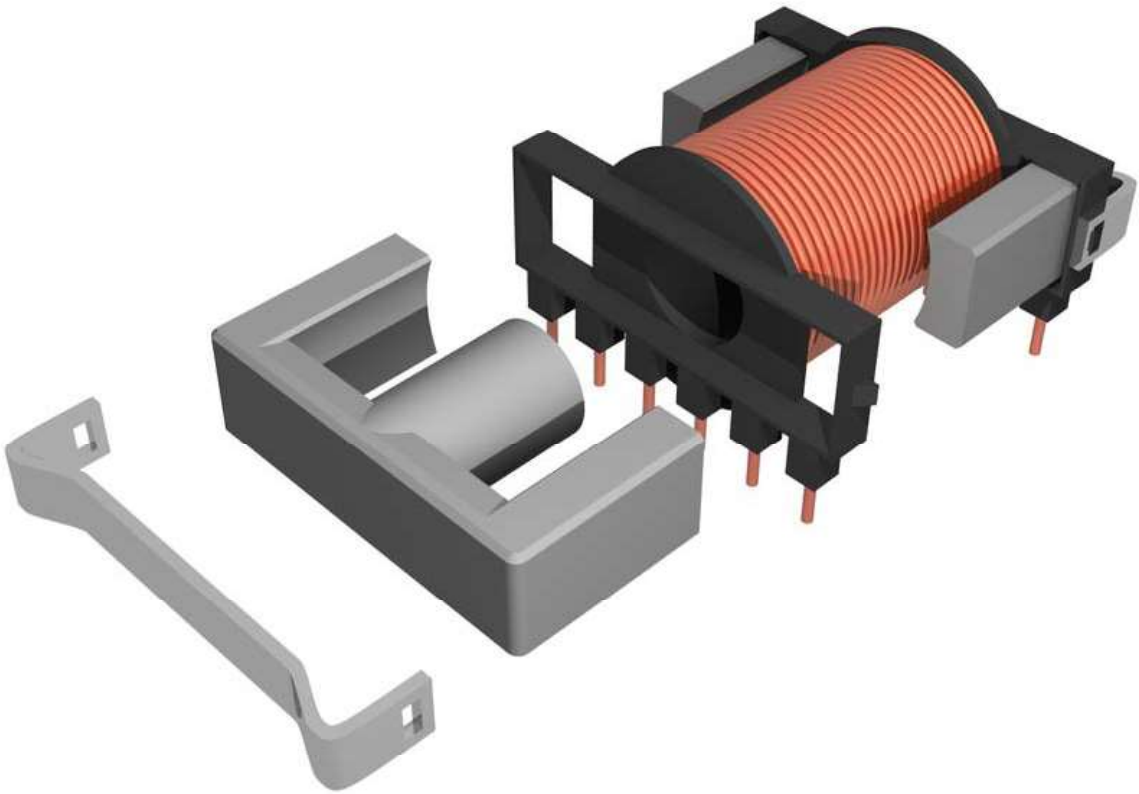
the *EP* core is halfways between a *E* and a *pot* core

"E" and "I" core

Sheets of suitable iron stamped out in shapes like the (sans-serif) letters "E" and "I", are stacked with the "I" against the open end of the "E" to form a 3-legged structure. Coils can be wound around any leg, but usually the center leg is used. This type of core is much used for power transformers, autotransformers, and inductors.



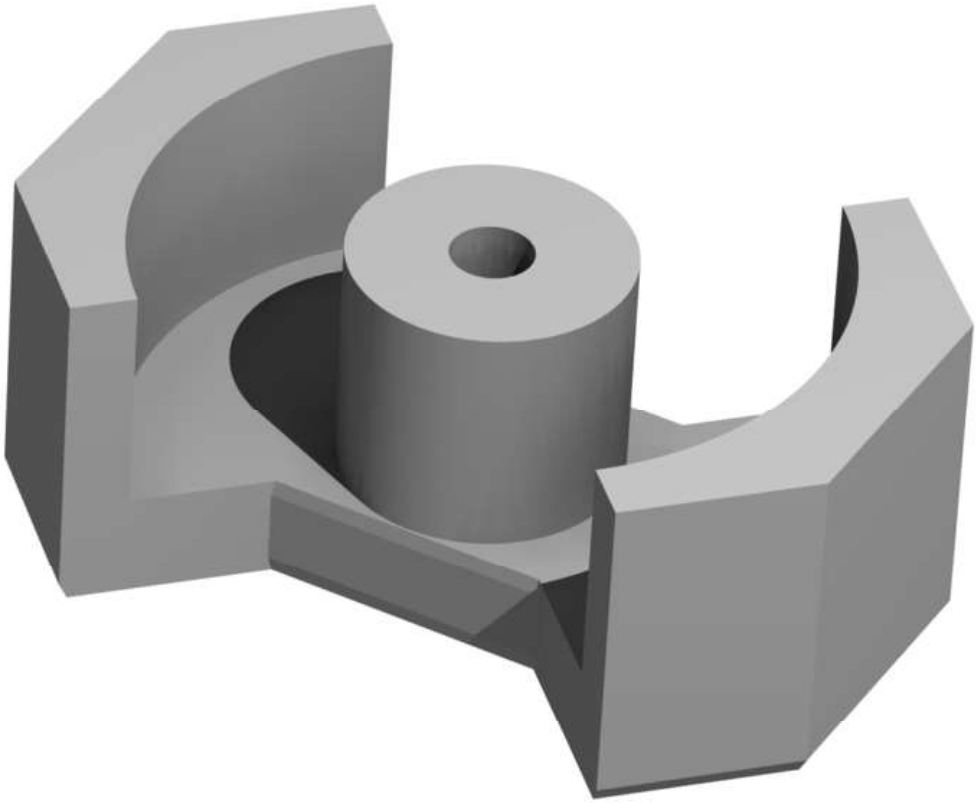
Construction of an inductor using two *ER* cores, a plastic bobbin and two clips. The bobbin has pins to be soldered to a printed circuit board.



Exploded view of the previous figure showing the structure

Pair of "E" cores

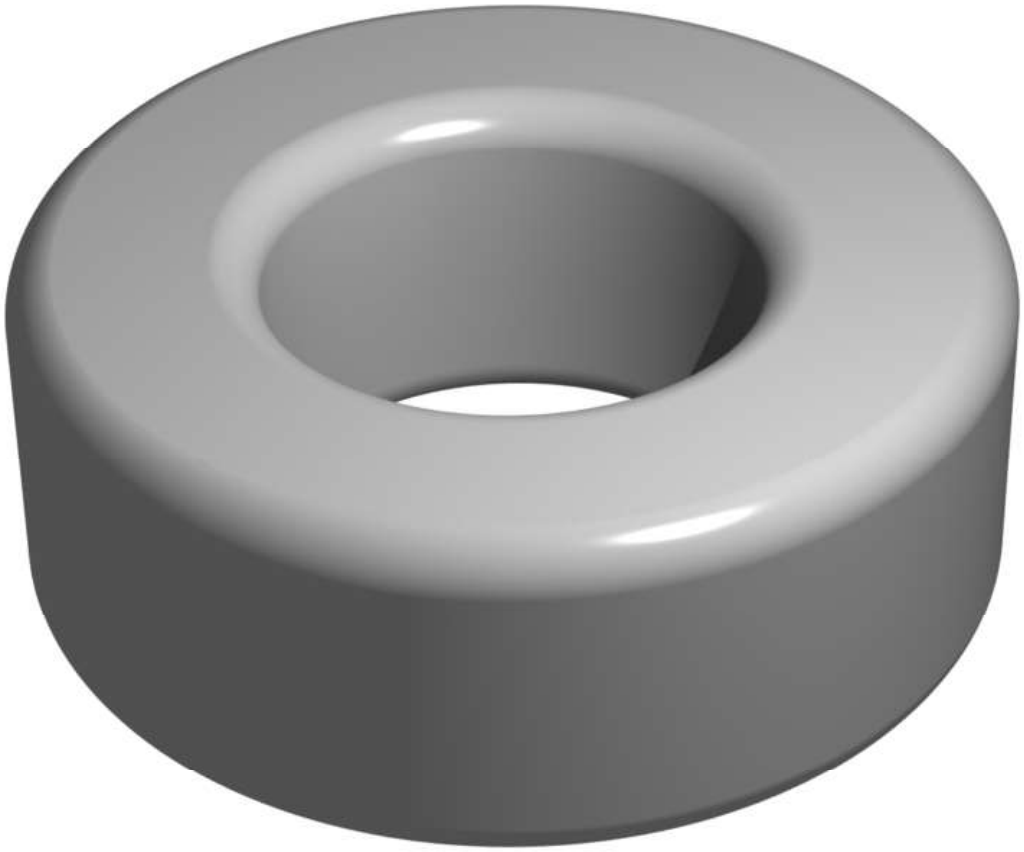
Again used for iron cores. Similar to using an "E" and "I" together, a pair of "E" cores will accommodate a larger coil former and can produce a larger inductor or transformer. If an air gap is required, the centre leg of the "E" is shortened so that the air gap sits in the middle of the coil to minimise fringing and reduce electromagnetic interference.



a pot core of 'RM' type

Pot core

Usually ferrite or similar. This is used for inductors and transformers. The shape of a pot core is round with an internal hollow that almost completely encloses the coil. Usually a pot core is made in two halves which fit together around a coil former (bobbin). This design of core has a shielding effect, preventing radiation and reducing electromagnetic interference.



A toroidal core

Toroidal core

This design is based on a toroid (the same shape as a doughnut). The coil is wound through the hole in the torus and around the outside. An ideal coil is distributed evenly all around the circumference of the torus. The symmetry of this geometry creates a magnetic field of circular loops inside the core, and the lack of sharp bends will constrain virtually all of the field to the core material. This not only makes a highly efficient transformer, but also reduces the electromagnetic interference radiated by the coil.

It is popular for applications where the desirable features are: high specific power per mass and volume, low mains hum, and minimal electromagnetic interference. One such application is the power supply for a hi-fi audio amplifier. The main drawback that limits their use for general purpose applications, is the inherent difficulty of winding wire through the center of a torus.

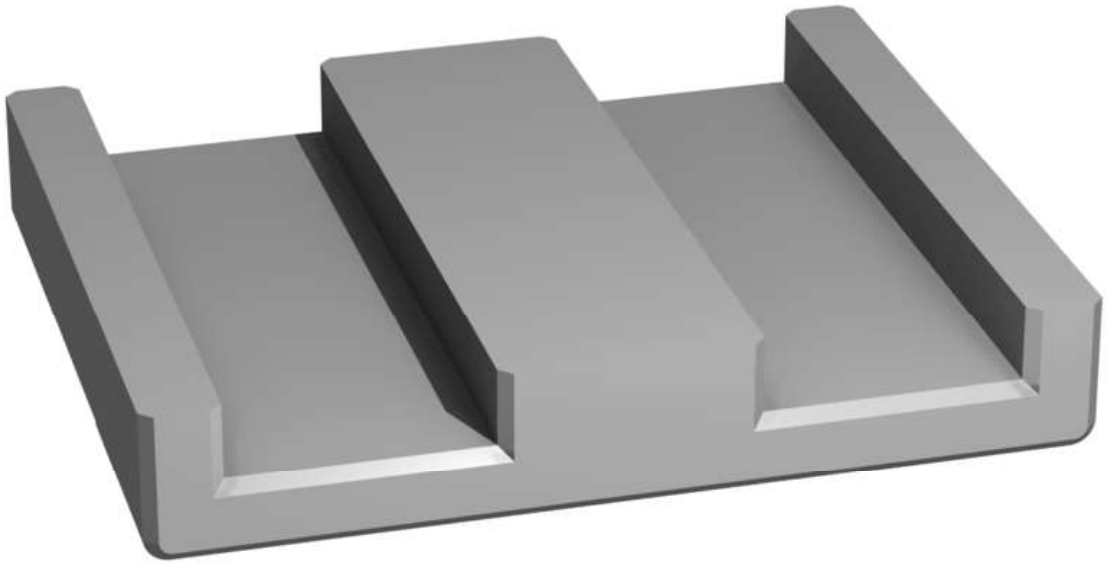
Unlike a split core (a core made of two elements, like a pair of *E* cores), specialized machinery is required for automated winding of a toroidal core. Toroids have less audible noise, such as mains hum, because the magnetic forces do not exert bending moment on the core. The core is only in compression or tension, and the circular shape is more stable mechanically.

Ring or bead

A ferrite ring on a computer data cable.

The ring is essentially identical in shape and performance to the toroid, except that inductors commonly pass only through the center of the core, without wrapping around the core multiple times.

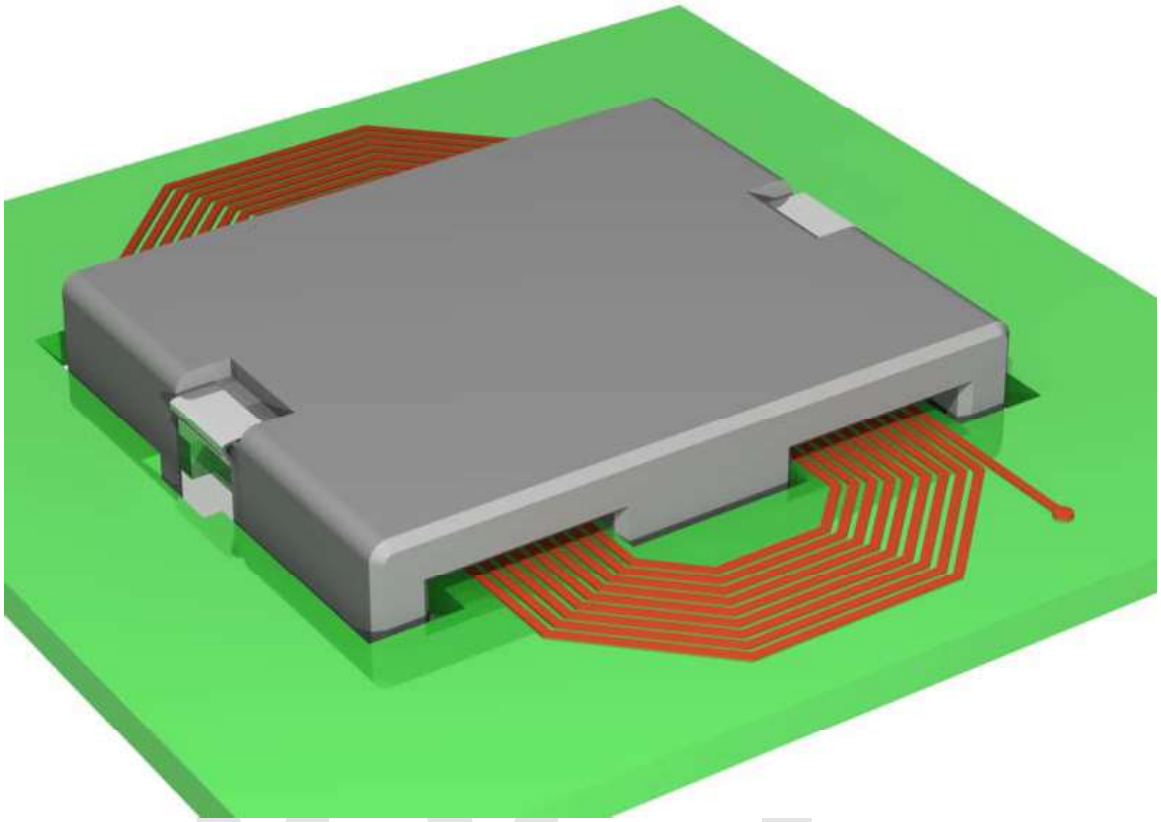
The ring core may also be composed of two separate C-shaped hemispheres secured together within a plastic shell, permitting it to be placed on finished cables with large connectors already installed, that would prevent threading the cable through the small inner diameter of a solid ring.



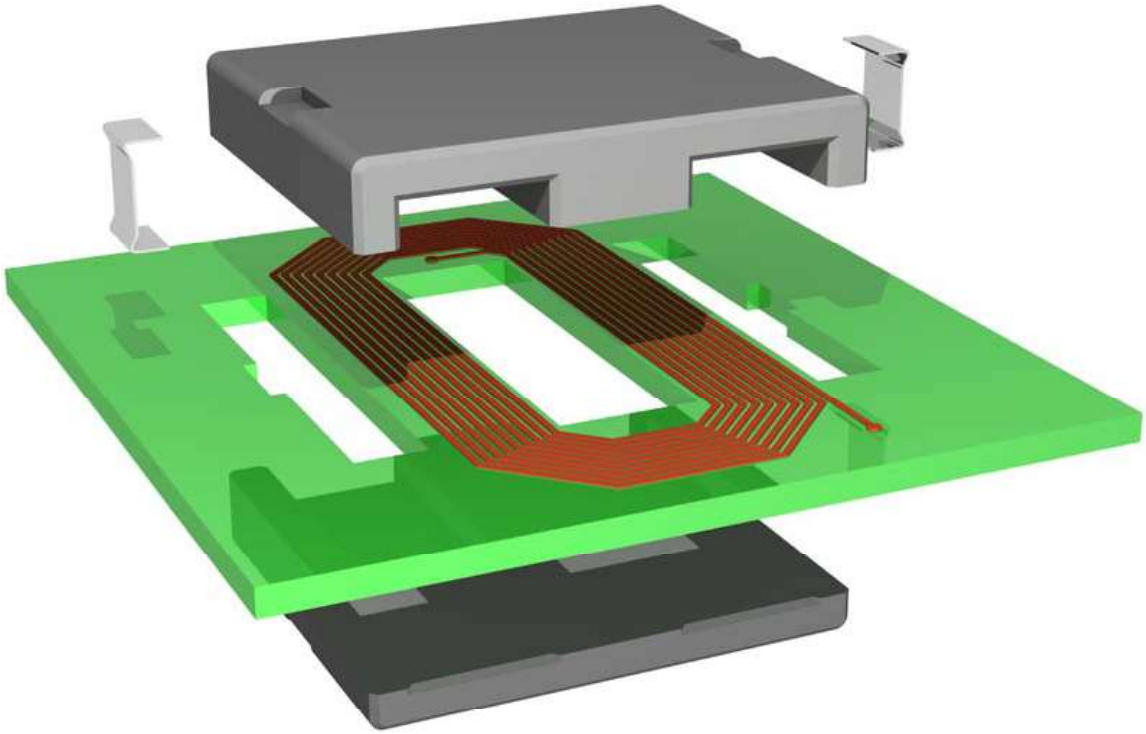
A planar 'E' core

Planar core

A planar core consists of two flat pieces of magnetic material, one above and one below the coil. It is typically used with a flat coil that is part of a printed circuit board. This design is excellent for mass production and allows a high power, small volume transformer to be constructed for low cost. It is not as ideal as either a **pot core** or **toroidal core** but costs less to produce.



A planar inductor



Exploded view that shows the spiral track made directly on the printed circuit board

Core loss

In a transformer or inductor, some of the power that would ideally be transferred through the device is lost in the core, resulting in heat and sometimes noise. There are various reasons for such losses, the primary ones being:

Hysteresis loss

When the magnetic field through the core changes, the magnetization of the core material changes by expansion and contraction of the tiny magnetic domains it is composed of, due to movement of the domain walls. This is a lossy process, because the domain walls get "snagged" on defects in the crystal structure and then "snap" past them, dissipating energy as heat. This is called hysteresis loss. It can be seen in the graph of the *B* field versus the *H* field for the material, which has the form of a closed loop. The amount of energy lost in the material in one cycle of the applied field is proportional to the area inside the hysteresis loop. Hysteresis loss increases with higher frequencies as more cycles are undergone per unit time.

Eddy current loss

The induction of eddy currents within the core causes a resistive loss. The higher the resistance of the core material the lower the loss. Lamination of the core material can reduce eddy current loss, and also making the core of a nonconductive magnetic material like ferrite.

Magnetic core materials

Having no magnetically active core material (an "air core") provides very low inductance in most situations, so a wide range of high-permeability materials are used to concentrate the field. Most high-permeability material are ferromagnetic or ferrimagnetic.

Soft iron

"Soft" iron is used in magnetic assemblies, electromagnets and in some electric motors; and it can create a concentrated field that is as much as 50,000 times more intense than an air core.

Iron is desirable to make magnetic cores, as it can withstand high levels of magnetic field without saturating (up to 2.16 teslas at ambient temperature.)

It is also used because, unlike "hard" iron, it does not remain magnetised when the field is removed, which is often important in applications where the magnetic field is required to be repeatedly switched.

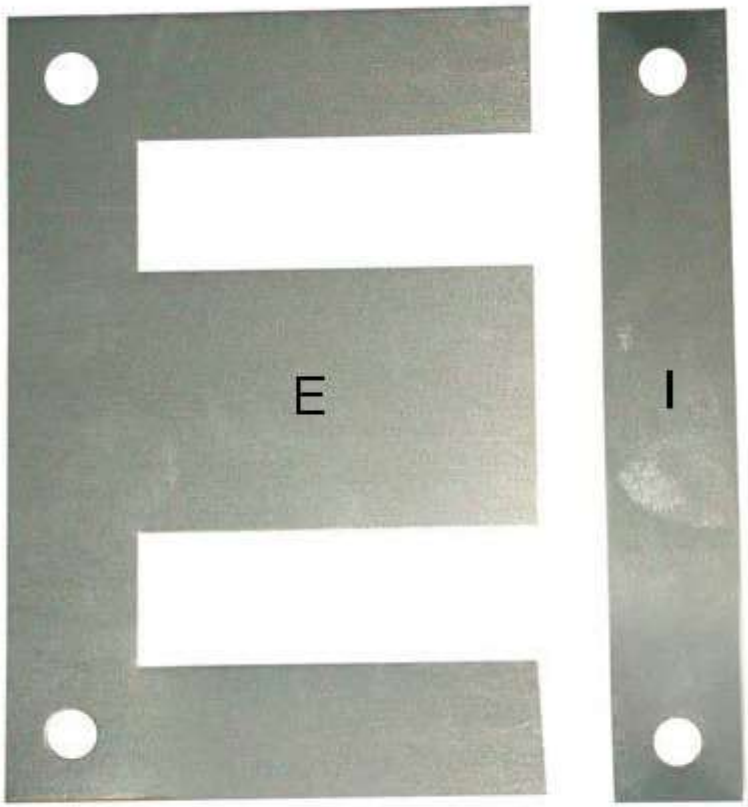
Unfortunately, due to the electrical conductivity of the metal, at AC frequencies a bulk block or rod of soft iron can often suffer from large eddy currents circulating within it that waste energy and cause undesirable heating of the iron.

Laminated silicon steel

Because iron is a relatively good conductor, it cannot be used in bulk form with a rapidly changing field, such as in a transformer, as intense eddy currents would appear due to the magnetic field, resulting in huge losses (this is used in induction heating).

Two techniques are commonly used together to increase the resistivity of iron: lamination and alloying of the iron with silicon.

Lamination



Typical EI Lamination.

Laminated magnetic cores are made of thin, insulated iron sheets. Using this technique, the magnetic core is equivalent to many individual magnetic circuits, each one receiving only a small fraction of the magnetic flux (because their section is a fraction of the whole core section). Furthermore, these circuits have a resistance that is higher than that of a non-laminated core, also because of their reduced section. From this, it can be seen that the thinner the laminations, the lower the eddy currents.

Silicon alloying

A small addition of silicon to iron (around 3%) results in a dramatic increase of the resistivity, up to four times higher. Further increase in silicon concentration impairs the steel's mechanical properties, causing difficulties for rolling due to brittleness.

Among the two types of silicon steel, grain-oriented (GO) and grain non-oriented (GNO), GO is most desirable for magnetic cores. It is anisotropic, offering better magnetic properties than

GNO in one direction. As the magnetic field in inductor and transformer cores is static (compared to that in electric motors), it is possible to use GO steel in the preferred orientation.

Carbonyl iron

Powdered cores made of carbonyl iron, a highly pure iron, have high stability of parameters across a wide range of temperatures and magnetic flux levels, with excellent Q factors between 50 kHz and 200 MHz. Carbonyl iron powders are basically constituted of micrometer-size spheres of iron coated in a thin layer of electrical insulation. This is equivalent to a microscopic laminated magnetic circuit, hence reducing the eddy currents, particularly at very high frequencies.

A popular application of carbonyl iron-based magnetic cores is in high-frequency and broadband inductors and transformers.

Iron powder

Powdered cores made of hydrogen reduced iron have higher permeability but lower Q. They are used mostly for electromagnetic interference filters and low-frequency chokes, mainly in switched-mode power supplies.

Ferrite

Ferrite ceramics are used for high-frequency applications. The ferrite materials can be engineered with a wide range of parameters.

Vitreous Metal

Amorphous metal is a variety of alloys that are non crystalline or glassy. These are being used to create high efficiency transformers. The materials can be highly responsive to magnetic fields for low hysteresis losses and they can also have lower conductivity to reduce eddy current losses. China is currently making wide spread industrial and power grid usage of these transformers for new installations.