Study Material for UG students

By

Dr. BIMAL KUMAR MANDAL

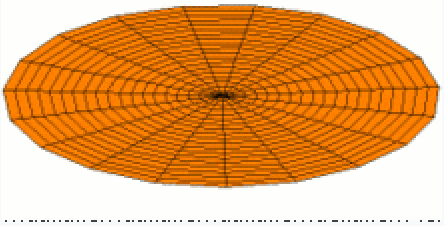
PART-I

Vibration is a mechanical phenomenon whereby oscillations occur about an equilibrium point. The word comes from Latin *vibrationem* ("shaking, brandishing"). The oscillations may be periodic, such as the motion of a pendulum—or random, such as the movement of a tire on a gravel road.

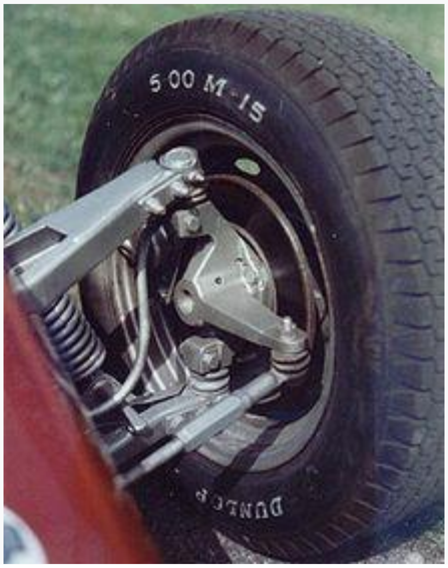
Vibration can be desirable: for example, the motion of a tuning fork, the reed in a woodwind instrument or harmonica, a mobile phone, or the cone of a loudspeaker.

In many cases, however, vibration is undesirable, wasting energy and creating unwanted sound. For example, the vibrational motions of engines, electric motors, or any mechanical device in operation are typically unwanted. Such vibrations could be caused by imbalances in the rotating parts, uneven friction, or the meshing of gear teeth. Careful designs usually minimize unwanted vibrations.

The studies of sound and vibration are closely related. Sound, or pressure waves, are generated by vibrating structures (e.g. vocal cords); these pressure waves can also induce the vibration of structures (e.g. ear drum). Hence, attempts to reduce noise are often related to issues of vibration.



One of the possible modes of vibration of a circular drum (see other modes).



Car Suspension: designing vibration control is undertaken as part of acoustic, automotive or mechanicalengineering.

Types of vibration[edit]

Free vibration occurs when a mechanical system is set in motion with an initial input and allowed to vibrate freely. Examples of this type of vibration are pulling a child back on a swing and letting go, or hitting a tuning fork and letting it ring. The mechanical system vibrates at one or more of its natural frequencies and damps down to motionlessness.

Forced vibration is when a time-varying disturbance (load, displacement or velocity) is applied to a mechanical system. The disturbance can be a periodic and steady-state input, a transient input, or a random input. The periodic input can be a harmonic or a non-harmonic disturbance. Examples of these types of vibration include a washing machine shaking due to an imbalance, transportation vibration caused by an engine or uneven road, or the vibration of a building during an earthquake. For linear systems, the frequency of the steady-state vibration response resulting from the application of a periodic, harmonic input is equal to the frequency of the applied force or motion, with the response magnitude being dependent on the actual mechanical system.

Damped vibration: When the energy of a vibrating system is gradually dissipated by friction and other resistances, the vibrations are said to be damped. The vibrations gradually reduce or change in frequency or intensity or cease and the system rests in its equilibrium positio

Vibration analysis[edit]

This section **does not cite any sources**. Please help improve this section by adding citations to reliable sources. Unsourced material may be challenged and removed. *(July 2013) (Learn how and when to remove this template message)*

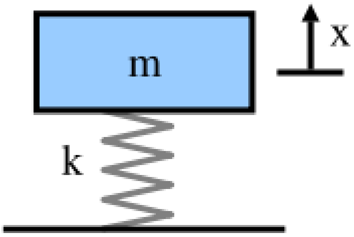
Vibration Analysis (VA), applied in an industrial or maintenance environment aims to reduce maintenance costs and equipment downtime by detecting equipment faults.[3][4] VA is a key component of a Condition Monitoring (CM) program, and is often referred to as Predictive Maintenance (PdM).[5] Most commonly VA is used to detect faults in rotating equipment (Fans, Motors, Pumps, and Gearboxes etc.) such as Unbalance, Misalignment, rolling element bearing faults and resonance conditions.

VA can use the units of Displacement, Velocity and Acceleration displayed as a Time Waveform (TWF), but most commonly the spectrum is used, derived from a Fast Fourier Transform of the TWF. The vibration spectrum provides important frequency information that can pinpoint the faulty component.

The fundamentals of vibration analysis can be understood by studying the simple mass–spring–damper model. Indeed, even a complex structure such as an automobile body can be modeled as a "summation" of simple mass–spring–damper models. The mass–spring–damper model is an example of a simple harmonic oscillator. The mathematics used to describe its behavior is identical to other simple harmonic oscillators such as the RLC circuit.

Note: This article does not include the step-by-step mathematical derivations, but focuses on major vibration analysis equations and concepts. Please refer to the references at the end of the article for detailed derivations.

Free vibration without damping[edit]



To start the investigation of the mass–spring–damper assume the damping is negligible and that there is no external force applied to the mass (i.e. free vibration). The force applied to the mass by the spring is proportional to the amount the spring is stretched "x" (assuming the spring is already compressed due to the weight of the mass). The proportionality constant, k, is the stiffness of the spring and has units of force/distance (e.g. lbf/in or N/m). The negative sign indicates that the force is always opposing the motion of the mass attached to it:

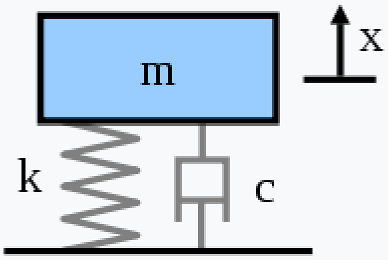
What causes the system to vibrate: from conservation of energy point of view[*edit*]

Vibrational motion could be understood in terms of conservation of energy. In the above example the spring has been

extended by a value of x and therefore some potential energy ( ) is stored in the spring. Once released, the spring tends to return to its un-stretched state (which is the minimum potential energy state) and in the process accelerates the mass. At the point where the spring has reached its un-stretched state all the potential energy that we supplied by stretching

it has been transformed into kinetic energy ( ). The mass then begins to decelerate because it is now compressing the spring and in the process transferring the kinetic energy back to its potential. Thus oscillation of the spring amounts to the transferring back and forth of the kinetic energy into potential energy. In this simple model the mass continues to oscillate forever at the same magnitude—but in a real system, *damping* always dissipates the energy, eventually bringing the spring to rest.

Free vibration with damping[edit]



Mass–spring–damper model

When a "viscous" damper is added to the model this outputs a force that is proportional to the velocity of the mass. The damping is called viscous because it models the effects of a fluid within an object. The proportionality constant *c* is called the damping coefficient and has units of Force over velocity (lbf s/ in or N s/m).

Damped and undamped natural frequencies[*edit*]

The major points to note from the solution are the exponential term and the cosine function. The exponential term defines how quickly the system “damps” down – the larger the damping ratio, the quicker it damps to zero. The cosine function is the oscillating portion of the solution, but the frequency of the oscillations is different from the undamped case.

The frequency in this case is called the "damped natural frequency", and is related to the undamped natural frequency by the following formula:

The damped natural frequency is less than the undamped natural frequency, but for many practical cases the damping ratio is relatively small and hence the difference is negligible. Therefore, the damped and undamped description are often dropped when stating the natural frequency (e.g. with 0.1 damping ratio, the damped natural frequency is only 1% less than the undamped).

The plots to the side present how 0.1 and 0.3 damping ratios effect how the system “rings” down over time. What is often done in practice is to experimentally measure the free vibration after an impact (for example by a hammer) and then determine the natural frequency of the system by measuring the rate of oscillation, as well as the damping ratio by measuring the rate of decay. The natural frequency and damping ratio are not only important in free vibration, but also characterize how a system behaves under forced vibration.

Forced vibration with damping[edit]

The behavior of the spring mass damper model varies with the addition of a harmonic force. A force of this type could, for example, be generated by a rotating imbalance.

Summing the forces on the mass results in the following ordinary differential equation:

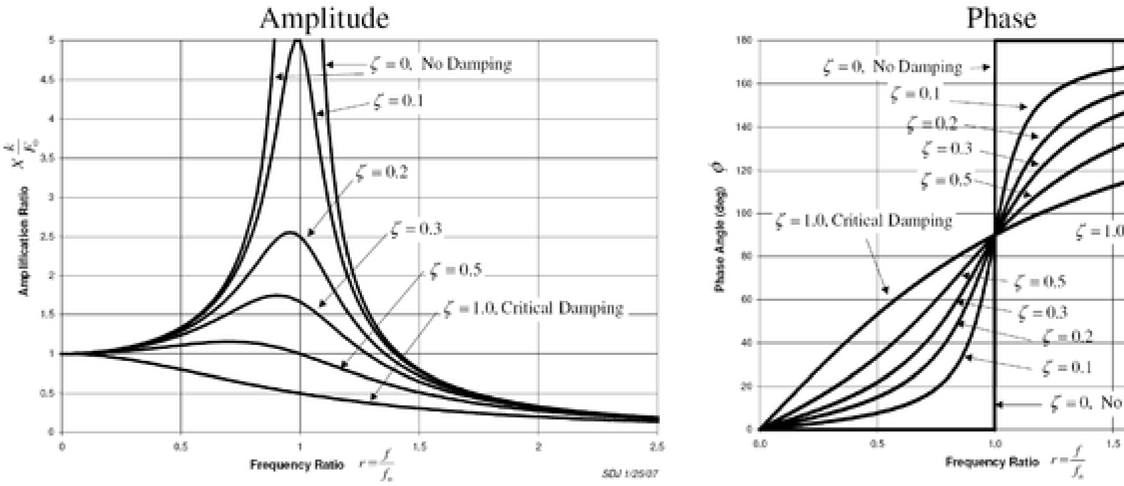
The [steady state](#) solution of this problem can be written as:

The result states that the mass will oscillate at the same frequency, *f*, of the applied force, but with a phase shift

The amplitude of the vibration “X” is defined by the following formula.

Where “r” is defined as the ratio of the harmonic force frequency over the undamped natural frequency of the mass–spring–damper model.

the phase shift, is defined by the following formula.



The plot of these functions, called "the frequency response of the system", presents one of the most important features in

forced vibration. In a lightly damped system when the forcing frequency nears the natural frequency ( ) the amplitude of the vibration can get extremely high. This phenomenon is called **[resonance](#)** (subsequently the natural frequency of a system is often referred to as the resonant frequency). In rotor bearing systems any rotational speed that excites a resonant frequency is referred to as a [critical speed](#).

If resonance occurs in a mechanical system it can be very harmful – leading to eventual failure of the system. Consequently, one of the major reasons for vibration analysis is to predict when this type of resonance may occur and then to determine what steps to take to prevent it from occurring. As the amplitude plot shows, adding damping can significantly reduce the magnitude of the vibration. Also, the magnitude can be reduced if the natural frequency can be shifted away from the forcing frequency by changing the stiffness or mass of the system. If the system cannot be changed, perhaps the forcing frequency can be shifted (for example, changing the speed of the machine generating the force).

PART-II

A **transistor** is a semiconductor device used to amplify or switch electronic signals and electrical power. It is composed of semiconductor material usually with at least three terminals for connection to an external circuit. A voltage or current applied to one pair of the transistor's terminals controls the current through another pair of terminals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal. Today, some transistors are packaged individually, but many more are found embedded in integrated circuits.

The transistor is the fundamental building block of modern electronic devices, and is ubiquitous in modern electronic systems. Julius Edgar Lilienfeld patented a field-effect transistor in 1926[1] but it was not possible to actually construct a working device at that time. The first practically implemented device was a point-contact transistor invented in 1947 by American physicists John Bardeen, Walter Brattain, and William Shockley. The transistor revolutionized the field of electronics, and paved the way for smaller and cheaper radios, calculators, and computers, among other things. The transistor is on the list of IEEE milestones in electronics,[2] and Bardeen, Brattain, and Shockley shared the 1956 Nobel Prize in Physics for their achievement.[3]

Importance[edit]

A Darlington transistor opened up so the actual transistor chip (the small square) can be seen inside. A Darlington transistor is effectively two transistors on the same chip. One transistor is much larger than the other, but both are large in comparison to transistors in large-scale integration because this particular example is intended for power applications.

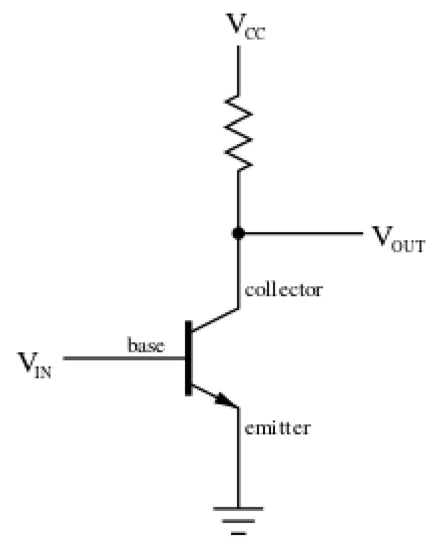
The transistor is the key active component in practically all modern electronics. Many consider it to be one of the greatest inventions of the 20th century.[29] Its importance in today's society rests on its ability to be mass-produced using a highly automated process (semiconductor device fabrication) that achieves astonishingly low per-transistor costs. The invention of the first transistor at Bell Labs was named an IEEE Milestone in 2009.[30]

Although several companies each produce over a billion individually packaged (known as *discrete*) transistors every year,[31] the vast majority of transistors are now produced in integrated circuits (often shortened to *IC*, *microchips* or simply *chips*), along with diodes, resistors, capacitors and other electronic components, to produce complete electronic circuits. A logic gate consists of up to about twenty transistors whereas an advanced microprocessor, as of 2009, can use as many as 3 billion transistors (MOSFETs).[32] "About 60 million transistors were built in 2002… for [each] man, woman, and child on Earth."[33]

The transistor's low cost, flexibility, and reliability have made it a ubiquitous device. Transistorized mechatronic circuits have replaced electromechanical devices in controlling appliances and machinery. It is often easier and cheaper to use a standard microcontroller and write a computer program to carry out a control function than to design an equivalent mechanical system to control that same function.

Simplified operation[edit]

This section **does not cite any sources**. Please help improve this section by adding citations to reliable sources. Unsourced material may be challenged and removed. *(November 2010) (Learn how and when to remove this template message)*



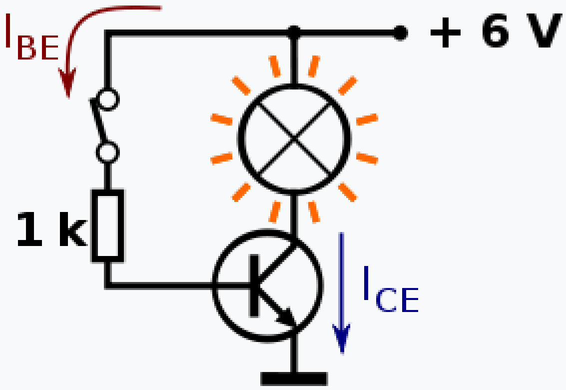
A simple circuit diagram to show the labels of a n–p–n bipolar transistor.

The essential usefulness of a transistor comes from its ability to use a small signal applied between one pair of its terminals to control a much larger signal at another pair of terminals. This property is called gain. It can produce a stronger output signal, a voltage or current, which is proportional to a weaker input signal; that is, it can act as an amplifier. Alternatively, the transistor can be used to turn current on or off in a circuit as an electrically controlled switch, where the amount of current is determined by other circuit elements.

There are two types of transistors, which have slight differences in how they are used in a circuit. A *bipolar transistor* has terminals labeled **base**, **collector**, and **emitter**. A small current at the base terminal (that is, flowing between the base and the emitter) can control or switch a much larger current between the collector and emitter terminals. For a *field-effect transistor*, the terminals are labeled **gate**, **source**, and **drain**, and a voltage at the gate can control a current between source and drain.

The image represents a typical bipolar transistor in a circuit. Charge will flow between emitter and collector terminals depending on the current in the base. Because internally the base and emitter connections behave like a semiconductor diode, a voltage drop develops between base and emitter while the base current exists. The amount of this voltage depends on the material the transistor is made from, and is referred to as V_{BE}.

Transistor as a switch[edit]



BJT used as an electronic switch, in grounded-emitter configuration.

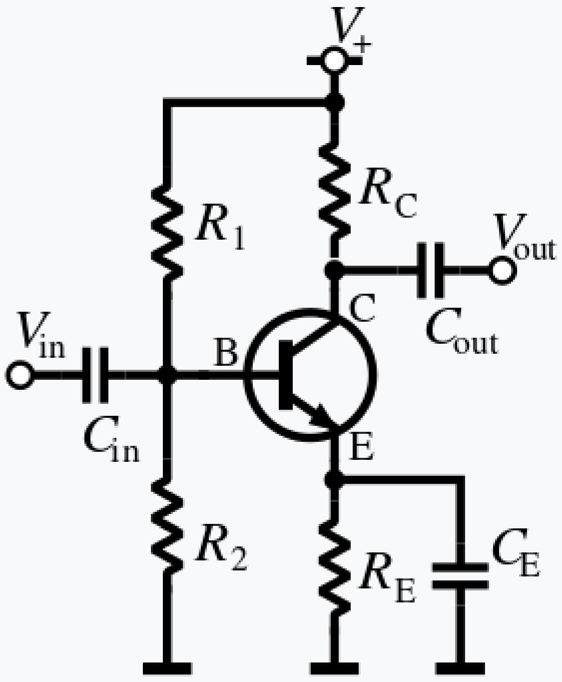
Transistors are commonly used in digital circuits as electronic switches which can be either in an "on" or "off" state, both for high-power applications such as switched-mode power supplies and for low-power applications such as logic gates. Important parameters for this application include the current switched, the voltage handled, and the switching speed, characterised by the rise and fall times.

In a grounded-emitter transistor circuit, such as the light-switch circuit shown, as the base voltage rises, the emitter and collector currents rise exponentially. The collector voltage drops because of reduced resistance from collector to emitter. If the voltage difference between the collector and emitter were zero (or near zero), the collector current would be limited only by the load resistance (light bulb) and the supply voltage. This is called *saturation* because current is flowing from collector to emitter freely. When saturated, the switch is said to be *on*.[34]

Providing sufficient base drive current is a key problem in the use of bipolar transistors as switches. The transistor provides current gain, allowing a relatively large current in the collector to be switched by a much smaller current into the base terminal. The ratio of these currents varies depending on the type of transistor, and even for a particular type, varies depending on the collector current. In the example light-switch circuit shown, the resistor is chosen to provide enough base current to ensure the transistor will be saturated.

In a switching circuit, the idea is to simulate, as near as possible, the ideal switch having the properties of open circuit when off, short circuit when on, and an instantaneous transition between the two states. Parameters are chosen such that the "off" output is limited to leakage currents too small to affect connected circuitry; the resistance of the transistor in the "on" state is too small to affect circuitry; and the transition between the two states is fast enough not to have a detrimental effect.

Transistor as an amplifier[edit]



Amplifier circuit, common-emitter configuration with a voltage-divider bias circuit.

The common-emitter amplifier is designed so that a small change in voltage (V_{in}) changes the small current through the base of the transistor; the transistor's current amplification combined with the properties of the circuit means that small swings in V_{in} produce large changes in V_{out}.

Various configurations of single transistor amplifier are possible, with some providing current gain, some voltage gain, and some both.

From mobile phones to televisions, vast numbers of products include amplifiers for sound reproduction, radio transmission, and signal processing. The first discrete-transistor audio amplifiers barely supplied a few hundred milliwatts, but power and audio fidelity gradually increased as better transistors became available and amplifier architecture evolved.

Modern transistor audio amplifiers of up to a few hundred watts are common and relatively inexpensive.

Comparison with vacuum tubes[edit]

Before transistors were developed, vacuum (electron) tubes (or in the UK "thermionic valves" or just "valves") were the main active components in electronic equipment.

Advantages[edit]

The key advantages that have allowed transistors to replace vacuum tubes in most applications are

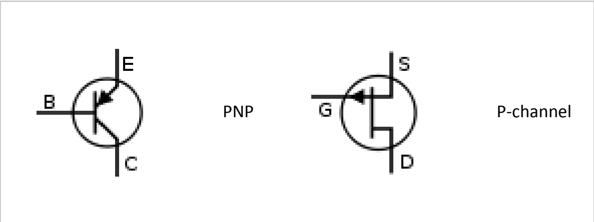
- no cathode heater (which produces the characteristic orange glow of tubes), reducing power consumption, eliminating delay as tube heaters warm up, and immune from cathode poisoning and depletion;
- very small size and weight, reducing equipment size;
- large numbers of extremely small transistors can be manufactured as a single integrated circuit;
- low operating voltages compatible with batteries of only a few cells;
- circuits with greater energy efficiency are usually possible. For low-power applications (e.g., voltage amplification) in particular, energy consumption can be very much less than for tubes;
- inherent reliability and very long life; tubes always degrade and fail over time. Some transistorized devices have been in service for more than 50 years[citation needed] ;
- complementary devices available, providing design flexibility including complementary-symmetry circuits, not possible with vacuum tubes;
- very low sensitivity to mechanical shock and vibration, providing physical ruggedness and virtually eliminating shock-induced spurious signals (e.g., microphonics in audio applications);
- not susceptible to breakage of a glass envelope, leakage, outgassing, and other physical damage.

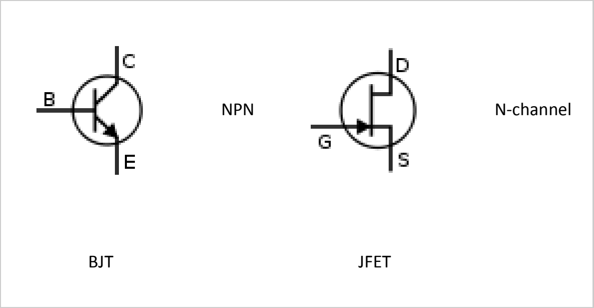
Limitations[edit]

Transistors have the following limitations:

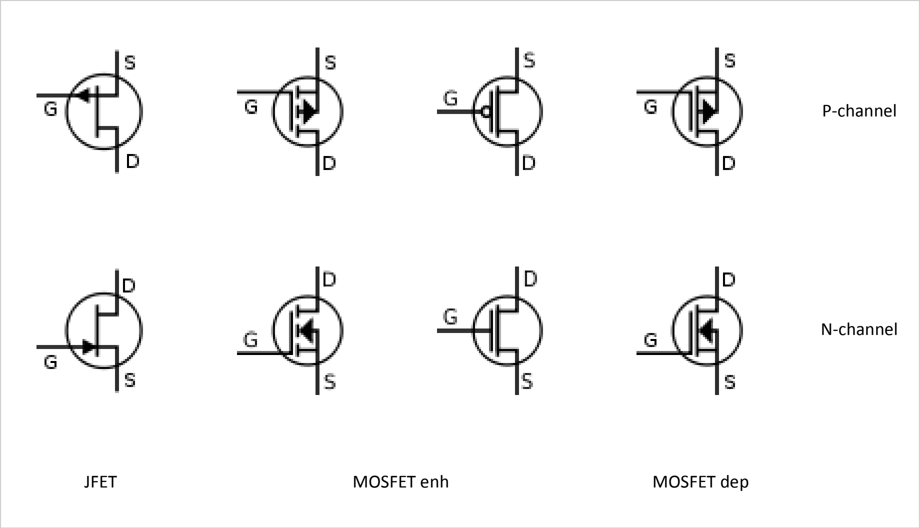
- silicon transistors can age and fail;[35]
- high-power, high-frequency operation, such as that used in over-the-air television broadcasting, is better achieved in vacuum tubes due to improved electron mobility in a vacuum;
- solid-state devices are susceptible to damage from very brief electrical and thermal events, including electrostatic discharge in handling; vacuum tubes are electrically much more rugged;
- sensitivity to radiation and cosmic rays (special radiation-hardened chips are used for spacecraft devices);
- vacuum tubes in audio applications create significant lower-harmonic distortion, the so-called tube sound, which some people prefer.[36]

Types[edit]





BJT and JFET symbols



JFET and MOSFET symbols

Transistors are categorized by

- semiconductor material: the metalloids germanium (first used in 1947) and silicon (first used in 1954)—in amorphous, polycrystalline and monocrystalline form—, the compounds gallium arsenide (1966) and silicon carbide (1997), the alloy silicon-germanium (1989), the allotrope of carbon graphene (research ongoing since 2004), etc. (see Semiconductor material);
- structure: BJT, JFET, IGFET (MOSFET), insulated-gate bipolar transistor, "other types";
- electrical polarity (positive and negative): n–p–n, p–n–p (BJTs), n-channel, p-channel (FETs);
- maximum power rating: low, medium, high;
- maximum operating frequency: low, medium, high, radio (RF), microwave frequency (the maximum effective frequency of a transistor in a common-emitter or common-source circuit is denoted by the term f_T, an abbreviation for transition frequency—the frequency of transition is the frequency at which the transistor yields unity voltage gain)
- application: switch, general purpose, audio, high voltage, super-beta, matched pair;
- physical packaging: through-hole metal, through-hole plastic, surface mount, ball grid array, power modules (see Packaging);
- amplification factor h_{FE}, β_F (transistor beta)[37] or g_m (transconductance).

Hence, a particular transistor may be described as *silicon, surface-mount, BJT, n–p–n, low-power, high-frequency switch*.

A popular way to remember which symbol represents which type of transistor is to look at the arrow and how it is arranged. Within an NPN transistor symbol, the arrow will Not Point iN. Conversely, within the PNP symbol you see that the arrow Points iN Proudly.

Bipolar junction transistor (BJT)[edit]

Main article: Bipolar junction transistor

Bipolar transistors are so named because they conduct by using both majority and minority carriers. The bipolar junction transistor, the first type of transistor to be mass-produced, is a combination of two junction diodes, and is formed of either a thin layer of p-type semiconductor sandwiched between two n-type semiconductors (an n–p–n transistor), or a thin layer of n-type semiconductor sandwiched between two p-type semiconductors (a p–n–p transistor). This construction produces two p–n junctions: a base–emitter junction and a base–collector junction, separated by a thin region of semiconductor known as the base region (two junction diodes wired together without sharing an intervening semiconducting region will not make a transistor).

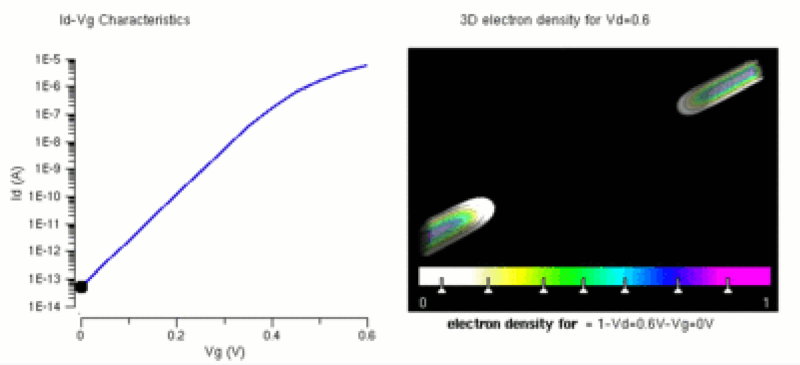
BJTs have three terminals, corresponding to the three layers of semiconductor—an *emitter*, a *base*, and a *collector*. They are useful in amplifiers because the currents at the emitter and collector are controllable by a relatively small base current.[38] In an n–p–n transistor operating in the active region, the emitter–base junction is forward biased (electrons and holes recombine at the junction), and electrons are injected into the base region. Because the base is narrow, most of these electrons will diffuse into the reverse-biased (electrons and holes are formed at, and move away from the junction) base–collector junction and be swept into the collector; perhaps one-hundredth of the electrons will recombine in the base, which is the dominant mechanism in the base current. By controlling the number of electrons that can leave the base, the number of electrons entering the collector can be controlled.[38] Collector current is approximately β (common-emitter current gain) times the base current. It is typically greater than 100 for small-signal transistors but can be smaller in transistors designed for high-power applications.

Unlike the field-effect transistor (see below), the BJT is a low-input-impedance device. Also, as the base–emitter voltage (V_{BE}) is increased the base–emitter current and hence the collector–emitter current (I_{CE}) increase exponentially according to the Shockley diode model and the Ebers-Moll model. Because of this exponential relationship, the BJT has a higher transconductance than the FET.

Bipolar transistors can be made to conduct by exposure to light, because absorption of photons in the base region generates a photocurrent that acts as a base current; the collector current is approximately β times the photocurrent. Devices designed for this purpose have a transparent window in the package and are called phototransistors.

Field-effect transistor (FET)[edit]

Main articles: Field-effect transistor, MOSFET, and JFET



Operation of a FET and its Id-Vg curve. At first, when no gate voltage is applied. There is no inversion electron in the channel, the device is OFF. As gate voltage increase, inversion electron density in the channel increase, current increase, the device turns on.

The *field-effect transistor*, sometimes called a *unipolar transistor*, uses either electrons (in *n-channel FET*) or holes (in *p-channel FET*) for conduction. The four terminals of the FET are named *source*, *gate*, *drain*, and *body* (*substrate*). On most FETs, the body is connected to the source inside the package, and this will be assumed for the following description.

In a FET, the drain-to-source current flows via a conducting channel that connects the *source* region to the *drain* region. The conductivity is varied by the electric field that is produced when a voltage is applied between the gate and source terminals; hence the current flowing between the drain and source is controlled by the voltage applied between the gate and source. As the gate–source voltage (V_{GS}) is increased, the drain–source current (I_{DS}) increases exponentially for V_{GS} below threshold, and then at a roughly quadratic rate ($I_{GS} \propto (V_{GS} - V_T)^2$) (where V_T is the threshold voltage at which drain current begins)[39] in the "space-charge-limited" region above threshold. A quadratic behavior is not observed in modern devices, for example, at the 65 nm technology node.[40]

For low noise at narrow bandwidth the higher input resistance of the FET is advantageous.

FETs are divided into two families: *junction FET* (JFET) and *insulated gate FET* (IGFET). The IGFET is more commonly known as a *metal–oxide–semiconductor FET* (MOSFET), reflecting its original construction from layers of metal (the gate), oxide (the insulation), and semiconductor. Unlike IGFETs, the JFET gate forms a p–n diode with the channel which lies between the source and drain. Functionally, this makes the n-channel JFET the solid-state equivalent of the vacuum tube triode which, similarly, forms a diode between its grid and cathode. Also, both devices operate in the *depletion mode*, they both have a high input impedance, and they both conduct current under the control of an input voltage.

Metal–semiconductor FETs (MESFETs) are JFETs in which the reverse biased p–n junction is replaced by a metal–semiconductor junction. These, and the HEMTs (high-electron-mobility transistors, or HFETs), in which a two-dimensional electron gas with very high carrier mobility is used for charge transport, are especially suitable for use at very high frequencies (microwave frequencies; several GHz).

FETs are further divided into *depletion-mode* and *enhancement-mode* types, depending on whether the channel is turned on or off with zero gate-to-source voltage. For enhancement mode, the channel is off at zero bias, and a gate potential can "enhance" the conduction. For the depletion mode, the channel is on at zero bias, and a gate potential (of the opposite polarity) can "deplete" the channel, reducing conduction. For either mode, a more positive gate voltage corresponds to a higher current for n-channel devices and a lower current for p-channel devices. Nearly all JFETs are depletion-mode because the diode junctions would forward bias and conduct if they were enhancement-mode devices; most IGFETs are enhancement-mode types.

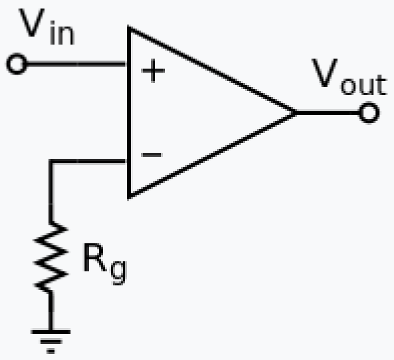
PART-III

An **operational amplifier** (often **op-amp** or **opamp**) is a DC-coupled high-gain electronic voltage amplifier with a differential input and, usually, a single-ended output.[1] In this configuration, an op-amp produces an output potential (relative to circuit ground) that is typically hundreds of thousands of times larger than the potential difference between its input terminals. Operational amplifiers had their origins in analog computers, where they were used to perform mathematical operations in many linear, non-linear and frequency-dependent circuits. The popularity of the op-amp as a building block in analog circuits is due to its versatility. Due to negative feedback, the characteristics of an op-amp circuit, its gain, input and output impedance, bandwidth etc. are determined by external components and have little dependence on temperature coefficients or manufacturing variations in the op-amp itself.

Op-amps are among the most widely used electronic devices today, being used in a vast array of consumer, industrial, and scientific devices. Many standard IC op-amps cost only a few cents in moderate production volume; however some integrated or hybrid operational amplifiers with special performance specifications may cost over $100 US in small quantities.[2] Op-amps may be packaged as components, or used as elements of more complex integrated circuits.

The op-amp is one type of differential amplifier. Other types of differential amplifier include the fully differential amplifier (similar to the op-amp, but with two outputs), the instrumentation amplifier (usually built from three op-amps), the isolation amplifier (similar to the instrumentation amplifier, but with tolerance to common-mode voltages that would destroy an ordinary op-amp), and negative feedback amplifier (usually built from one or more op-amps and a resistive feedback network).

Operation[edit]



An op-amp without negative feedback (a comparator)

The amplifier's differential inputs consist of a non-inverting input (+) with voltage V_+ and an inverting input (–) with voltage V_-; ideally the op-amp amplifies only the difference in voltage between the two, which is called the *differential input voltage*. The output voltage of the op-amp V_{out} is given by the equation:

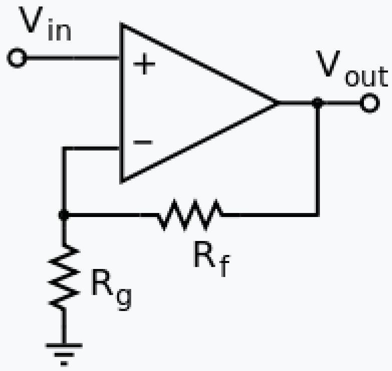
where A_{OL} is the open-loop gain of the amplifier (the term "open-loop" refers to the absence of a feedback loop from the output to the input).

Open loop amplifier[edit]

The magnitude of A_{OL} is typically very large—100,000 or more for integrated circuit op-amps—and therefore even a quite small difference between V_+ and V_- drives the amplifier output nearly to the supply voltage. Situations in which the output voltage is equal to or greater than the supply voltage are referred to as *saturation* of the amplifier. The magnitude of A_{OL} is not well controlled by the manufacturing process, and so it is impractical to use an open loop amplifier as a stand-alone differential amplifier.

Without negative feedback, and perhaps with positive feedback for regeneration, an op-amp acts as a comparator. If the inverting input is held at ground (0 V) directly or by a resistor R_g, and the input voltage V_{in} applied to the non-inverting input is positive, the output will be maximum positive; if V_{in} is negative, the output will be maximum negative. Since there is no feedback from the output to either input, this is an *open loop* circuit acting as a comparator.

Closed loop[edit]



An op-amp with negative feedback (a non-inverting amplifier)

If predictable operation is desired, negative feedback is used, by applying a portion of the output voltage to the inverting input. The *closed loop* feedback greatly reduces the gain of the circuit. When negative feedback is used, the circuit's overall gain and response becomes determined mostly by the feedback network, rather than by the op-amp characteristics. If the feedback network is made of components with values small relative to the op amp's input impedance, the value of the op-amp's open loop response A_{OL} does not seriously affect the circuit's performance. The response of the op-amp circuit with its input, output, and feedback circuits to an input is characterized mathematically by a transfer function; designing an op-amp circuit to have a desired transfer function is in the realm of electrical engineering. The transfer functions are important in most applications of op-amps, such as in analog computers. High input impedance at the input terminals and low output impedance at the output terminal(s) are particularly useful features of an op-amp.

In the non-inverting amplifier on the right, the presence of negative feedback via the voltage divider R_f, R_g determines the *closed-loop gain*$A_{CL} = V_{out} / V_{in}$. Equilibrium will be established when V_{out} is just sufficient to "reach around and pull" the inverting input to the same voltage as V_{in}. The voltage gain of the entire circuit is thus $1 + R_f/R_g$. As a simple example, if V_{in} = 1 V and $R_f = R_g$, V_{out} will be 2 V, exactly the amount required to keep V_- at 1 V. Because of the feedback provided by the R_f, R_g network, this is a *closed loop* circuit.

Another way to analyze this circuit proceeds by making the following (usually valid) assumptions:[3]

- When an op-amp operates in linear (i.e., not saturated) mode, the difference in voltage between the non-inverting (+) pin and the inverting (−) pin is negligibly small.
- The input impedance between (+) and (−) pins is much larger than other resistances in the circuit.

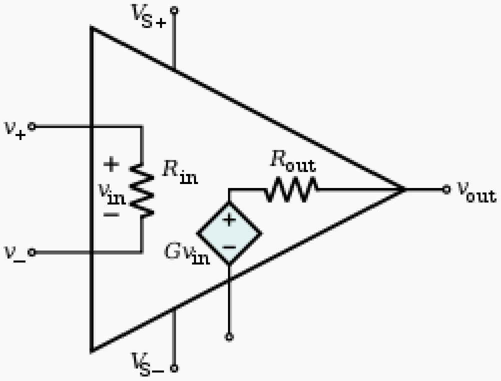
The input signal V_{in} appears at both (+) and (−) pins, resulting in a current i through R_g equal to V_{in}/R_g.

> Since Kirchhoff's current law states that the same current must leave a node as enter it, and since the impedance into the (−) pin is near infinity, we can assume practically all of the same current i flows through R_f, creating an output voltage

> > By combining terms, we determine the closed-loop gain A_{CL}:

Op-amp characteristics[edit]

Ideal op-amps[edit]



An equivalent circuit of an operational amplifier that models some resistive non-ideal parameters.

An ideal op-amp is usually considered to have the following characteristics:[4][5]

- Infinite open-loop gain G = v_{out} / v_{in}
- Infinite input impedance R_{in}, and so zero input current
- Zero input offset voltage
- Infinite output voltage range
- Infinite bandwidth with zero phase shift and infinite slew rate
- Zero output impedance R_{out}
- Zero noise
- Infinite common-mode rejection ratio (CMRR)
- Infinite power supply rejection ratio.

These ideals can be summarized by the two "golden rules":

I. **In a closed loop** the output attempts to do whatever is necessary to make the voltage difference between the inputs zero.
II. The inputs draw no current.[6]:177

The first rule only applies in the usual case where the op-amp is used in a closed-loop design (negative feedback, where there is a signal path of some sort feeding back from the output to the inverting input). These rules are commonly used as a good first approximation for analyzing or designing op-amp circuits.[6]:177

None of these ideals can be perfectly realized. A real op-amp may be modeled with non-infinite or non-zero parameters using equivalent resistors and capacitors in the op-amp model. The designer can then include these effects into the overall performance of the final circuit. Some parameters may turn out to have negligible effect on the final design while others represent actual limitations of the final performance that must be evaluated.

Real op-amps[edit]

Real op-amps differ from the ideal model in various aspects.

DC imperfections[edit]

Real operational amplifiers suffer from several non-ideal effects:

Finite gain

Open-loop gain is infinite in the ideal operational amplifier but finite in real operational amplifiers. Typical devices exhibit open-loop DC gain ranging from 100,000 to over 1 million. So long as the loop gain (i.e., the product of open-loop and feedback gains) is very large, the circuit gain will be determined entirely by the amount of negative feedback (i.e., it will be independent of open-loop gain). In cases where closed-loop gain must be very high, the feedback gain will be very low, and the low feedback gain causes low loop gain; in these cases, the operational amplifier will cease to behave ideally.

Finite input impedances

The *differential input impedance* of the operational amplifier is defined as the impedance *between* its two inputs; the *common-mode input impedance* is the impedance from each input to ground. MOSFET-input operational amplifiers often have protection circuits that effectively short circuit any input differences greater than a small threshold, so the input impedance can appear to be very low in some tests. However, as long as these operational amplifiers are used in a typical high-gain negative feedback application, these protection circuits will be inactive. The input bias and leakage currents described below are a more important design parameter for typical operational amplifier applications.

Non-zero output impedance

Low output impedance is important for low-impedance loads; for these loads, the voltage drop across the output impedance effectively reduces the open loop gain. In configurations with a voltage-sensing negative feedback, the output impedance of the amplifier is effectively lowered; thus, in linear applications, op-amp circuits usually exhibit a very low output impedance indeed.

Low-impedance outputs typically require high quiescent (i.e., idle) current in the output stage and will dissipate more power, so low-power designs may purposely sacrifice low output impedance.

Input current

Due to biasing requirements or leakage, a small amount of current (typically ~10 nanoamperes for bipolar op-amps, tens of picoamperes (pA) for JFET input stages, and only a few pA for MOSFET input stages) flows into the inputs. When large resistors or sources with high output impedances are used in the circuit, these small currents can produce large unmodeled voltage drops. If the input currents are matched, *and* the impedance looking *out* of *both* inputs are matched, then the voltages produced at each input will be equal. Because the operational amplifier operates on the *difference* between its inputs, these matched voltages will have no effect. It is more common for the input currents to be slightly mismatched. The difference is called input offset current, and even with matched resistances a small *offset voltage* (different from the input offset voltage below) can be produced. This offset voltage can create offsets or drifting in the operational amplifier.

Input offset voltage

This voltage, which is what is required across the op-amp's input terminals to drive the output voltage to zero.[7][nb 1] In the perfect amplifier, there would be no input offset voltage. However, it exists in actual op-amps because of imperfections in the differential amplifier that constitutes the input stage of the vast majority of these devices. Input offset voltage creates two problems: First, due to the amplifier's high voltage gain, it virtually assures that the amplifier output will go into saturation if it is operated without negative feedback, even when the input terminals are wired together. Second, in a closed loop, negative feedback configuration, the input offset voltage is amplified along with the signal and this may pose a problem if high precision DC amplification is required or if the input signal is very small.[nb 2]

Common-mode gain

A perfect operational amplifier amplifies only the voltage difference between its two inputs, completely rejecting all voltages that are common to both. However, the differential input stage of an operational amplifier is never perfect, leading to the amplification of these common voltages to some degree. The standard measure of this defect is called the common-mode rejection ratio (denoted CMRR). Minimization of common mode gain is usually important in non-inverting amplifiers (described below) that operate at high amplification.

Power-supply rejection

The output of a perfect operational amplifier will be completely independent from its power supply. Every real operational amplifier has a finite power supply rejection ratio (PSRR) that reflects how well the op-amp can reject changes in its supply voltage.

emperature effects

All parameters change with temperature. Temperature drift of the input offset voltage is especially important.

Drift

Real op-amp parameters are subject to slow change over time and with changes in temperature, input conditions, etc.

AC imperfections[edit]

The op-amp gain calculated at DC does not apply at higher frequencies. Thus, for high-speed operation, more sophisticated considerations must be used in an op-amp circuit design.

Finite bandwidth

All amplifiers have finite bandwidth. To a first approximation, the op-amp has the frequency response of an integrator with gain. That is, the gain of a typical op-amp is inversely proportional to frequency and is characterized by its gain–bandwidth product (GBWP). For example, an op-amp with a GBWP of 1 MHz would have a gain of 5 at 200 kHz, and a gain of 1 at 1 MHz. This dynamic response coupled with the very high DC gain of the op-amp gives it the characteristics of a first-order low-pass filter with very high DC gain and low cutoff frequency given by the GBWP divided by the DC gain.

The finite bandwidth of an op-amp can be the source of several problems, including:

Stability

Associated with the bandwidth limitation is a phase difference between the input signal and the amplifier output that can lead to oscillation in some feedback circuits. For example, a sinusoidal output signal meant to interfere destructively with an input signal of the same frequency will interfere constructively if delayed by 180 degrees forming positive feedback. In these cases, the feedback circuit can be stabilized by means of frequency compensation, which increases the gain or phase margin of the open-loop circuit. The circuit designer can implement this compensation externally with a separate circuit component. Alternatively, the compensation can be implemented within the operational amplifier with the addition of a dominant pole that sufficiently attenuates the high-frequency gain of the operational amplifier. The location of this pole may be fixed internally by the manufacturer or configured by the circuit designer using methods specific to the op-amp. In general, dominant-pole frequency compensation reduces the bandwidth of the op-amp even further. When the desired closed-loop gain is high, op-amp frequency compensation is often not needed because the requisite open-loop gain is sufficiently low; consequently, applications with high closed-loop gain can make use of op-amps with higher bandwidths.

Distortion, and other effects

Limited bandwidth also results in lower amounts of feedback at higher frequencies, producing higher distortion, and output impedance as the frequency increases.

Typical low-cost, general-purpose op-amps exhibit a GBWP of a few megahertz. Specialty and high-speed op-amps exist that can achieve a GBWP of hundreds of megahertz. For very high-frequency circuits, a current-feedback operational amplifier is often used.

Noise

Amplifiers generate random voltage at the output even when there is no signal applied. This can be due to thermal noise and flicker noise of the devices. For applications with high gain or high bandwidth, noise becomes a very important consideration.

Input capacitance

Most important for high frequency operation because it reduces input impedance and may cause phase shifts.

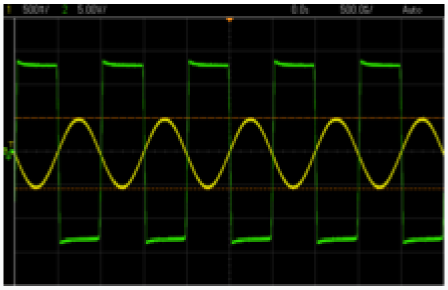
Common-mode gain

See DC imperfections, above.

Power-supply rejection

With increasing frequency the power-supply rejection usually gets worse. So it can be important to keep the supply clean of higher frequency ripples and signals, e.g. by the use of bypass capacitors.

Non-linear imperfections[*edit*]



The input (yellow) and output (green) of a saturated op amp in an inverting amplifier

Saturation

Output voltage is limited to a minimum and maximum value close to the power supply voltages.[nb 3] The output of older op-amps can reach to within one or two volts of the supply rails. The output of newer so-called "rail to rail" op-amps can reach to within millivolts of the supply rails when providing low output currents.

Slewing

The amplifier's output voltage reaches its maximum rate of change, the slew rate, usually specified in volts per microsecond. When slewing occurs, further increases in the input signal have no effect on the rate of change of the output. Slewing is usually caused by the input stage saturating; the result is a constant current i driving a capacitance C in the amplifier (especially those capacitances used to implement its frequency compensation); the slew rate is limited by $dv/dt=i/C$.

Slewing is associated with the *large-signal* performance of an op-amp. Consider, for example, an op-amp configured for a gain of 10. Let the input be a 1 V, 100 kHz sawtooth wave. That is, the amplitude is 1 V and the period is 10 microseconds. Accordingly, the rate of change (i.e., the slope) of the input is 0.1 V per microsecond. After 10x amplification, the output should be a 10 V, 100 kHz sawtooth, with a corresponding slew rate of 1 V per microsecond. However, the classic 741 op-amp has a 0.5 V per microsecond slew rate specification, so that its output can rise to no more than 5 V in the sawtooth's 10 microsecond period. Thus, if one were to measure the output, it would be a 5 V, 100 kHz sine-wave, rather than a 10 V, 100 kHz sawtooth.

Next consider the same amplifier and 100 kHz sawtooth, but now the input amplitude is 100 mV rather than 1 V. After 10x amplification the output is a 1 V, 100 kHz sawtooth with a corresponding slew rate of 0.1 V per microsecond. In this instance the 741 with its 0.5 V per microsecond slew rate will amplify the input properly.

Modern high speed op-amps can have slew rates in excess of 5,000 V per microsecond. However, it is more common for op-amps to have slew rates in the range 5-100 V per microsecond. For example, the general purpose TL081 op-amp has a slew rate of 13 V per microsecond. As a general rule, low power and small bandwidth op-amps have low slew rates. As an example, the LT1494 micropower op-amp consumes 1.5 microamp but has a 2.7 kHz gain-bandwidth product and a 0.001 V per microsecond slew rate.