Electronics

In order to manipulate and store sounds, we need a representation that can be easily transferred from point to point, altered to suit our wishes, and stored in a permanent fashion. A simple and flexible such representation is the flow of electrons, or electricity. We can use an electromechanical device, a microphone, to convert the air pressure variations of sound into an analogous flow of electrons. It is then simple to distribute the electrons through a wire to other places. It is also possible to convert the flow of electrons into magnetic flux that may then be stored for a long time, retained as magnetic patterns on media like analog tape and digital hard disks.

Devices including recorders, amplifiers, mixers, equalizers, dynamic range processors, and effects processors are able to reproduce, mix, and alter the electrical information in specific ways, giving an engineer the ability to manipulate artistically the reproduction of the sounds that were recorded. While these devices are complex, their operation is based on a just a few elementary concepts of analog electronics. Analog electronics refers to systems in which electric signal amplitude varies continuously in direct proportion to the intensity of the original sound vibrations that were converted to electricity by a transducing device. Until recently, this was the only method available for processing audio data. Digital electronics now allow discrete measurements of the transducer signal to be created, stored, and manipulated in binary form by computers at a very high rate of speed, producing a fast but non-continuous numerical representation of the original sound. We will be able to understand all these audio devices more thoroughly by first exploring the basic concepts of analog electronics.

Electricity is simply the movement of charge (symbol q). Charge is a fundamental property of matter: it has two possible states, positive and negative. Opposite charges can combine and cancel. Charge was theorized before the structure of the atom was understood, so it was assumed that positive charge flowed through electric circuits. Electric current is now known to be due to the flow of negatively charged electrons, but current is still sometimes thought of as positive charge flow in scientific and engineering literature. The flow of charge can be through a resistive medium like air (lightning), or through a solid conductor like a wire. When considering electronics, the path through which the charge moves is called a circuit; the charge moves from a source through a loop of circuit elements and back to the source. For direct current (DC), the charge flows in only one direction (although its magnitude may continuously vary), while for alternating current (AC) the charge flows back and forth in both directions.

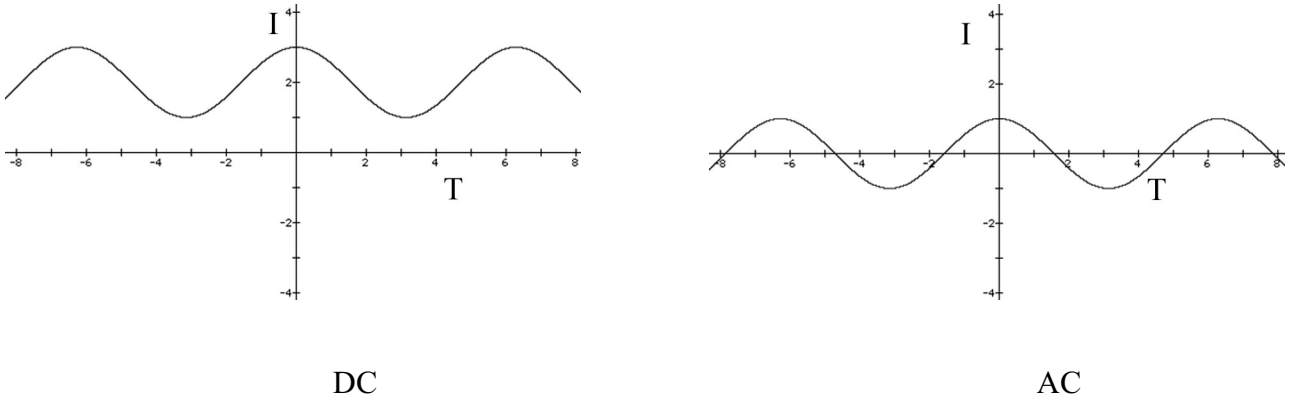


Fig. 1 Direct current (DC) and alternating current (AC)

Basic electrical principles include:

Charge (q): Charge is the carrier of electric current; it is generally carried by electrons of the atoms composing the material through which the charge flows or on which it is stored, although in solid state devices currents may be carried by missing electrons that effectively create positive charge currents. Charge can also be static, which occurs if an excess of one charge, positive or negative, accumulates on an object, attracting oppositely-charged dust and pet hair, for example. Charge is measured in coulombs, a quantity of electric charge equivalent to 6.414×10^{18} electrons, the amount of charge transferred by one ampere in one second.

Current (i): Current is the flow of charge. Current is defined as $I=dq/dt$, or the amount of charge flow (Δq) per unit time (Δt): it is the rate of charge transfer. Current is measured in amperes, with the units of coulombs/sec.

Voltage (v): Voltage is the force propelling charge through a circuit, like pressure in a fluid system. Voltage is defined as $v = -W_{\infty}/q$, where W_{∞} is the work expended to move a charge from infinity to the point of its measurement; volts have the units of joule/coulomb. Voltage is also called electromotive force (emf) as an indication of its nature as the driving force of electricity. Because of the amount of work potential a voltage possesses, it is also often referred to as potential. Voltages are capable of doing considerable work: they drive current through circuits that can fill an auditorium with sound, weld metals or cook food. Audio signals are most often represented as time-varying voltages.

Impedance ($\boldsymbol{Z}$): Impedance is an opposition to the flow of current (analogous in effect to the diameter of a pipe in a fluid flow system) and is measured in ohms for resistance (R) and reactance ($\boldsymbol{X}$), two types of impedance. Resistance does not depend on the frequency of the signal while reactance does. The equivalent resistance of a reactance is a function of frequency, however for a given frequency, the magnitude of a reactance can be considered much like a resistance in a circuit with regard to determining signal amplitude. At the same time, any reactive impedance also has an effect on the propagation time through the circuit and can alter the relative phase of a sine wave signal. This complicates the analysis of reactive circuits with time-varying signals.

Electric and magnetic fields ($\vec{E}$ and $\vec{B}$): Every charged particle creates an electric field ($\vec{E}$) that diminishes in magnitude as it radiates outward in all directions. Any two separated unlike charges, called a dipole, generate between them an electric field whose strength and direction depend on the spatial location of the measurement relative to the charges. This field is described mathematically as a vector field, a large collection of individual vectors each of which represents the electric force magnitude and direction at that particular point in space. The field exerts a force on any charged particle within it. Connecting lines of equal force radiate out in every direction from the charges and the longer the path, the lower the strength of the field along that path.

While we normally regard a wire simply as a conduit for electrical current flow, there is a very important phenomenon generated by current flowing in a wire: namely the creation of a magnetic field ($\vec{B}$) that varies with the changing flow of current. For a current flowing in a hypothetical infinite-length straight wire, the magnetic field magnitude B at a distance R is given by:

$$B = \frac{\mu_0 i}{2\pi R},$$

where μ_0 is the permeability constant (1.26 x 10^{-6} Tesla-meter/ampere) and i is current. As the current increases, so does the field strength and as the distance increases, the field strength decreases. Whenever current flows, it sets up a magnetic field, and any time a wire moves through a magnetic field, a current flow is produced in the wire. Known as induction, this phenomenon allows circuits to be coupled with no physical connection in a transformer, where two coils share an overlapping magnetic field. The inherent connection between electric current and magnetic fields is at the heart of two critical stages of sound recording: the conversion of sound into electric current in microphones (and the converse in loudspeakers) and the magnetic recording process used in both analog and digital recorders. It can also create problems by providing an avenue for unwanted coupling of signals in certain situations.

<u>Signal</u>: The term denotes a time-varying voltage or current that encodes information: a voltage or current that varies in proportion to a measured quantity, like air pressure, for example.

The simplest electronic components are the passive devices: resistors, capacitors, and inductors. Passive means they do not require external power to function, only the power contained in a signal itself. (Active components like transistors, vacuum tubes, and integrated circuits require external power as well as the energy of the signal.) Electronic devices are often characterized by their current-voltage (*i-v*) relationships: as we vary the current through the device, what does the voltage across them do? While keeping in mind that these circuit elements all exhibit slight deviations from the "ideal" behavior, these differences are small enough to ignore at first.

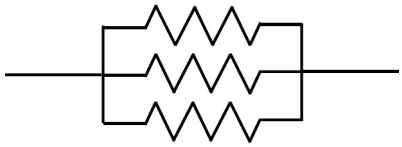
Resistors

$$v = iR$$

Resistors are passive devices that have constant impedance regardless of the frequency of the current flowing. They oppose the flow of current by an amount directly related to the resistance (like a constriction in a hose resists the flow of liquid). Resistances are the simplest of the passive devices and working with purely resistive circuits is a good place to start to build an understanding of electronic circuits. Resistances in series simply add arithmetically,

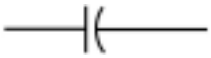
$$R_{total} = R_1 + R_2 + ... + R_n,$$

while parallel resistances add as :



$$\frac{1}{R_{total}} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}.$$

Using just resistors, circuits are only able to attenuate signals linearly, but the variable resistor, or potentiometer, is a purely resistive device at the center of most audio processes: it is the volume control or fader. The potentiometer ("pot") consists of a fixed resistance from end-to-end with a movable wiper than can connect anywhere between the ends of the resistive element (see Fig. 2 below).



Capacitors

$$v = \int i dt / C, \; (i = C\, dv/dt), \; X_c = 1/2\pi f C$$

A capacitor consists of two electrically charged conductors, usually flat conductive surfaces called plates, placed close together with a dielectric (non-conducting) material between them. When a charge is applied to one plate, it repels like charges on the opposite plate, leaving a net opposite charge. The voltage that builds across a capacitor is related to the charge that resides on the capacitor's plates by the equation $q=Cv$, so the voltage is equal to the charge divided by the capacitance ($v=q/C$). Since current is the rate of charge flow, the larger the current, the faster the voltage builds. For direct current, the capacitor charges up with a time constant that depends on the capacitance value and the impedance through which the current flows into the capacitor, since that impedance will determine the maximum amount of current flowing into the capacitor from a given voltage input. Once the capacitor is fully charged, no further current flows. This means that the capacitor is an effective block for direct current. For alternating current (like audio signals), the response is more complicated: the voltage that develops across the capacitor depends on how quickly the current is changing in magnitude and direction. Since it takes time for the charge to build up, the result is a frequency dependent delay (or phase shift) in the output signal since the driving voltage, and therefore the charging current, is changing while the charge accumulates

The unit of capacitance is the farad (F). One farad equals one coulomb/volt. Until the advent of monster car stereo systems, capacitors of this size were unusual: in electronic circuits values of micro-farads ($\mu F=10^{-6}$ F), nanofarads ($nF=10^{-9}F$), and picofarads ($pF=10^{-12}F$) are most common. A charged capacitor acts like a battery, though it can only generate a current until it is fully discharged.

Inductors

$$v = L\, di/dt, \; X_L = 2\pi f L$$

An inductor is most often a simple coil of wire, which can be wrapped around either an air or metal core. As current flows into an inductor, a magnetic field is created around the coil. When the current stops, the magnetic field collapses, generating an induced current flow in the coil in the opposite direction to the original current. Low-frequency currents flow easily through the inductor, but as the alternating current frequency increases, the impedance of the inductor increases. Like the capacitor, the inductor introduces a phase shift. An inductor is conceptually similar to a water tank: it "stores" a magnetic field rather than water, with the magnetic field storing energy that can be fed back into a circuit. Resonant circuits with inductors are sometimes referred to as "tank" circuits.

Transformers are special types of inductors, where two separate coils share overlapping magnetic fields. When the primary coil is driven, it generates a magnetic field that induces current to flow in the secondary coil. Since there is no physical connection between the primary and secondary coil wires, the two circuits are physically isolated from each other. Often, transformers consist of an iron core wrapped with two or more coils, which couple through the magnetically susceptible metal. Transformers are used to get voltage gain (at the expense of current reduction) and to step down power line voltages for power supplies. Transformers are also used to match impedances between devices and to provide ground isolation.

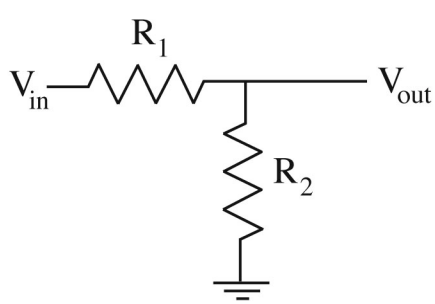
<u>Ohm's Law</u>: Ohm's law is one of the simplest, yet most important principles of electronics. It states that:

$$v = i \times R$$

(Strictly speaking, Ohm's law says that R is independent of v and therefore i versus v must be linear.) The voltage drop across a resistor is the current multiplied by the resistance. This holds true for the impedance of inductors and capacitors as well, if we take into account their frequency-dependent nature. The amount of work done per unit time in a circuit is given by:

$$P = i \times v = v^2/R = i^2 \times R, \quad \text{where Ohm's law has been substituted for either } i \text{ or } v.$$

Although most electronic devices are full of active components like op-amps and transistors, much of the actual circuit function is accomplished by simple arrangements of the passive elements: resistors, capacitors, and inductors. Understanding these simple circuits will allow an audio device user to examine the schematic diagram and quickly gain knowledge about the function of the device. It also makes troubleshooting possible.



$$v_{out} = v_{in}(R_2/(R_1 + R_2))$$

Potentiometer (variable resistor)

Fig. 2 Voltage divider and variable resistor

The simplest functional circuit, but one of the most important, is the voltage divider (Figure 2.) The voltage divider is so-called because the input voltage divides in proportion to the resistances of the circuit. If we measure the voltage drop across R_2, it is the input voltage times the ratio of R_2 to the total resistance of the circuit (ignoring the effects of other devices connected to the circuit.) The variable resistor is in fact an adjustable voltage divider that lets us control levels by continuously varying the ratio between R_1 and R_2.

The voltage divider helps explain a general principle relating to the proper interconnection of devices: impedance matching. When we connect two audio devices, we wish for the signal voltage at the output of the first device to be received at the input of the second device with as large amplitude as possible. In order to guarantee this, the input impedance of the second device must be greater than the output impedance of the first device. These impedances can be modeled by the two resistors above with R_1 representing the output impedance and R_2 representing the input impedance. We can see that when R_2 is much larger than R_1, most of the voltage will drop across R_2 as we desire: in order to efficiently transfer voltage, the input (load) impedance must be large relative to the output (source) impedance. The table below shows relative signal transfer, using a 1 volt input, as a function of the ratio of the two resistances, where R_S (source or output resistance) is equivalent to R_1 above and R_L (load or input resistance) corresponds to R_2:

R_S (ohms)	R_L (ohms)	V_O (volts)
100	1	0.0099
100	10	0.091
100	100	0.5
100	1,000	0.91
100	10,000	0.99

The voltage divider circuit, when composed of frequency-selective passive elements (capacitors or inductors) will act as a filter, a circuit that allows some frequencies to pass while others are attenuated. Filters in use today rely mainly on resistors and capacitors. Inductors are less frequently used because they are physically large in an era of tiny surface-mount electronic elements, susceptible to electromagnetic interference, and aren't necessary in many applications. Inductors do impart a characteristic sound that is sometimes desired and are again becoming popular as a way of intentionally altering the sound in many audio devices. The frequency above (or below) which attenuation occurs depends on the value of resistance and capacitance (or inductance).

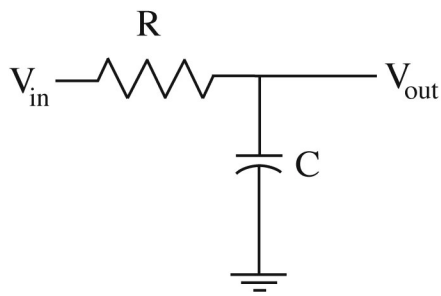


Fig. 3 Low-pass filter

Figure 3 shows a low-pass filter: low frequency signals pass un-attenuated. As the signal frequency increases, the capacitive reactance decreases. At the frequency at which the capacitive reactance just equals the resistance, the output signal is reduced by $1/\sqrt{2}$ (-3dB). This is known as the corner frequency of the filter and is determined by:

$$f_0 = 1/(2\pi RC)$$

[It might appear that the amplitude of the output signal at the corner frequency should be half the input and therefore the output should be down by 1/2, or -6 dB, but there is another consideration: the capacitor has an effect on the phase of the signal as well as its amplitude. As the frequency of the signal increases, the time response of the capacitor begins to shift the phase of a sine wave signal as it flows through the capacitor. We must use a vector description of the circuit, one that involves imaginary numbers to deal simultaneously with the amplitude and phase of the signal. When the vector description of the impedances is used, the magnitude part of the vector sum is:

$$\mathbf{Z}_{total} = \sqrt{(R^2 + \mathbf{X}_c^2)}$$

So, substituting in the voltage divider equation, we get:

$$v_{out} = \left(\frac{R}{\sqrt{R^2 + \mathbf{X}_C^2}}\right) v_{in}$$

So at the frequency where $X_c = R$,

$$\frac{v_{out}}{v_{in}} = \frac{R}{\sqrt{2R^2}} = \frac{1}{\sqrt{2}} \cdot]$$

A similar analysis can be applied to the high-pass filter, shown below in Figure 4.

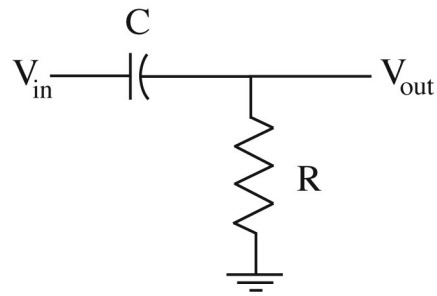


Fig. 4 High-pass filter

In this case, the capacitor effectively blocks DC and low frequency components, causing most of the input voltage to drop across the high capacitive reactance. As the frequency of the signal increases, the impedance of the capacitor decreases and more of the input voltage appears across the resistor.

Frequently, an intuitive understanding of the operation of circuit elements is as helpful as a complete engineering analysis. Most circuits can be understood on a superficial basis, since one is not trying to design a circuit, but simple appreciate what a circuit is doing to the signal in a general way, for example, boosting high frequencies, mixing signals, and buffering or impedance matching. Often, a circuit can be well understood just using Ohm's law and the R-C circuit theory introduced above. As users of audio devices, we are not concerned as much with engineering precision as we are with understanding general circuit function.

Conceptually, capacitors have high impedance for low frequencies and inductors have high impedance for high frequencies. Exactly what constitutes high and low frequency depends on the relative values of the circuit elements. For a more precise quantitative understanding of passive device circuits and sinusoidal signals, the use of phasors and complex number descriptions of impedances are in order.

As mentioned above, the impedances of capacitors and inductors may be treated like resistances, but with an added dependence on the frequency of the signal not present in resistors. Since reactance is a function of frequency, the exact equivalent resistance changes with frequency but for a given frequency has a fixed value. We may treat all impedances similarly if we have a way of dealing with the frequency-dependent nature of reactance, which we represent using complex numbers (in bold type). For example, the impedance of a capacitor is given by:

$$\mathbf{Z}_C = 1/j\omega C, \text{ where } j = \sqrt{-1} \text{ and } \omega = 2\pi f.$$

The impedance of an inductor is:

$$\mathbf{Z}_L = j\omega L.$$

If we use these expressions for the impedances of reactive circuit elements, circuit analysis is simplified.

Simple series and parallel combinations of impedances are often encountered in useful circuits like equalizers. By selecting the arrangement and values of these components, we can make frequency

selective circuits to modify the frequency composition of a signal by attenuating some frequencies more than others. (Note that using passive components alone, we can only attenuate and not amplify: amplification requires active circuit elements that often follow these passive circuits to provide gain.) The high-pass and low-pass circuits are the simplest of these circuits: using more impedance elements in clever geometries, we can make filters that boost or cut all frequencies above or below a corner frequency (shelving filters), boost or cut only the frequencies within a certain range (peaking filter), and cut very sharp, narrow bands of frequencies (notch filter). With the concepts of how impedances combine in series and parallel, we can begin to analyze and understand these more complicated circuits.

In series, impedances simply add linearly. In parallel, they combine as:

$$\frac{1}{\mathbf{Z}_{total}} = \frac{1}{\mathbf{Z}_1} + \frac{1}{\mathbf{Z}_2} + \ldots + \frac{1}{\mathbf{Z}_n},$$

For two parallel impedances:

$$\mathbf{Z}_{total} = \frac{\mathbf{Z}_1 \times \mathbf{Z}_2}{(\mathbf{Z}_1 + \mathbf{Z}_2)}.$$

By using the complex number representation of the reactive impedances, we can calculate the behavior of useful circuits more easily. But even without employing the mathematical analysis, we can grasp the function of a circuit by considering the relative contribution of each element's impedance as it relates to the whole circuit. Consider the circuit below, a useful low-frequency tone controlling circuit:

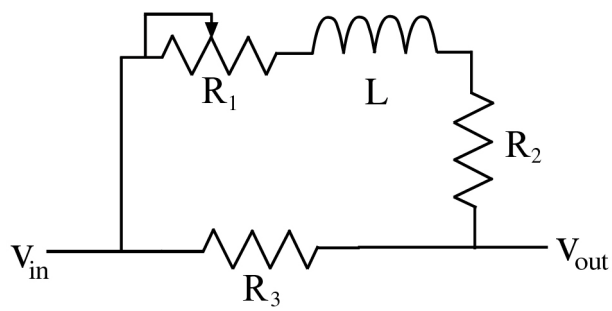


Fig. 5 Low boost shelf filter

The elements R_1, L, and R_2 provide a parallel path to R_3. When the parallel pathway has low impedance relative to R_3, it makes the total impedance low and the signal is passed through with little attenuation. Since for high frequencies X_L is large relative to R_3, the circuit will behave as if R_3 is the only element in the circuit at high frequencies. At low frequencies' X_L is low enough to ignore and the total impedance of the circuit will be determined by R_1 plus R_2 in parallel with R_3. Since R_1 is adjustable, it will determine how much the parallel path shunts (bypasses) R_3 and therefore determines the overall impedance with R_2 setting the minimum impedance. For low frequencies, the total impedance is

adjusted by R_1 and for high frequencies it is determined by R_3. This gives us a useful filter: a low shelf where every frequency below a corner frequency is increased an adjustable amount relative to the higher frequencies. It should be remembered, however, that no gain is available in such a circuit and we are actually attenuating the "boosted" frequencies less. Without considering the values of the components we cannot tell exactly where the line between low and high frequencies will occur, but we can determine the general function of the circuit by inspection, keeping in mind the basic operation of the passive electronic elements.

As we have seen, passive electronic components can be used to alter the frequency content of signals but cannot amplify, or increase, the signal level. For this we must rely on active devices, ones that can convert large applied DC voltages into amplified versions of smaller input voltages. These devices fall mainly into two categories: vacuum tubes and solid-state devices based on silicon or similar semi-conducting materials. Each is able to use a small voltage or current to control a larger voltage or current, thereby producing amplification or gain. The physical processes involved in these alternative systems are different but each has advantages and disadvantages that allow us to choose one or the other for a particular task based on these relative strengths and weaknesses.

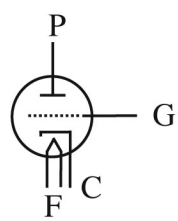


Fig. 6 Triode

Vacuum or thermionic tubes were the earliest devices allowing amplification. They are somewhat like a modified light bulb: an evacuated glass envelope with several metal electrodes. The cathode electrode is heated by electric current, through a filament (F), until it is hot enough to emit electrons. By placing a large positive voltage between the cathode (C) and a second electrode, the anode or plate (P), a current can be made to flow as the free electrons are attracted to the plate. If a third electrode, the control grid (G), is placed between the cathode and the plate and made slightly negative relative to the cathode, a small voltage can be made to control the current from cathode to plate, the plate current. When the plate current is made to flow through a fixed resistance, a larger voltage proportional to the grid voltage can be produced. This three-electrode tube is called a triode. Further refinement of the triode created the tetrode and pentode, with additional electrodes added to improve some aspects of the system. The tetrode adds a screen grid between the grid and plate that reduces the capacitance of the tube and makes the plate current less dependent on the plate voltage. The pentode also adds a suppressor grid connected to the cathode between the screen grid and plate to address the secondary emission of electrons that occurs as the plate current electrons strike the plate and dislodge stray electrons.

Vacuum tubes share the advantages and disadvantages of high-impedance devices: they have inherently high input impedances and can produce large linear amplification, but they are plagued by thermal noise and capacitance issues and are vulnerable to mechanical vibration pickup ("microphonics") and air leakage. They also require external power supplied at over 300 VDC and must be heated to incandescence. Tubes can produce clean amplification if they are used in proper circuits and with careful shielding. When overdriven, they distort by current limiting rather than by voltage limiting at

the output, producing harmonics but without the relative harshness of abrupt voltage-limited clipping. Many tube amplifier circuit designs employ transformers to match the desired input and output impedances to the high voltages and impedances of the tube circuits and these components also contribute to the classic sound of favored tube amplifiers.

A newer amplification device has revolutionized electronics and spawned the age of computers: the transistor. Unlike vacuum tubes, transistors can be ultra-miniaturized, require little applied power, and are practically made of sand. These solid-state devices are based on the behavior of materials known as semiconductors. Metals are easily able to conduct electricity: they do so by releasing loosely bound outer electrons freely, allowing current flow as the electrons move in response to external electric fields. Insulating materials have no such electrons available and hence do not allow current to flow. Materials such as silicon and germanium, possessing a limited number of free outer electrons and exhibiting an intermediate ability to conduct current, are known as semiconductors. Solid-state devices all rely on the same basic structure: junctions of semiconducting layers, each containing specific chemical impurities (called dopants) that confer either a net surplus or deficit of electrons. The areas with surplus electrons are designated n (negative) and the layers lacking electrons p (positive). Places where electrons are missing in p-type material are considered to behave as mobile positive charges and are called holes. Both electrons and holes diffuse freely through the semiconductor material and their movement is influenced by electric fields as well as by their concentration gradients. Solid-state devices are constructed by joining p and n regions of semiconducting materials in various combinations, leading to different behaviors that result from the physical construction and doping materials that are added to form the different devices. Joining one n and one p layer creates a diode junction, which passes current easily in one direction but not in the other. Stacking three layers in alternating fashion creates a bipolar junction transistor.

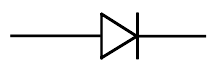
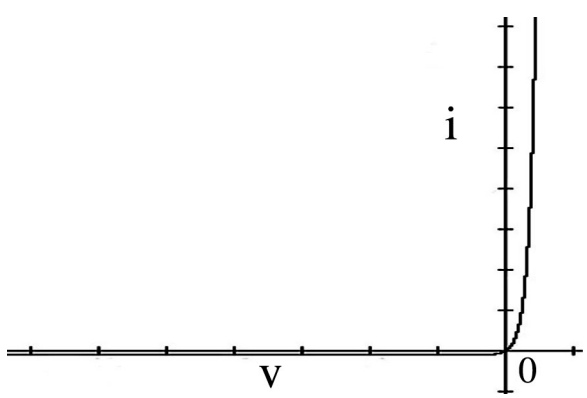


Fig. 7 Diode

A diode is constructed from two adjoining areas of doped semiconductor, one p-type and one n-type. When a voltage is applied that makes the p area positive relative to the n region (forward bias), current flows and the diode conducts. When the voltage is reversed (reverse bias), the surplus electrons and holes are pulled away from the junction and little current flows. The I-V description of the ideal diode is:



$$i = I_S(e^{(qv/kT)} - 1)$$

Fig. 8 Diode I-V curve

where I_S is the saturation current, which is determined by the construction of the diode, q is the electric charge, v is the voltage (positive on the p side), k is Boltzmann's constant, and T is temperature. This relationship results in very small current flow when v is negative, essentially I_S, which is in the nanoamp range. As the voltage becomes positive, the junction begins conduction and at around 0.5 V, current increases dramatically for very small increments in the voltage. This non-linear I-V behavior of a p-n junction differs from the linear behavior of resistors, whose impedance (I-V plot slope) is independent of the applied voltage, and therefore the diode does not obey Ohm's Law.

While the theoretical description of p-n junction behavior does predict real-world behavior, it is an incomplete description of real diode in-circuit performance, especially with regard to complex signals like music. Fortunately, a simplified conceptual understanding of diodes will be sufficient to understand their uses in audio equipment. One significant deviation from ideal occurs with large reverse-bias voltages, which cause the diode to begin conducting even though the junction is reverse-biased. This is known as the breakdown voltage and its value is determined by the physical construction of the diode. While this can lead to thermal damage, it can also be used to generate a voltage reference as long as the power dissipating capability of the diode is not exceeded. Special diodes known as Zener diodes are used in this mode to provide voltages that don't change as the current through them changes, generating stable voltage references in power supplies, for example. A common use for diodes in audio circuits is to rectify AC voltages in order to generate DC control voltages derived from signal amplitude envelopes, for use in compressors and expanders as well as in visual amplitude displays. The light-emitting diode (LED) is a special variety of diode that emits photons of light in response to current flow and is of no small importance in the metering of audio signals.

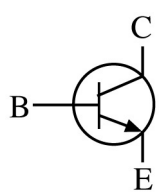


Fig. 9 Bipolar NPN transistor

By constructing two diodes back to back sharing a central electrode, a transistor is created. The behavior of this device depends on the mobility of the free charges that can diffuse through the semiconductor and respond to the influence of electric fields that are created at the junctions between the p and n areas. Since the free electrons and holes are constantly in motion powered by thermal energy, they diffuse randomly through the silicon like ions in solution even with no externally applied voltage. At the p-n junction, the electrons and holes are able to cross the junction, following their concentration gradients. The result is an equilibrium distribution of electrons and holes counterbalanced by the resulting electric field generated at the p-n junction by the separation of charges. This electric field is only present in the immediate vicinity of the p-n junction and does not affect the bulk of the semiconductor, however charges that happen to drift close to the junction are forced across it by the field. Since electron and hole pairs are continually combining, neutralizing their separated charges, as well as being generated anew by thermal energy, there is significant random movement of charge in the materials even without external voltages, a potential source of noise in active circuits.

In the most common type, the bipolar junction transistor, the emitter is doped more heavily than is the base or the collector. In NPN transistors, this excess doping provides extra electrons that diffuse toward the emitter-base p-n junction. When a positive voltage is applied between the base and emitter, the junction is forward biased and electrons flow into the base. Electrons enter and diffuse around the base as minority carriers, since holes are the majority carrier in the p-type base. The base is made small so that all of the electrons injected into the base by the emitter don't just combine with holes. When the collector is made positive relative to the base, the junction is reverse-biased, but electrons may still be swept across the base-collector junction by the electric field. Most of the emitter current flows out the collector lead, with only 1% or so of this current flowing in the base lead. If the base current is controlled externally (as by a signal), the collector and emitter currents will be amplified versions of the base current and we have a method of increasing signal levels. PNP transistors work similarly but with reversed polarity: holes are emitted instead of electrons.

The bipolar transistor operates as a current device: the current into the base (B) controls a larger current flowing from collector (C) to emitter (E), which is an analog of the base current amplified by the current gain (β or h_{fe}) of the transistor. Transistors have inherently low input impedance at the base unlike the high input impedance at the grid of a vacuum tube. The output voltage depends on the resistor used in the circuit, as the transistor's output is a current that causes a voltage drop in the load resistor. If a resistor is placed in series with the collector (C), the circuit is an inverting voltage amplifier. If it is placed in the emitter lead (E), the circuit gives current gain but no voltage gain: it functions as a non-inverting impedance converter, or buffer. The transistor can also be connected as a current gain device, as it often appears in power supply circuits where the base current is regulated to control the larger collector/emitter currents. Like diodes, transistors are non-linear devices: they require a minimum voltage input to force the transistor into conducting current. There is a linear range that can be used for audio amplification if the input voltage is biased, or added to a fixed DC voltage, to center the input in the linear region of the I-V plot. When used in properly designed and constructed circuits, transistors can give clean, low-noise, high-gain performance in microphone preamplifiers and many other critical audio devices. They can also be used to create larger circuits contained on a single silicon substrate, making functional circuits like operational amplifiers (op-amps), voltage-controlled gain devices, level detectors, oscillators and filters as well as incredibly complicated digital circuits available small packages.

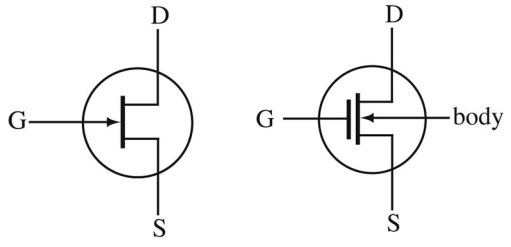


Fig. 10 J-FET and MOS-FET (n-channel)

Another type of transistor often encountered in audio circuits is the field effect transistor or FET. FETs can be either MOS-FET or J-FET types; while constructed differently both make use of the internal electric field to control the flow of charge through the device, hence the name. The junction FET (J-FET) is constructed with the source and drain leads connected to opposite ends of a channel of one semiconductor type embedded in a bulk surround of the opposite type. When a voltage is applied

between the source and the surrounding semiconductor (the gate), the electric field (e-> in Figure 8) at the p-n junction causes the depletion of carriers in the channel and thus acts to control the flow of charge through the channel, changing its resistance as a function of the gate voltage. The drain-source resistance can vary from nearly infinite down to tens or hundreds of ohms, a very useful range. The p-n junction must never become forward or the gate will begin to conduct current as the diode junction switches on, therefore the gate voltage must remain negative relative to the most negative end of the source/drain channel. By placing a resistor between the drain and the power supply voltage, a small gate voltage can be amplified as the drain-source current flows through it. Unlike the bipolar transistor, the FET is a voltage input device: the gate voltage controls the drain-source current and the gate provides a high-resistance input as well.

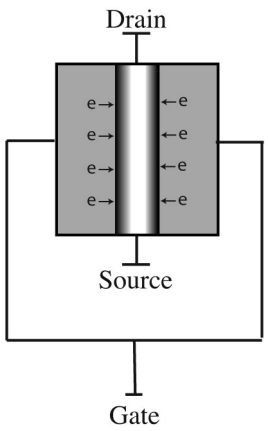


Fig. 11 JFET operation

In the MOS (metal oxide semiconductor) type of FET, two areas of the same type are slightly separated, embedded in a substrate consisting of the opposite type. For an n-channel MOS-FET, the device is constructed on a substrate of p-type material into which are doped a pair of separated n channel areas, the source and drain. The substrate surface between the two n areas is layered with an insulating oxide layer and then a metal electrode, the gate. Because the gate is insulated from the semiconductor materials, the MOS-FET gate has very high input impedance. The gate electrode and the substrate semiconductor beneath the insulating oxide layer form a tiny capacitor, on the order of 1 pF, which results in the accumulation of charges in the substrate area under the gate between the source and drain. The voltage at the gate thereby causes a channel of conduction between the source and drain to be formed, by electrons attracted to the positive charges on the gate in the case of the n-channel FET. The gate voltage acts to control the flow of current from source to drain by altering the number of available carriers induced to join the conduction channel. Because of the very high input impedance of the MOS-FET, relatively small static charges reaching the gate can damage the device, making static protection necessary to prevent damage to these devices.

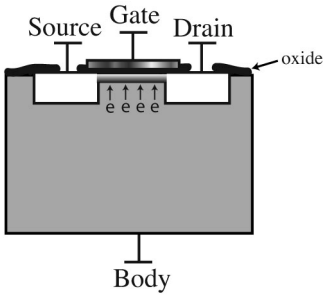


Fig. 12 MOS-FET operation

FETs are used where high input impedances are desired and also in power amplifiers and in digital circuits. The FET can be used as a variable resistance device as well as an amplifier or switch, hence FETs are commonly found in digitally controlled audio switches, providing a way of switching audio signal routing and muting under software control, and in compressors and gates as voltage-controlled variable resistance elements. FETs are useful as preamplifiers in that they have high input impedances that are desirable for high-impedance devices such as instrument pickups where they function as impedance converters as well as amplifiers.

The availability of different amplification devices allows for a variety of circuit designs to accomplish a particular audio task. How a device will sound depends on many smaller influences as well, circuit interactions not always considered in our first approximation descriptions of the circuit elements. For example, the physical construction of a device may affect the sound it produces by creating small, stray capacitances and inductances in high-impedance circuits that cause instability at high frequencies. There are subtle and hard-to-define aspects of audio devices that may improve or degrade sound quality performance and resist an easy scientific analysis. There is an element of art in the design of audio circuits that comes with experience: both circuit design and implementation can contribute to potential differences in sound between two similar-looking pieces of gear. With the development of large integrated circuits, entire functional circuits on a single chip, it has become easy to use off-the-shelf components to build equipment very inexpensively. While more carefully crafted devices benefit from using fewer individual transistors and carefully selected circuit elements, simple audio circuits using a standard integrated circuit amplifier device known as an operational amplifier have made it possible to create audio devices that are easy to design and can sound quite good.

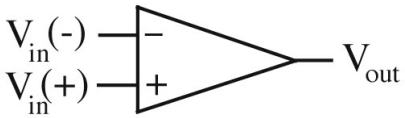


Fig. 13 Operational amplifier (op-amp)

The operational amplifier was originally designed for use in analog computers, where circuits were created to do real-time computation, using analog voltages to solve complicated problems. The operational amplifier, or op-amp, is a high-gain amplifier with two inputs that can be used to add (non-inverting) and subtract (inverting), or it can be used as a linear amplifier whose characteristics are programmed by a network of attached components. Because the output is the amplified difference between the two input voltages, the amplifier is known as a differential amplifier. This will prove to be

a particularly important feature in high-quality audio applications, since it can be used to eliminate interference in complex wiring installations. The flexibility of the op-amp makes it a tempting choice for general-purpose designs and it is found in all performance levels of audio equipment. Op-amps can be made using bipolar transistors alone or in conjunction with field effect transistors, offering a range of possible configurations of these devices and allowing op-amps to be optimized for a range of different applications. While many op-amps are integrated circuits, there are also op-amps custom-made from selected transistors and other components and sealed in plastic modules found in some high-performance audio equipment.

The ideal op-amp device is assumed to have an infinite input impedance and infinite gain. By using a voltage divider connected from the output to the inverting input to provide negative feedback (a portion of the output is fed back to the (-) input), the gain of an inverting amplifier can be set. The ideal op-amp is assumed to have no current flowing into or out of either input and to have no voltage difference between the inputs. While in reality these conditions are not strictly met, they are often close enough to ideal to allow simple implementation of op-amp audio circuits, especially using audio-optimized op-amps now readily available. Since op-amps are so common, it is worth understanding the various common op-amp circuits and we will examine them in detail when we discuss amplifiers and other audio circuitry.

Suggested Reading:

The Art of Electronics, Horowitz, Paul & Winfield Hill, 2nd Ed., Cambridge University Press, 1989
ISBN: 0521370957

© 2004 Jay Kadis (May be used for educational purposes only)