

Vacuum:

Fundamental Physics Concept

(Concepts and Applications)



Jacqueline Steel

First Edition, 2012

ISBN 978-81-323-1675-6

WWT

© All rights reserved.

Published by:

Learning Press

4735/22 Prakashdeep Bldg,

Ansari Road, Darya Ganj,

Delhi - 110002

Email: info@wtbooks.com

WORLD TECHNOLOGIES

Table of Contents

Chapter 1 - Introduction to Vacuum

Chapter 2 - Outer Space

Chapter 3 - Vacuum Pump

Chapter 4 - Pressure Measurement

Chapter 5 - Vacuum Tube

Chapter 6 - Vacuum State

WWT

Chapter- 1

Introduction to Vacuum



Pump to demonstrate vacuum

In everyday usage, **vacuum** is a volume of space that is essentially empty of matter, such that its gaseous pressure is much less than atmospheric pressure. The word comes from the Latin term for "empty". Even putting aside the complexities of the quantum vacuum, the classical notion of a **perfect vacuum** with gaseous pressure of exactly zero is only a philosophical concept and is never observed in practice. Physicists often discuss ideal test results that would occur in a perfect vacuum, which they simply call "vacuum" or "free

space", and use the term **partial vacuum** to refer to real vacuum. The Latin term **in vacuo** is also used to describe an object as being in what would otherwise be a vacuum.

The *quality* of a vacuum refers to how closely it approaches a perfect vacuum. Residual gas pressure is a primary indicator of quality, and is most commonly measured in units called torr, even in metric contexts. Lower pressures indicate higher quality, although other variables must also be taken into account.

Vacuum has been a frequent topic of philosophical debate since Ancient Greek times, but was not studied empirically until the 17th century. Evangelista Torricelli produced the first laboratory vacuum in 1643, and other experimental techniques were developed as a result of his theories of atmospheric pressure. A **torricellian vacuum** is created by filling a tall glass container closed at one end with mercury and then inverting the container into a bowl to contain the mercury.

Vacuum became a valuable industrial tool in the 20th century with the introduction of incandescent light bulbs and vacuum tubes, and a wide array of vacuum technology has since become available. The recent development of human spaceflight has raised interest in the impact of vacuum on human health, and on life forms in general.



A large vacuum chamber

Etymology

From Latin **vacuum** (*an empty space, void*) noun use of neuter of *vacuus* (*empty*) related to *vacare* (*be empty*).

"Vacuum" is one of the few words in the English language that contains two consecutive 'u's.

Uses



Light bulbs contain a partial vacuum, usually backfilled with argon, which protects the tungsten filament

Vacuum is useful in a variety of processes and devices. Its first widespread use was in the incandescent light bulb to protect the filament from chemical degradation. The chemical inertness produced by a vacuum is also useful for electron beam welding, cold welding, vacuum packing and vacuum frying. Ultra-high vacuum is used in the study of atomically clean substrates, as only a very good vacuum preserves atomic-scale clean surfaces for a

reasonably long time (on the order of minutes to days). High to ultra-high vacuum removes the obstruction of air, allowing particle beams to deposit or remove materials without contamination. This is the principle behind chemical vapor deposition, physical vapor deposition, and dry etching which are essential to the fabrication of semiconductors and optical coatings, and to surface science. The reduction of convection provides the thermal insulation of thermos bottles. Deep vacuum lowers the boiling point of liquids and promotes low temperature outgassing which is used in freeze drying, adhesive preparation, distillation, metallurgy, and process purging. The electrical properties of vacuum make electron microscopes and vacuum tubes possible, including cathode ray tubes. The elimination of air friction is useful for flywheel energy storage and ultracentrifuges.

Vacuum driven machines

Vacuums are commonly used to produce suction, which has an even wider variety of applications. The Newcomen steam engine used vacuum instead of pressure to drive a piston. In the 19th century, vacuum was used for traction on Isambard Kingdom Brunel's experimental atmospheric railway. Vacuum brakes were once widely used on trains in the UK but, except on heritage railways, they have been replaced by air brakes.

Manifold vacuum can be used to drive accessories on automobiles. The best-known application is the vacuum servo, used to provide power assistance for the brakes. Obsolete applications include vacuum-driven windscreen wipers and fuel pumps.

Quantum-mechanical definition

In quantum mechanics, the vacuum is defined as the state (i.e. solution to the equations of the theory) with the lowest energy. To first approximation, this is simply a state with no particles, hence the name.

Even an ideal vacuum, thought of as the complete absence of anything, will not in practice remain empty. Consider a vacuum chamber that has been completely evacuated, so that the (classical) particle concentration is zero. The walls of the chamber will emit light in the form of black body radiation. This light carries momentum, so the vacuum does have a radiation pressure. This limitation applies even to the vacuum of interstellar space. Even if a region of space contains no particles, the cosmic microwave background fills the entire universe with black body radiation.

An ideal vacuum cannot exist even inside of a molecule. Each atom in the molecule exists as a probability function of space, which has a certain non-zero value everywhere in a given volume. Thus, even "between" the atoms there is a certain probability of finding a particle, so the space cannot be said to be a vacuum.

More fundamentally, quantum mechanics predicts that vacuum energy will be different from its naive, classical value. The quantum correction to the energy is called the zero-point energy and consists of energies of virtual particles that have a brief existence. This

is called vacuum fluctuation. Vacuum fluctuations may also be related to the so-called cosmological constant in cosmology. The best evidence for vacuum fluctuations is the Casimir effect and the Lamb shift.

In quantum field theory and string theory, the term "vacuum" is used to represent the ground state in the Hilbert space, that is, the state with the lowest possible energy. In free (non-interacting) quantum field theories, this state is analogous to the ground state of a quantum harmonic oscillator. If the theory is obtained by quantization of a classical theory, each stationary point of the energy in the configuration space gives rise to a single vacuum. String theory is believed to have a huge number of vacua - the so-called string theory landscape.

Outgassing

Outgassing (sometimes called **offgassing**, particularly when in reference to indoor air quality) is the release of a gas that was dissolved, trapped, frozen, absorbed or adsorbed in some material. It can include sublimation and evaporation which are phase transitions of a substance into a gas, as well as desorption, seepage from cracks or internal volumes and gaseous products of slow chemical reactions. Boiling is generally thought of as a separate phenomenon from outgassing because it consists of a phase transition of a liquid into a vapor made of the same substance.

Outgassing in a vacuum

Outgassing is a challenge to creating and maintaining clean high-vacuum environments. NASA maintains a list of low-outgassing materials to be used for spacecraft, as outgassing products can condense onto optical elements, thermal radiators, or solar cells and obscure them. Materials not normally considered absorbent can release enough light-weight molecules to interfere with industrial or scientific vacuum processes. Moisture, sealants, lubricants, and adhesives are the most common sources, but even metals and glasses can release gases from cracks or impurities. The rate of outgassing increases at higher temperatures because the vapour pressure and rate of chemical reaction increases. For most solid materials, the method of manufacture and preparation can reduce the level of outgassing significantly. Cleaning surfaces or baking individual components or the entire assembly before use can drive off volatiles.

NASA's Stardust spaceprobe suffered reduced image quality due to an unknown contaminant that had condensed on the CCD sensor of the navigation camera. A similar problem affected the Cassini-Huygens spaceprobe's Narrow Angle Camera, but was corrected by repeatedly heating the system to 4 degrees Celsius.

Outgassing from rock

Outgassing is the source of many tenuous atmospheres of terrestrial planets or moons. Many materials are volatile relative to the extreme vacuum of space, such as around the

Earth's Moon, and may evaporate or even boil at ambient temperature. Materials on the lunar surface have completely outgassed and been ripped away by solar winds long ago, but volatile materials may remain at depth. Once released, gases almost always are less dense than the surrounding rocks and sand and seep toward the surface. The lunar atmosphere probably originates from outgassing of warm material below the surface. At the Earth's tectonic divergent boundaries where new crust is being created, helium and carbon dioxide are some of the volatiles being outgassed from mantle magma.

Outgassing in a closed environment

Outgassing can be significant if it collects in a closed environment where air is stagnant or recirculated. This is, for example, the origin of new car smell. Even a nearly odourless material such as wood may build up a strong smell if kept in a closed box for months. There is some concern that softeners and solvents that are released from many industrial products, especially plastics, may be harmful to human health.. Some types of RTV sealants outgas the poison cyanide for weeks after application. These outgassing poisons are of great concern in the design of submarines and space stations.

Quality

The quality of a vacuum is indicated by the amount of matter remaining in the system, so that a high quality vacuum is one with very little matter left in it. Vacuum is primarily measured by its absolute pressure, but a complete characterization requires further parameters, such as temperature and chemical composition. One of the most important parameters is the **mean free path** (MFP) of residual gases, which indicates the average distance that molecules will travel between collisions with each other. As the gas density decreases, the MFP increases, and when the MFP is longer than the chamber, pump, spacecraft, or other objects present, the continuum assumptions of fluid mechanics do not apply. This vacuum state is called *high vacuum*, and the study of fluid flows in this regime is called particle gas dynamics. The MFP of air at atmospheric pressure is very short, 70 nm, but at 100 mPa ($\sim 1 \times 10^{-3}$ Torr) the MFP of room temperature air is roughly 100 mm, which is on the order of everyday objects such as vacuum tubes. The Crookes radiometer turns when the MFP is larger than the size of the vanes.

Vacuum quality is subdivided into ranges according to the technology required to achieve it or measure it. These ranges do not have universally agreed definitions, but a typical distribution is as follows:

	pressure (Torr)	pressure (Pa)
Atmospheric pressure	760	101.3 kPa
Low vacuum	760 to 25	100 kPa to 3 kPa
Medium vacuum	25 to 1×10^{-3}	3 kPa to 100 mPa
High vacuum	1×10^{-3} to 1×10^{-9}	100 mPa to 100 nPa
Ultra high vacuum	1×10^{-9} to 1×10^{-12}	100 nPa to 100 pPa

Extremely high vacuum	$<1 \times 10^{-12}$	<100 pPa
Outer Space	1×10^{-6} to $<3 \times 10^{-17}$	100 μ Pa to <3 fPa
Perfect vacuum	0	0 Pa

- **Atmospheric pressure** is variable but standardized at 101.325 kPa (760 Torr)
- **Low vacuum**, also called *rough vacuum* or *coarse vacuum*, is vacuum that can be achieved or measured with rudimentary equipment such as a vacuum cleaner and a liquid column manometer.
- **Medium vacuum** is vacuum that can be achieved with a single pump, but the pressure is too low to measure with a liquid or mechanical manometer. It can be measured with a McLeod gauge, thermal gauge or a capacitive gauge.
- **High vacuum** is vacuum where the MFP of residual gases is longer than the size of the chamber or of the object under test. High vacuum usually requires multi-stage pumping and ion gauge measurement. Some texts differentiate between high vacuum and *very high vacuum*.
- **Ultra high vacuum** requires baking the chamber to remove trace gases, and other special procedures. British and German standards define ultra high vacuum as pressures below 10^{-6} Pa (10^{-8} Torr).
- **Deep space** is generally much more empty than any artificial vacuum. It may or may not meet the definition of high vacuum above, depending on what region of space and astronomical bodies are being considered. For example, the MFP of interplanetary space is smaller than the size of the solar system, but larger than small planets and moons. As a result, solar winds exhibit continuum flow on the scale of the solar system, but must be considered as a bombardment of particles with respect to the Earth and Moon.
- **Perfect vacuum** is an ideal state that cannot be obtained in a laboratory, nor can it be found or obtained anywhere else in the universe, apart from possibly the singularity of a black hole, or the (potentially large) spaces between atoms in lesser vacuums.
- **Hard vacuum** and **Soft vacuum** are terms that are defined with a dividing line defined differently by different sources, such as 5 psia, one Torr, or 0.1 Torr the common denominator being that a hard vacuum is a higher vacuum than a soft one.

Examples

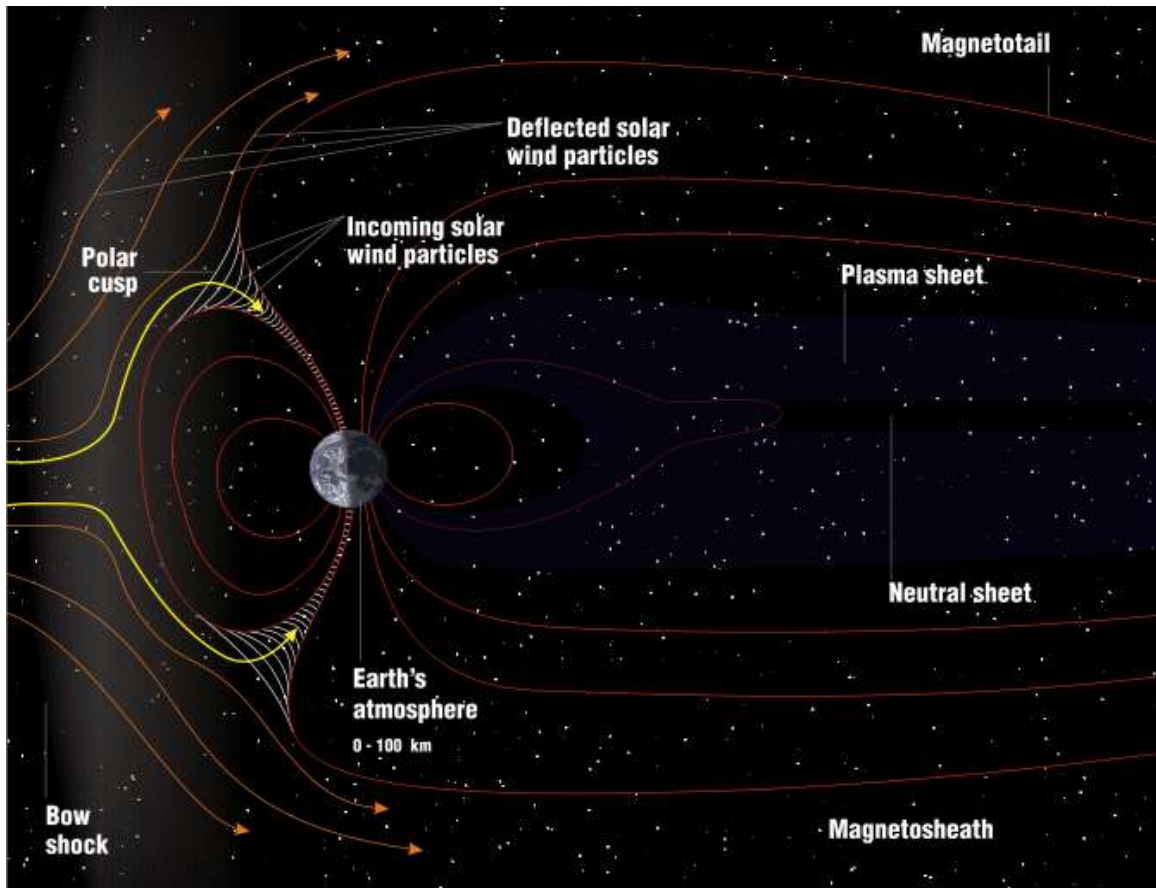
	pressure (Pa)	pressure (Torr)	mean free path	molecules per cm^3
Vacuum cleaner	approximately 80 kPa	600	70 nm	10^{19}
liquid ring vacuum pump	approximately 3.2 kPa	24		
freeze drying	100 to 10 Pa	1 to 0.1	100 μ m to 1 mm	10^{16} to 10^{15}

rotary vane pump	100 Pa to 100 mPa	1 to 10^{-3}	100 μm to 10 cm	10^{16} to 10^{13}
Incandescent light bulb	10 to 1 Pa	0.1 to 0.01	1 mm to 1 cm	10^{15} to 10^{14}
Thermos bottle	1 to 0.01 Pa	10^{-2} to 10^{-4}	1 cm to 1 m	10^{14} to 10^{12}
Earth thermosphere	1 Pa to 100 nPa	10^{-2} to 10^{-9}	1 cm to 100 km	10^{14} to 10^7
Vacuum tube	10 μPa to 10 nPa	10^{-7} to 10^{-10}		
Cryopumped MBE chamber	100 nPa to 1 nPa	10^{-9} to 10^{-11}	100 to 10,000 km	10^7 to 10^5
Pressure on the Moon	approximately 1 nPa	10^{-11}		4×10^5
Interplanetary space				10
Interstellar space				1
Intergalactic space				

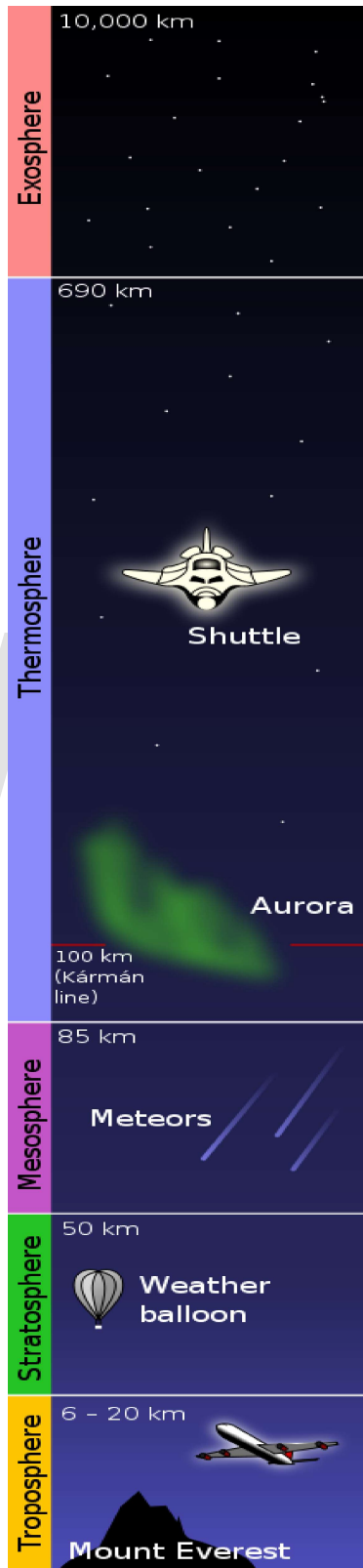


Chapter- 2

Outer Space



Outer space is not a perfect vacuum, but a tenuous plasma awash with charged particles, electromagnetic fields, and the occasional star.



The boundaries between the Earth's surface and outer space, at the Kármán line, 100 km (62 mi) and exosphere at 690 km (430 mi).

Outer space (often simply called **space**) is the void that exists beyond any celestial body including the Earth. It is not completely empty (i.e. a perfect vacuum), but contains a low density of particles, predominantly hydrogen plasma, as well as electromagnetic radiation, magnetic fields, and neutrinos. Theoretically, it also contains dark matter and dark energy.

Discovery

In 350 BC, Greek philosopher Aristotle suggested that *nature abhors a vacuum*, a principle that became known as the *horror vacui*. Based on this idea that a vacuum could not exist, it was widely held for many centuries that space could not be empty. As late as the seventeenth century, the French philosopher René Descartes argued that the entirety of space must be filled. It became known to Galileo Galilei that air had weight and so was subject to gravity. He also demonstrated that there was an established force that resisted the formation of a vacuum. However, it would remain for his pupil Evangelista Torricelli to create an apparatus that would produce a vacuum. At the time this experiment created a scientific sensation in Europe. The French mathematician Blaise Pascal reasoned that if the column of mercury was suspended by air then the column ought to be shorter at higher altitude. His brother in law, Florin Périer, repeated the experiment on the Puy-de-Dôme mountain in central France and found that the column was shorter by three inches. This decrease in pressure was further demonstrated by carrying a half-full balloon up a mountain and watching it gradually inflate, then deflate upon descent. These and other experiments were used to overthrow the principle of *horror vacui*.

Further work on the physics of the vacuum was performed by Otto von Guericke. He correctly noted that the atmosphere of the Earth surrounds the planet like a shell, with the density gradually declining with altitude. He concluded that there must be a vacuum between the Earth and the Moon.

Early speculations as to the infinite dimension of space was performed in the sixteenth century by the Italian philosopher Giordano Bruno. He extended the Copernican heliocentric cosmology to the concept of an infinite universe that is filled with a substance he called aether, which did not cause resistance to the motions of heavenly bodies. English philosopher William Gilbert arrived at a similar conclusion, arguing that the stars are visible to us only because they are surrounded by a thin aether or a void. This concept of an aether originated with ancient Greek philosophers, including Aristotle, who conceived of it as the medium through which the heavenly bodies moved.

The concept of a universe filled with a luminiferous aether remained in vogue among some scientists up until the twentieth century. This form of aether was viewed as the medium through which light could propagate. In 1887, the Michelson-Morley experiment was carried out as an attempt to detect the Earth's motion through this medium by looking for changes in the speed of light based on the direction of the planet's motion. However, the null result indicated something was wrong with the concept. Since then the idea of the luminiferous aether had essentially been abandoned, to be replaced by Albert Einstein's

theory of special relativity. The latter held that the speed of light is a constant in a vacuum, regardless of the observer's motion or frame of reference.

The first professional astronomer to support the concept of an infinite universe was the Englishman Thomas Digges in 1576. However, the true scale of the universe remained unknown until the first successful measurement of the distance to a nearby star was performed in 1838 by the German astronomer Friedrich Bessel. He showed that the star 61 Cygni had a parallax of just 0.31 arcseconds (compared to the modern value of 0.287"). This corresponded to a distance of over 10 light years. The distance scale to the Andromeda galaxy was determined in 1923 by American astronomer Edwin Hubble when he measured the brightness of cepheid variables within that galaxy. This established that the Andromeda galaxy, and by extension all galaxies, lay well outside the Milky Way.

The modern concept of outer space is based upon the Big Bang cosmology, which was first proposed in 1931 by the Belgian physicist Georges Lemaître. This theory holds that the observable universe originated from a very compact form that has since undergone continuous expansion. Matter that remained following the initial expansion has since undergone gravitational collapse to create stars, galaxies and other astronomical objects, leaving behind a deep vacuum that forms what is now called outer space.

The term *outer space* was first recorded by the English poet Lady Emmeline Stuart-Wortley in her poem "The Maiden of Moscow" in 1842, and later popularized in the writings of HG Wells in 1901. The shorter term *space* is actually older, first used to mean the region beyond Earth's sky in John Milton's *Paradise Lost* in 1667.

Environment

Outer space is the closest natural approximation of a perfect vacuum. It has effectively no friction, allowing stars, planets and moons to move freely along ideal gravitational trajectories. However, even in the deep vacuum of intergalactic space there are still a few hydrogen atoms per cubic meter. By comparison, the air we breathe contains about 10^{25} molecules per cubic meter. The sparse density of matter in outer space means that electromagnetic radiation can travel great distances without being scattered; the mean free path for a photon in intergalactic space is about 10^{23} km, or 10 billion light years. The deep vacuum of space could make it an attractive environment for certain industrial processes, for instance those that require ultraclean surfaces.

Stars, planets, asteroids and moons keep their atmospheres by gravitational attraction, and as such, atmospheres have no clearly delineated boundary: the density of atmospheric gas simply decreases with distance from the object. The Earth's atmospheric pressure drops to about 1 Pa at 100 kilometres (62 mi) of altitude. This is known as the Kármán line, a common definition of the boundary with outer space. Beyond this line, isotropic gas pressure rapidly becomes insignificant when compared to radiation pressure from the Sun and the dynamic pressure of the solar wind, so the definition of pressure becomes difficult to interpret. The thermosphere in this range has large gradients of pressure,

temperature and composition, and varies greatly due to space weather. Astrophysicists prefer to use number density to describe these environments, in units of particles per cubic centimeter.

Temperature

All of the observable Universe is filled with photons that were created during the Big Bang, which is known as the cosmic microwave background radiation (CMB). There is quite likely a correspondingly large number of neutrinos called the cosmic neutrino background. The current black body temperature of this photon radiation is about 3 K ($-270\text{ }^{\circ}\text{C}$; $-454\text{ }^{\circ}\text{F}$). Some regions of outer space can contain highly energetic particles that have a much higher temperature than the CMB.

Effect on human bodies

Contrary to popular belief, a person suddenly exposed to the vacuum would not explode, freeze to death or die from boiling blood. Air would immediately leave the lungs due to the enormous pressure gradient. Any oxygen dissolved in the blood would empty into the lungs to try to equalize the partial pressure gradient. Once the deoxygenated blood arrived at the brain, death would quickly follow.

Humans and animals exposed to vacuum will lose consciousness after a few seconds and die of hypoxia within minutes. Blood and other body fluids do boil when their pressure drops below 6.3 kPa, the vapor pressure of water at body temperature. This condition is called ebullism. The steam may bloat the body to twice its normal size and slow circulation, but tissues are elastic and porous enough to prevent rupture. Ebullism is slowed by the pressure containment of blood vessels, so some blood remains liquid. Swelling and ebullism can be reduced by containment in a flight suit. Shuttle astronauts wear a fitted elastic garment called the Crew Altitude Protection Suit (CAPS) which prevents ebullism at pressures as low as 2 kPa. Water vapor would also rapidly evaporate off from exposed areas such as the lungs, cornea of the eye and mouth, cooling the body. Rapid evaporative cooling of the skin will create frost, particularly in the mouth, but this is not a significant hazard. Space may be cold, but it's mostly vacuum and transfers heat ineffectively; as a result the main temperature regulation concern for space suits is how to get rid of naturally generated body heat.

Cold or oxygen-rich atmospheres can sustain life at pressures much lower than atmospheric, as long as the density of oxygen is similar to that of standard sea-level atmosphere. The colder air temperatures found at altitudes of up to 3 kilometres (1.9 mi) generally compensate for the lower pressures there. Above this altitude, oxygen enrichment is necessary to prevent altitude sickness, and spacesuits are necessary to prevent ebullism above 19 kilometres (12 mi). Most spacesuits use only 20 kPa of pure oxygen, just enough to sustain full consciousness. This pressure is high enough to prevent ebullism, but simple evaporation of blood can still cause decompression sickness and gas embolisms if not managed.

Rapid decompression can be much more dangerous than vacuum exposure itself. Even if the victim does not hold his breath, venting through the windpipe may be too slow to prevent the fatal rupture of the delicate alveoli of the lungs. Eardrums and sinuses may be ruptured by rapid decompression, soft tissues may bruise and seep blood, and the stress of shock will accelerate oxygen consumption leading to hypoxia. Injuries caused by rapid decompression are called barotrauma. A pressure drop as small as 13 kPa, which produces no symptoms if it is gradual, may be fatal if it occurs suddenly.

Boundary

There is no clear boundary between Earth's atmosphere and space, as the density of the atmosphere gradually decreases as the altitude increases. There are several designated scientific boundaries, namely:

- The Fédération Aéronautique Internationale has established the Kármán line at an altitude of 100 kilometres (62 mi) as a working definition for the boundary between aeronautics and astronautics. This is used because above an altitude of roughly 100 km, as Theodore von Kármán calculated, a vehicle would have to travel faster than orbital velocity in order to derive sufficient aerodynamic lift from the atmosphere to support itself.
- The United States designates people who travel above an altitude of 50 miles (80 km) as astronauts.
- NASA's mission control uses 76 miles (122 km) as their re-entry altitude, which roughly marks the boundary where atmospheric drag becomes noticeable (depending on the ballistic coefficient of the vehicle), thus leading shuttles to switch from steering with thrusters to maneuvering with air surfaces.

In 2009, scientists at the University of Calgary reported detailed measurements with an instrument called the Supra-Thermal Ion Imager (an instrument that measures the direction and speed of ions), which allowed them to determine that space begins 118 kilometres (73 mi) above Earth. The boundary represents the midpoint of a gradual transition over tens of kilometers from the relatively gentle winds of the Earth's atmosphere to the more violent flows of charged particles in space, which can reach speeds well over 600 miles per hour (1,000 km/h).

This was only the second time that direct measurements of charged particle flows have been conducted at this region, which is too high for balloons and too low for satellites. It was however the first study to include all the relevant elements for this kind of determination – for example, the upper atmospheric winds.

The instrument was carried by the JOULE-II rocket on January 19, 2007, and traveled to an altitude of about 124 miles (200 km). From there it collected data while it was moving through the "edge of space".

Legal status

The Outer Space Treaty provides the basic framework for international space law. This treaty covers the legal use of outer space by nation states, and includes in its definition of *outer space* the Moon and other celestial bodies. The treaty states that outer space is free for all nation states to explore and is not subject to claims of national sovereignty. It also prohibits the deployment of nuclear weapons in outer space. The treaty was passed by the United Nations General Assembly in 1963 and signed in 1967 by the USSR, the United States of America and the United Kingdom. As of January 1, 2008 the treaty has been signed by 98 states and ratified by an additional 27 states.

Between 1958 and 2008, outer space has been the subject of multiple resolutions by the United Nations General Assembly. Of these, more than 50 have been concerning the international co-operation in the peaceful uses of outer space and preventing an arms race in space. Four additional space law treaties have been negotiated and drafted by the UN's Committee on the Peaceful Uses of Outer Space. The 1979 Moon Treaty turned the jurisdiction of all heavenly bodies (including the orbits around such bodies) over to the international community. However, this treaty has not been ratified by any nation that currently practices manned spaceflight.

Space versus orbit

To perform an orbit, a spacecraft must travel faster than a sub-orbital spaceflight. A spacecraft has not entered orbit until it is traveling with a sufficiently great horizontal velocity such that the acceleration due to gravity on the spacecraft is less than or equal to the centripetal acceleration being caused by its horizontal velocity. So to enter orbit, a spacecraft must not only reach space, but must also achieve a sufficient orbital speed (angular velocity). For a low-Earth orbit, this is about 7,900 m/s (28,400 km/h; 17,700 mph); by contrast, the fastest airplane speed ever achieved (excluding speeds achieved by deorbiting spacecraft) was 2,200 m/s (7,900 km/h; 4,900 mph) in 1967 by the North American X-15.

Konstantin Tsiolkovsky was the first person to realize that, given the energy available from any available chemical fuel, a several-stage rocket would be required. The escape velocity to pull free of Earth's gravitational field altogether and move into interplanetary space is about 11,000 m/s (39,600 km/h; 24,600 mph) The energy required to reach velocity for low Earth orbit (32 MJ/kg) is about twenty times the energy required simply to climb to the corresponding altitude (10 kJ/(km·kg)).

There is a major difference between sub-orbital and orbital spaceflights. The minimum altitude for a stable orbit around Earth (that is, one without significant atmospheric drag) begins at around 350 kilometres (220 mi) above mean sea level. A common misunderstanding about the boundary to space is that orbit occurs simply by reaching this altitude. Achieving orbital speed can theoretically occur at any altitude, although atmospheric drag precludes an orbit that is too low. At sufficient speed, an airplane would

need a way to keep it from flying off into space, but at present, this speed is several times greater than anything within reasonable technology.

A common misconception is that people in orbit are outside Earth's gravity because they are "floating". They are floating because they are in "free fall": they are accelerating toward Earth, along with their spacecraft, but are simultaneously moving sideways fast enough that the "fall" away from a straight-line path merely keeps them in orbit at a constant distance above Earth's surface. Earth's gravity reaches out far past the Van Allen belt and keeps the Moon in orbit at an average distance of 384,403 kilometres (238,857 mi).

Regions

Space is not a perfect vacuum: its different regions are defined by the various atmospheres and "winds" that dominate within them, and extend to the point at which those winds give way to those beyond. Geospace extends from Earth's atmosphere to the outer reaches of Earth's magnetic field, whereupon it gives way to the solar wind of interplanetary space. Interplanetary space extends to the heliopause, whereupon the solar wind gives way to the winds of the interstellar medium. Interstellar space then continues to the edges of the galaxy, where it fades into the intergalactic void.

Geospace



Aurora australis observed by *Discovery*, on STS-39, May 1991 (orbital altitude: 260 km)

Geospace is the region of outer space near the Earth. Geospace includes the upper region of the atmosphere, as well as the ionosphere and magnetosphere. The Van Allen radiation belts also lie within the geospace. The region between Earth's atmosphere and the Moon is sometimes referred to as **cis-lunar space**.

Although it meets the definition of outer space, the atmospheric density within the first few hundred kilometers above the Kármán line is still sufficient to produce significant drag on satellites. Most artificial satellites operate in this region called low earth orbit and must fire their engines every few days to maintain orbit. The drag here is low enough that it could theoretically be overcome by radiation pressure on solar sails, a proposed propulsion system for interplanetary travel.

Geospace is populated by electrically charged particles at very low densities, the motions of which are controlled by the Earth's magnetic field. These plasmas form a medium from which storm-like disturbances powered by the solar wind can drive electrical currents into the Earth's upper atmosphere.

During geomagnetic storms two regions of geospace, the radiation belts and the ionosphere, can become strongly disturbed. These storms increase fluxes of energetic electrons that can permanently damage satellite electronics, disrupting telecommunications and GPS technologies, and can also be a hazard to astronauts, even in low-Earth orbit. They also create aurorae seen near the magnetic poles.

Geospace contains material left over from previous manned and unmanned launches that are a potential hazard to spacecraft. Some of this debris re-enters Earth's atmosphere periodically.

The absence of air makes geospace (and the surface of the Moon) ideal locations for astronomy at all wavelengths of the electromagnetic spectrum, as evidenced by the spectacular pictures sent back by the Hubble Space Telescope, allowing light from about 13.7 billion years ago — almost to the time of the Big Bang — to be observed.

The outer boundary of geospace is the interface between the magnetosphere and the solar wind. The inner boundary is the ionosphere. Alternately, geospace is the region of space between the Earth's upper atmosphere and the outermost reaches of the Earth's magnetic field.

Interplanetary

Interplanetary space, the space around the Sun and planets of the Solar System, is the region dominated by the interplanetary medium, which extends out to the heliopause where the influence of the galactic environment starts to dominate over the magnetic field and particle flux from the Sun. Interplanetary space is defined by the solar wind, a continuous stream of charged particles emanating from the Sun that creates a very tenuous atmosphere (the heliosphere) for billions of miles into space. This wind has a particle density of 5–10 protons/cm³ and is moving at a velocity of 350–400 km/s. The

distance and strength of the heliopause varies depending on the activity level of the solar wind. The discovery since 1995 of extrasolar planets means that other stars must possess their own interplanetary media.

The volume of interplanetary space is an almost pure vacuum, with a mean free path of about one astronomical unit at the orbital distance of the Earth. However, this space is not completely empty, and is sparsely filled with cosmic rays, which include ionized atomic nuclei and various subatomic particles. There is also gas, plasma and dust, small meteors, and several dozen types of organic molecules discovered to date by microwave spectroscopy.

Interplanetary space contains the magnetic field generated by the Sun. There are also magnetospheres generated by planets such as Jupiter, Saturn and the Earth that have their own magnetic fields. These are shaped by the influence of the solar wind into the approximation of a teardrop shape, with the long tail extending outward behind the planet. These magnetic fields can trap particles from the solar wind and other sources, creating belts of magnetic particles such as the Van Allen Belts. Planets without magnetic fields, such as Mars and Mercury, but excluding Venus, have their atmospheres gradually eroded by the solar wind.

Interstellar

Interstellar space is the physical space within a galaxy not occupied by stars or their planetary systems. The interstellar medium resides—by definition—in interstellar space.

Intergalactic

Intergalactic space is the physical space between galaxies. Generally free of dust and debris, intergalactic space is very close to a total vacuum. The space between galaxy clusters, called the voids, is probably nearly empty. Some theories put the average density of the Universe as the equivalent of one hydrogen atom per cubic meter. The density of the universe, however, is clearly not uniform; it ranges from relatively high density in galaxies (including very high density in structures within galaxies, such as planets, stars, and black holes) to conditions in vast voids that have much lower density than the universe's average.

Surrounding and stretching between galaxies, there is a rarefied plasma that is thought to possess a cosmic filamentary structure and that is slightly denser than the average density in the universe. This material is called the *intergalactic medium (IGM)* and is mostly ionized hydrogen; i.e. a plasma consisting of equal numbers of electrons and protons. The IGM is thought to exist at a density of 10 to 100 times the average density of the universe (10 to 100 hydrogen atoms per cubic meter). It reaches densities as high as 1000 times the average density of the universe in rich clusters of galaxies.

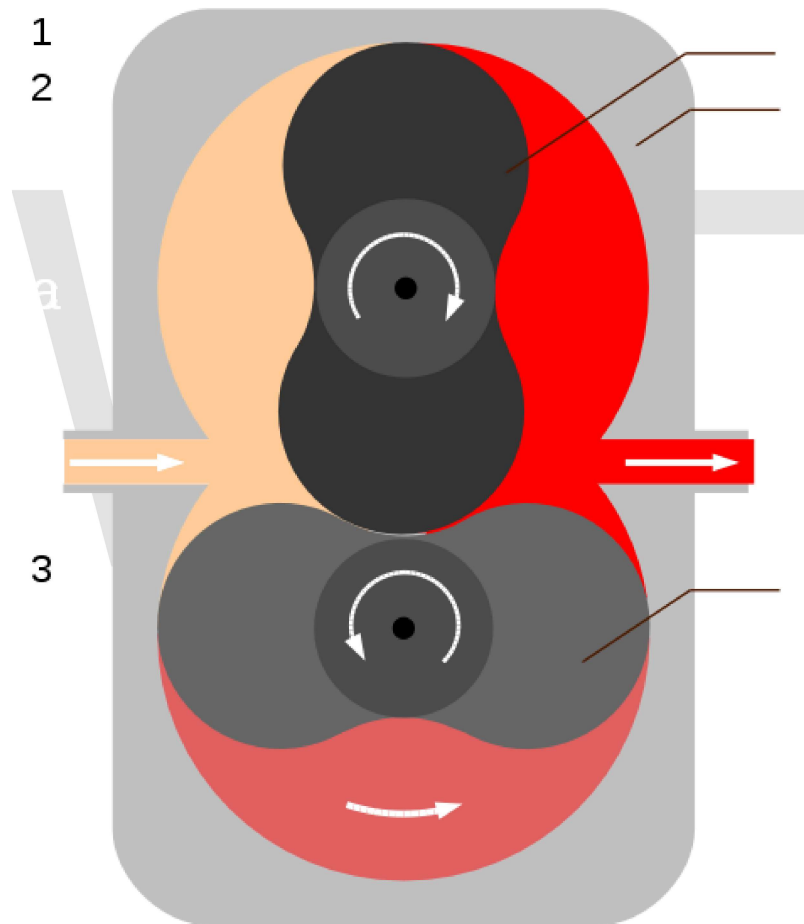
The reason the IGM is thought to be mostly ionized gas is that its temperature is thought to be quite high by terrestrial standards (though some parts of it are only "warm" by

astrophysical standards). As gas falls into the Intergalactic Medium from the voids, it heats up to temperatures of 10^5 K to 10^7 K, which is high enough for the bound electrons to escape from the hydrogen nuclei upon collisions. At these temperatures, it is called the Warm-Hot Intergalactic Medium (WHIM). Computer simulations indicate that on the order of half the atomic matter in the universe might exist in this warm-hot, rarefied state. When gas falls from the filamentary structures of the WHIM into the galaxy clusters at the intersections of the cosmic filaments, it can heat up even more, reaching temperatures of 10^8 K and above.

WWT

Chapter- 3

Vacuum Pump



The Roots blower is one example of a vacuum pump

A **vacuum pump** is a device that removes gas molecules from a sealed volume in order to leave behind a partial vacuum. The vacuum pump was invented in 1650 by Otto von Guericke.

Types

Pumps can be broadly categorized according to three techniques:

- **Positive displacement** pumps use a mechanism to repeatedly expand a cavity, allow gases to flow in from the chamber, seal off the cavity, and exhaust it to the atmosphere.
- **Momentum transfer** pumps, also called **molecular pumps**, use high speed jets of dense fluid or high speed rotating blades to knock gas molecules out of the chamber.
- **Entrapment** pumps capture gases in a solid or adsorbed state. This includes **cryopumps**, **getters**, and **ion pumps**.

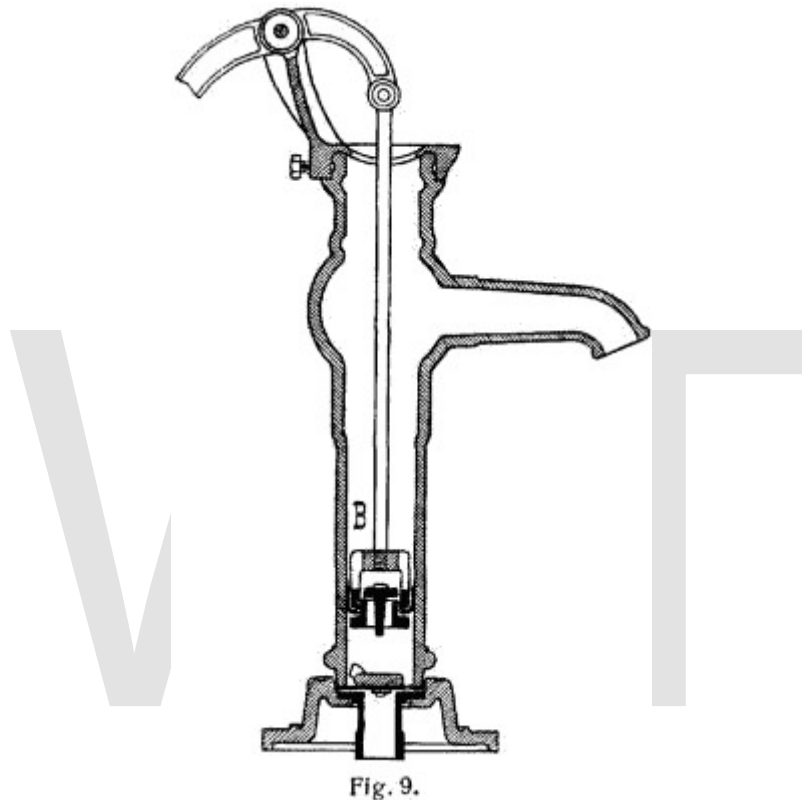
Positive displacement pumps are the most effective for low vacuums. Momentum transfer pumps in conjunction with one or two positive displacement pumps are the most common configuration used to achieve high vacuums. In this configuration the positive displacement pump serves two purposes. First it obtains a rough vacuum in the vessel being evacuated before the momentum transfer pump can be used to obtain the high vacuum, as momentum transfer pumps cannot start pumping at atmospheric pressures. Second the positive displacement pump backs up the momentum transfer pump by evacuating to low vacuum the accumulation of displaced molecules in the high vacuum pump. Entrapment pumps can be added to reach ultrahigh vacuums, but they require periodic regeneration of the surfaces that trap air molecules or ions. Due to this requirement their available operational time can be unacceptably short in low and high vacuums, thus limiting their use to ultrahigh vacuums. Pumps also differ in details like manufacturing tolerances, sealing material, pressure, flow, admission or no admission of oil vapor, service intervals, reliability, tolerance to dust, tolerance to chemicals, tolerance to liquids and vibration.

Performance measures

- **Pumping speed** refers to the volume flow rate of a pump at its inlet, often measured in volume per unit of time. Momentum transfer and entrapment pumps are more effective on some gases than others, so the pumping rate can be different for each of the gases being pumped, and the average volume flow rate of the pump will vary depending on the chemical composition of the gases remaining in the chamber.
- **Throughput** refers to the pumping speed multiplied by the gas pressure at the inlet, and is measured in units of pressure·volume/unit time. At a constant temperature, throughput is proportional to the number of molecules being pumped per unit time, and therefore to the mass flow rate of the pump. When discussing a leak in the system or backstreaming through the pump, **throughput** refers to the volume leak rate multiplied by the pressure at the vacuum side of the leak, so the leak throughput can be compared to the pump throughput.

Positive displacement and momentum transfer pumps have a constant volume flow rate, (pumping speed,) but as the chamber's pressure drops, this volume contains less and less mass. So although the pumping speed remains constant, the throughput and mass flow rate drop exponentially. Meanwhile, the leakage, evaporation, sublimation and backstreaming rates continue to produce a constant throughput into the system.

Positive displacement



The manual water pump draws water up from a well by creating a vacuum that water rushes in to fill. In a sense, it acts to evacuate the well, although the high leakage rate of dirt prevents a high quality vacuum from being maintained for any length of time.



Mechanism of a scroll pump

Fluids cannot be pulled, so it is technically impossible to create a vacuum by suction. Suction is the movement of fluids into a vacuum under the effect of a higher external pressure, but the vacuum has to be created first. The easiest way to create an artificial vacuum is to expand the volume of a container. For example, the diaphragm muscle expands the chest cavity, which causes the volume of the lungs to increase. This expansion reduces the pressure and creates a partial vacuum, which is soon filled by air pushed in by atmospheric pressure

To continue evacuating a chamber indefinitely without requiring infinite growth, a compartment of the vacuum can be repeatedly closed off, exhausted, and expanded again. This is the principle behind *positive displacement pumps*, like the manual water pump for example. Inside the pump, a mechanism expands a small sealed cavity to create a deep vacuum. Because of the pressure differential, some fluid from the chamber (or the well, in our example) is pushed into the pump's small cavity. The pump's cavity is then sealed from the chamber, opened to the atmosphere, and squeezed back to a minute size.

More sophisticated systems are used for most industrial applications, but the basic principle of cyclic volume removal is the same:

- Rotary vane pump, the most common
- Diaphragm pump, zero oil contamination
- Liquid ring pump
- Piston pump, cheapest
- Scroll pump, highest speed dry pump
- Screw pump (10 Pa)
- Wankel pump
- External vane pump
- Roots blower, also called a **booster pump**, has highest pumping speeds but low compression ratio
- Multistage Roots pump that combine several stages providing high pumping speed with better compression ratio
- Toepler pump
- Lobe pump

The base pressure of a rubber- and plastic-sealed piston pump system is typically 1 to 50 kPa, while a scroll pump might reach 10 Pa (when new) and a rotary vane oil pump with a clean and empty metallic chamber can easily achieve 0.1 Pa.

A positive displacement vacuum pump moves the same volume of gas with each cycle, so its pumping speed is constant unless it is overcome by backstreaming.

Momentum transfer



A cutaway view of a turbomolecular high vacuum pump

In a **momentum transfer** pump, gas molecules are accelerated from the vacuum side to the exhaust side (which is usually maintained at a reduced pressure by a positive displacement pump). Momentum transfer pumping is only possible below pressures of about 0.1 kPa. Matter flows differently at different pressures based on the laws of fluid dynamics. At atmospheric pressure and mild vacuums, molecules interact with each other and push on their neighboring molecules in what is known as viscous flow. When the distance between the molecules increases, the molecules interact with the walls of the chamber more often than the other molecules, and **molecular pumping** becomes more effective than positive displacement pumping. This regime is generally called **high vacuum**.

Molecular pumps sweep out a larger area than mechanical pumps, and do so more frequently, making them capable of much higher pumping speeds. They do this at the expense of the seal between the vacuum and their exhaust. Since there is no seal, a small

pressure at the exhaust can easily cause backstreaming through the pump; this is called **stall**. In high vacuum, however, pressure gradients have little effect on fluid flows, and molecular pumps can attain their full potential.

The two main types of molecular pumps are the diffusion pump and the turbomolecular pump. Both types of pumps blow out gas molecules that diffuse into the pump by imparting momentum to the gas molecules. Diffusion pumps blow out gas molecules with jets of oil or mercury, while turbomolecular pumps use high speed fans to push the gas. Both of these pumps will stall and fail to pump if exhausted directly to atmospheric pressure, so they must be exhausted to a lower grade vacuum created by a mechanical pump.

As with positive displacement pumps, the base pressure will be reached when leakage, outgassing, and backstreaming equal the pump speed, but now minimizing leakage and outgassing to a level comparable to backstreaming becomes much more difficult.

- Diffusion pump
- Turbomolecular pump

Entrapment

Entrapment pumps may be cryopumps, which use cold temperatures to condense gases to a solid or adsorbed state, **chemical** pumps, which react with gases to produce a solid residue, or ionization pumps, which use strong electrical fields to ionize gases and propel the ions into a solid substrate. A cryomodule uses cryopumping.

- Ion pump
- Cryopump
- Sorption pump
- Non-evaporative getter

Other pump types

- Venturi vacuum pump (aspirator) (10 to 30 kPa)
- Steam ejector (vacuum depends on the number of stages, but can be very low)

Techniques

Vacuum pumps are combined with chambers and operational procedures into a wide variety of vacuum systems. Sometimes more than one pump will be used (in series or in parallel) in a single application. A partial vacuum, or rough vacuum, can be created using a positive displacement pump that transports a gas load from an inlet port to an outlet (exhaust) port. Because of their mechanical limitations, such pumps can only achieve a low vacuum. To achieve a higher vacuum, other techniques must then be used, typically in series (usually following an initial fast pump down with a positive displacement

pump). Some examples might be use of an oil sealed rotary vane pump (the most common positive displacement pump) backing a diffusion pump, or a dry scroll pump backing a turbomolecular pump. There are other combinations depending on the level of vacuum being sought.

Achieving high vacuum is difficult because all of the materials exposed to the vacuum must be carefully evaluated for their outgassing and vapor pressure properties. For example, oils, and greases, and rubber, or plastic gaskets used as seals for the vacuum chamber must not boil off when exposed to the vacuum, or the gases they produce would prevent the creation of the desired degree of vacuum. Often, all of the surfaces exposed to the vacuum must be baked at high temperature to drive off adsorbed gases.

Outgassing can also be reduced simply by desiccation prior to vacuum pumping. High vacuum systems generally require metal chambers with metal gasket seals such as Klein flanges or ISO flanges, rather than the rubber gaskets more common in low vacuum chamber seals. The system must be clean and free of organic matter to minimize outgassing. All materials, solid or liquid, have a small vapour pressure, and their outgassing becomes important when the vacuum pressure falls below this vapour pressure. As a result, many materials that work well in low vacuums, such as epoxy, will become a source of outgassing at higher vacuums. With these standard precautions, vacuums of 1 mPa are easily achieved with an assortment of molecular pumps. With careful design and operation, 1 μ Pa is possible.

Several types of pumps may be used in sequence or in parallel. In a typical pumpdown sequence, a positive displacement pump would be used to remove most of the gas from a chamber, starting from atmosphere (760 Torr, 101 kPa) to 25 Torr (3 kPa). Then a sorption pump would be used to bring the pressure down to 10^{-4} Torr (10 mPa). A cryopump or turbomolecular pump would be used to bring the pressure further down to 10^{-8} Torr (1 μ Pa). An additional ion pump can be started below 10^{-6} Torr to remove gases which are not adequately handled by a cryopump or turbo pump, such as helium or hydrogen.

Ultra high vacuum generally requires custom-built equipment, strict operational procedures, and a fair amount of trial-and-error. Ultra-high vacuum systems are usually made of stainless steel with metal-gasketed conflat flanges. The system is usually baked, preferably under vacuum, to temporarily raise the vapour pressure of all outgassing materials in the system and boil them off. If necessary, this outgassing of the system can also be performed at room temperature, but this takes much more time. Once the bulk of the outgassing materials are boiled off and evacuated, the system may be cooled to lower vapour pressures to minimize residual outgassing during actual operation. Some systems are cooled well below room temperature by liquid nitrogen to shut down residual outgassing and simultaneously cryopump the system.

In ultra-high vacuum systems, some very odd leakage paths and outgassing sources must be considered. The water absorption of aluminium and palladium becomes an unacceptable source of outgassing, and even the absorptivity of hard metals such as

stainless steel or titanium must be considered. Some oils and greases will boil off in extreme vacuums. The porosity of the metallic chamber walls may have to be considered, and the grain direction of the metallic flanges should be parallel to the flange face.

The impact of molecular size must be considered. Smaller molecules can leak in more easily and are more easily absorbed by certain materials, and molecular pumps are less effective at pumping gases with lower molecular weights. A system may be able to evacuate nitrogen (the main component of air) to the desired vacuum, but the chamber could still be full of residual atmospheric hydrogen and helium. Vessels lined with a highly gas-permeable material such as palladium (which is a high-capacity hydrogen sponge) create special outgassing problems.

Uses of vacuum pumps

Vacuum pumps are used in many industrial and scientific processes including:

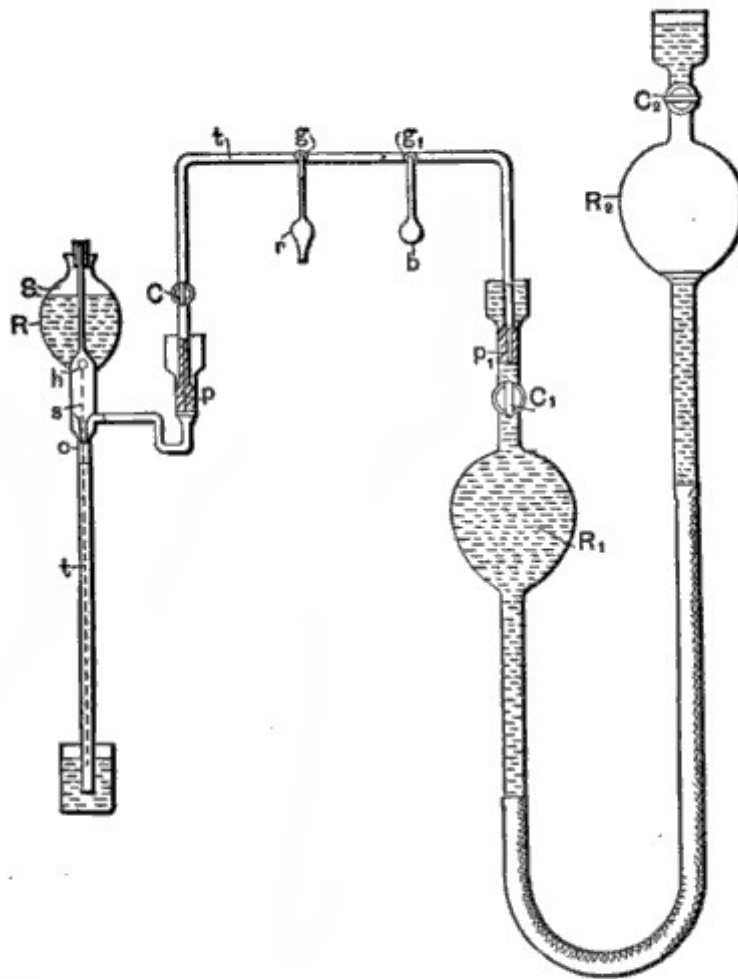
- Composite Plastic moulding processes (VRTM)
- Driving some of the flight instruments in most aircraft.
- The production of most types of electric lamps, vacuum tubes, and CRTs where the device is either left evacuated or re-filled with a specific gas or gas mixture
- Semiconductor processing, notably ion implantation, dry etch and PVD, ALD, PECVD and CVD deposition and soon in photolithography
- Electron microscopy
- Medical processes that require suction
- Medical applications such as such Radiotherapy, Radiosurgery, Radiopharmacy
- Analytical instrumentation to analyse gas, liquid, solid, surface and bio materials
- Mass spectrometers to create an ultra high vacuum between the ion source and the detector
- Vacuum Coating for decoration, for durability, for energy saving
- Glass coating for low e glass
- Hard coating for engine (as in Formula One)
- Ophthalmic coating
- Milking Machines
- Air conditioning service - removing all contaminants from the system before charging with refrigerant
- Trash compactor
- Vacuum engineering
- Sewage systems
- Freeze Drying
- As the main source of vacuum in dairy shed plant

Vacuum may be used to power, or provide assistance to mechanical devices. In diesel engined motor vehicles, a pump fitted on the engine (usually on the camshaft) is used to produce vacuum. In petrol engines, instead, vacuum is obtained as a side-effect of the operation of the engine and the flow restriction created by the throttle plate. This vacuum may then be used to power:

- The vacuum servo booster for the hydraulic brakes
- Motors that move dampers in the ventilation system
- The throttle driver in the cruise control servomechanism

In an aircraft, the vacuum source is often used to power gyroscopes in the various flight instruments. To prevent the complete loss of instrumentation in the event of an electrical failure, the instrument panel is deliberately designed with certain instruments powered by electricity and other instruments powered by the vacuum source.

History of the vacuum pump



APPARATUS USED FOR OBTAINING HIGH DEGREES OF EXHAUSTION.

The predecessor to the vacuum pump was the suction pump, which was known to the romans. Dual-action suction pumps were found in the city of Pompeii. Arabic engineer Al-Jazari also described suction pumps in the 13th century. He said that his model was a

larger version of the siphons the Bizantines used to discharge the Greek fire. The suction pump later reappeared in Europe from the 15th century.

By the 17th century, water pump designs had improved to the point that they produced measurable vacuums, but this was not immediately understood. What was known was that suction pumps could not pull water beyond a certain height: 18 Florentine yards according to a measurement taken around 1635. (The conversion to metres is uncertain, but it would be about 9 or 10 metres.) This limit was a concern to irrigation projects, mine drainage, and decorative water fountains planned by the Duke of Tuscany, so the Duke commissioned Galileo to investigate the problem. Galileo advertised the puzzle to other scientists, including Gaspar Berti who replicated it by building the first water barometer in Rome in 1639. Berti's barometer produced a vacuum above the water column, but he could not explain it. The breakthrough was made by Evangelista Torricelli in 1643. Building upon Galileo's notes, he built the first mercury barometer and wrote a convincing argument that the space at the top was a vacuum. The height of the column was then limited to the maximum weight that atmospheric pressure could support. Some people believe that although Torricelli's experiment was crucial, it was Blaise Pascal's experiments that proved the top space really contained vacuum.

In 1654, Otto von Guericke invented the first vacuum pump and conducted his famous Magdeburg hemispheres experiment, showing that teams of horses could not separate two hemispheres from which the air had been evacuated. Robert Boyle improved Guericke's design and conducted experiments on the properties of vacuum. Robert Hooke also helped Boyle produce an air pump which helped to produce the vacuum. The study of vacuum then lapsed until 1855, when Heinrich Geissler invented the mercury displacement pump and achieved a record vacuum of about 10 Pa (0.1 Torr). A number of electrical properties become observable at this vacuum level, and this renewed interest in vacuum. This, in turn, led to the development of the vacuum tube.

In the 19th century, Nikola Tesla designed the apparatus, imaged to the right, that contains a Sprengel Pump to create a high degree of exhaustion.

Hazards



Student of Smolny Institute Catherine Molchanova with vacuum pump, by Dmitry Levitzky, 1776

Old vacuum-pump oils that were produced before circa 1980 often contain a mixture of several different dangerous polychlorinated biphenyls (PCBs), which are highly toxic, carcinogenic, persistent organic pollutants.

Chapter- 4

Pressure Measurement



The construction of a bourdon tube gauge, construction elements are made of brass

Many techniques have been developed for the measurement of pressure and vacuum. Instruments used to measure pressure are called **pressure gauges** or **vacuum gauges**.

A **manometer** could also be referring to a pressure measuring instrument, usually limited to measuring pressures near to atmospheric. The term *manometer* is often used to refer specifically to liquid column hydrostatic instruments.

A **vacuum gauge** is used to measure the pressure in a vacuum—which is further divided into two subcategories: high and low vacuum (and sometimes ultra-high vacuum). The applicable pressure range of many of the techniques used to measure vacuums have an overlap. Hence, by combining several different types of gauge, it is possible to measure system pressure continuously from 10 mbar down to 10^{-11} mbar.

Absolute, gauge and differential pressures - zero reference

Although pressure is an absolute quantity, everyday pressure measurements, such as for tire pressure, are usually made relative to ambient air pressure. In other cases measurements are made relative to a vacuum or to some other ad hoc reference. When distinguishing between these zero references, the following terms are used:

- **Absolute pressure** is zero referenced against a perfect vacuum, so it is equal to gauge pressure plus atmospheric pressure.
- **Gauge pressure** is zero referenced against ambient air pressure, so it is equal to absolute pressure minus atmospheric pressure. Negative signs are usually omitted.
- **Differential pressure** is the difference in pressure between two points.

The zero reference in use is usually implied by context, and these words are only added when clarification is needed. Tire pressure and blood pressure are gauge pressures by convention, while atmospheric pressures, deep vacuum pressures, and altimeter pressures must be absolute. Differential pressures are commonly used in industrial process systems. Differential pressure gauges have two inlet ports, each connected to one of the volumes whose pressure is to be monitored. In effect, such a gauge performs the mathematical operation of subtraction through mechanical means, obviating the need for an operator or control system to watch two separate gauges and determine the difference in readings. Moderate vacuum pressures are often ambiguous, as they may represent absolute pressure or gauge pressure without a negative sign. Thus a vacuum of 26 inHg gauge is equivalent to an absolute pressure of 30 inHg (typical atmospheric pressure) $- 26 \text{ inHg} = 4 \text{ inHg}$.

Atmospheric pressure is typically about 100 kPa at sea level, but is variable with altitude and weather. If the absolute pressure of a fluid stays constant, the gauge pressure of the same fluid will vary as atmospheric pressure changes. For example, when a car drives up a mountain (atmospheric air pressure decreases), the (gauge) tire pressure goes up. Some standard values of atmospheric pressure such as 101.325 kPa or 100 kPa have been defined, and some instruments use one of these standard values as a constant zero reference instead of the actual variable ambient air pressure. This impairs the accuracy of these instruments, especially when used at high altitudes.

Use of the atmosphere as reference is usually signified by a (g) after the pressure unit e.g. 30 psi g, which means that the pressure measured is the total pressure minus atmospheric pressure. There are two types of gauge reference pressure: vented gauge (vg) and sealed gauge (sg).

A vented gauge pressure transmitter for example allows the outside air pressure to be exposed to the negative side of the pressure sensing diaphragm, via a vented cable or a hole on the side of the device, so that it always measures the pressure referred to ambient barometric pressure. Thus a vented gauge reference pressure sensor should always read zero pressure when the process pressure connection is held open to the air.

A sealed gauge reference is very similar except that atmospheric pressure is sealed on the negative side of the diaphragm. This is usually adopted on high pressure ranges such as hydraulics where atmospheric pressure changes will have a negligible effect on the accuracy of the reading, so venting is not necessary. This also allows some manufacturers to provide secondary pressure containment as an extra precaution for pressure equipment safety if the burst pressure of the primary pressure sensing diaphragm is exceeded.

There is another way of creating a sealed gauge reference and this is to seal a high vacuum on the reverse side of the sensing diaphragm. Then the output signal is offset so the pressure sensor reads close to zero when measuring atmospheric pressure.

A sealed gauge reference pressure transducer will never read exactly zero because atmospheric pressure is always changing and the reference in this case is fixed at 1 bar.

An absolute pressure measurement is one that is referred to absolute vacuum. The best example of an absolute referenced pressure is atmospheric or barometric pressure.

To produce an absolute pressure sensor the manufacturer will seal a high vacuum behind the sensing diaphragm. If the process pressure connection of an absolute pressure transmitter is open to the air, it will read the actual barometric pressure.

Units

Pressure units						
	Pascal (Pa)	Bar (bar)	Technical atmosphere (at)	Atmosphere (atm)	Torr (Torr)	Pound- force per square inch (psi)
1 Pa	$\equiv 1 \text{ N/m}^2$	10^{-5}	1.0197×10^{-5}	9.8692×10^{-6}	7.5006×10^{-3}	145.04×10^{-6}
1 bar	100,000	$\equiv 10^6 \text{ dyn/cm}^2$	1.0197	0.98692	750.06	14.5037744
1 at	98,066.5	0.980665	$\equiv 1 \text{ kgf/cm}^2$	0.96784	735.56	14.223
1 atm	101,325	1.01325	1.0332	$\equiv 1 \text{ atm}$	760	14.696
1 torr	133.322	1.3332×10^{-3}	1.3595×10^{-3}	1.3158×10^{-3}	$\equiv 1 \text{ Torr};$ $\approx 1 \text{ mmHg}$	19.337×10^{-3}
1 psi	6.894×10^3	68.948×10^{-3}	70.307×10^{-3}	68.046×10^{-3}	51.715	$\equiv 1 \text{ lbf/in}^2$

Example reading: $1 \text{ Pa} = 1 \text{ N/m}^2 = 10^{-5} \text{ bar} = 10.197 \times 10^{-6} \text{ at} = 9.8692 \times 10^{-6} \text{ atm} = 7.5006 \times 10^{-3} \text{ torr} = 145.04 \times 10^{-6} \text{ psi}$ etc.

The SI unit for pressure is the pascal (Pa), equal to one newton per square metre ($\text{N}\cdot\text{m}^{-2}$ or $\text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-2}$). This special name for the unit was added in 1971; before that, pressure in SI was expressed in units such as N/m^2 . When indicated, the zero reference is stated in parenthesis following the unit, for example 101 kPa (abs). The pound per square inch (psi) is still in widespread use in the US and Canada, notably for cars. A letter is often appended to the psi unit to indicate the measurement's zero reference; psia for absolute, psig for gauge, psid for differential, although this practice is discouraged by the NIST.

Because pressure was once commonly measured by its ability to displace a column of liquid in a manometer, pressures are often expressed as a depth of a particular fluid (e.g. inches of water). The most common choices are mercury (Hg) and water; water is nontoxic and readily available, while mercury's density allows for a shorter column (and so a smaller manometer) to measure a given pressure.

Fluid density and local gravity can vary from one reading to another depending on local factors, so the height of a fluid column does not define pressure precisely. When 'millimetres of mercury' or 'inches of mercury' are quoted today, these units are not based on a physical column of mercury; rather, they have been given precise definitions that can be expressed in terms of SI units. The water-based units usually assume one of the older definitions of the kilogram as the weight of a litre of water.

Although no longer favoured by measurement experts, these **manometric units** are still encountered in many fields. Blood pressure is measured in millimetres of mercury in most of the world, and lung pressures in centimeters of water are still common. Natural gas pipeline pressures are measured in inches of water, expressed as "'WC' ('Water Column'). Scuba divers often use a manometric rule of thumb: the pressure exerted by ten metres depth of water is approximately equal to one atmosphere. In vacuum systems, the units torr, micrometre of mercury (micron), and inch of mercury (inHg) are most commonly used. Torr and micron usually indicates an absolute pressure, while inHg usually indicates a gauge pressure.

Atmospheric pressures are usually stated using kilopascal (kPa), or atmospheres (atm), except in American meteorology where the hectopascal (hPa) and millibar (mbar) are preferred. In American and Canadian engineering, stress is often measured in kip. Note that stress is not a true pressure since it is not scalar. In the cgs system the unit of pressure was the barye (ba), equal to $1 \text{ dyn}\cdot\text{cm}^{-2}$. In the mts system, the unit of pressure was the pieze, equal to 1 sthene per square metre.

Many other hybrid units are used such as mmHg/cm² or grams-force/cm² (sometimes as kg/cm² and g/mol² without properly identifying the force units). Using the names kilogram, gram, kilogram-force, or gram-force (or their symbols) as a unit of force is forbidden in SI; the unit of force in SI is the newton (N).

Static and dynamic pressure

Static pressure is uniform in all directions, so pressure measurements are independent of direction in an immovable (static) fluid. Flow, however, applies additional pressure on surfaces perpendicular to the flow direction, while having little impact on surfaces parallel to the flow direction. This directional component of pressure in a moving (dynamic) fluid is called dynamic pressure. An instrument facing the flow direction measures the sum of the static and dynamic pressures; this measurement is called the total pressure or stagnation pressure. Since dynamic pressure is referenced to static pressure, it is neither gauge nor absolute; it is a differential pressure.

While static gauge pressure is of primary importance to determining net loads on pipe walls, dynamic pressure is used to measure flow rates and airspeed. Dynamic pressure can be measured by taking the differential pressure between instruments parallel and perpendicular to the flow. Pitot-static tubes, for example perform this measurement on airplanes to determine airspeed. The presence of the measuring instrument inevitably acts to divert flow and create turbulence, so its shape is critical to accuracy and the calibration curves are often non-linear.

Applications

- Altimeter
- Barometer
- MAP sensor
- Pitot tube
- Sphygmomanometer

Instruments

Many instruments have been invented to measure pressure, with different advantages and disadvantages. Pressure range, sensitivity, dynamic response and cost all vary by several orders of magnitude from one instrument design to the next. The oldest type is the liquid column (a vertical tube filled with mercury) manometer invented by Evangelista Torricelli in 1643. The U-Tube was invented by Christian Huygens in 1661.

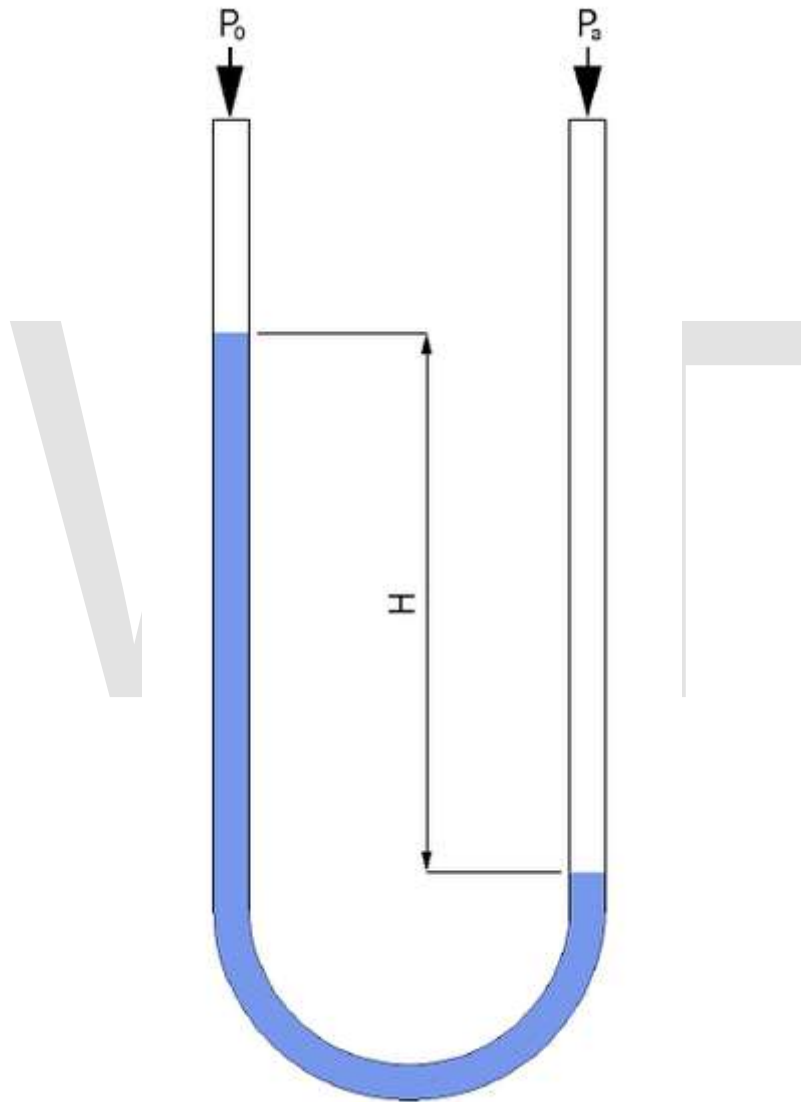
Hydrostatic

Hydrostatic gauges (such as the mercury column manometer) compare pressure to the hydrostatic force per unit area at the base of a column of fluid. Hydrostatic gauge measurements are independent of the type of gas being measured, and can be designed to have a very linear calibration. They have poor dynamic response.

Piston

Piston-type gauges counterbalance the pressure of a fluid with a solid weight or a spring. Another name for piston gauge is deadweight tester. For example, dead-weight testers used for calibration or tire-pressure gauges.

Liquid column



The difference in fluid height in a liquid column manometer is proportional to the

$$H = \frac{P_a - P_o}{g\rho}$$

pressure difference.

Liquid column gauges consist of a vertical column of liquid in a tube whose ends are exposed to different pressures. The column will rise or fall until its weight is in equilibrium with the pressure differential between the two ends of the tube. A very simple version is a U-shaped tube half-full of liquid, one side of which is connected to the region of interest while the reference pressure (which might be the atmospheric pressure or a vacuum) is applied to the other. The difference in liquid level represents the applied pressure. The pressure exerted by a column of fluid of height h and density ρ is given by the hydrostatic pressure equation, $P = h\rho g$. Therefore the pressure difference between the applied pressure P_a and the reference pressure P_0 in a U-tube manometer can be found by solving $P_a - P_0 = h\rho g$. If the fluid being measured is significantly dense, hydrostatic corrections may have to be made for the height between the moving surface of the manometer working fluid and the location where the pressure measurement is desired.

Although any fluid can be used, mercury is preferred for its high density (13.534 g/cm^3) and low vapour pressure. For low pressure differences well above the vapour pressure of water, water is commonly used (and "inches of water" is a common pressure unit). Liquid-column pressure gauges are independent of the type of gas being measured and have a highly linear calibration. They have poor dynamic response. When measuring vacuum, the working liquid may evaporate and contaminate the vacuum if its vapor pressure is too high. When measuring liquid pressure, a loop filled with gas or a light fluid must isolate the liquids to prevent them from mixing. Simple hydrostatic gauges can measure pressures ranging from a few Torr (a few 100 Pa) to a few atmospheres. (Approximately 1,000,000 Pa)

A single-limb liquid-column manometer has a larger reservoir instead of one side of the U-tube and has a scale beside the narrower column. The column may be inclined to further amplify the liquid movement. Based on the use and structure following type of manometers are used

1. Simple Manometer
2. Micromanometer
3. Differential manometer
4. Inverted differential manometer



A McLeod gauge, drained of mercury

McLeod gauge

A McLeod gauge isolates a sample of gas and compresses it in a modified mercury manometer until the pressure is a few mmHg. The gas must be well-behaved during its compression (it must not condense, for example). The technique is slow and unsuited to continual monitoring, but is capable of good accuracy.

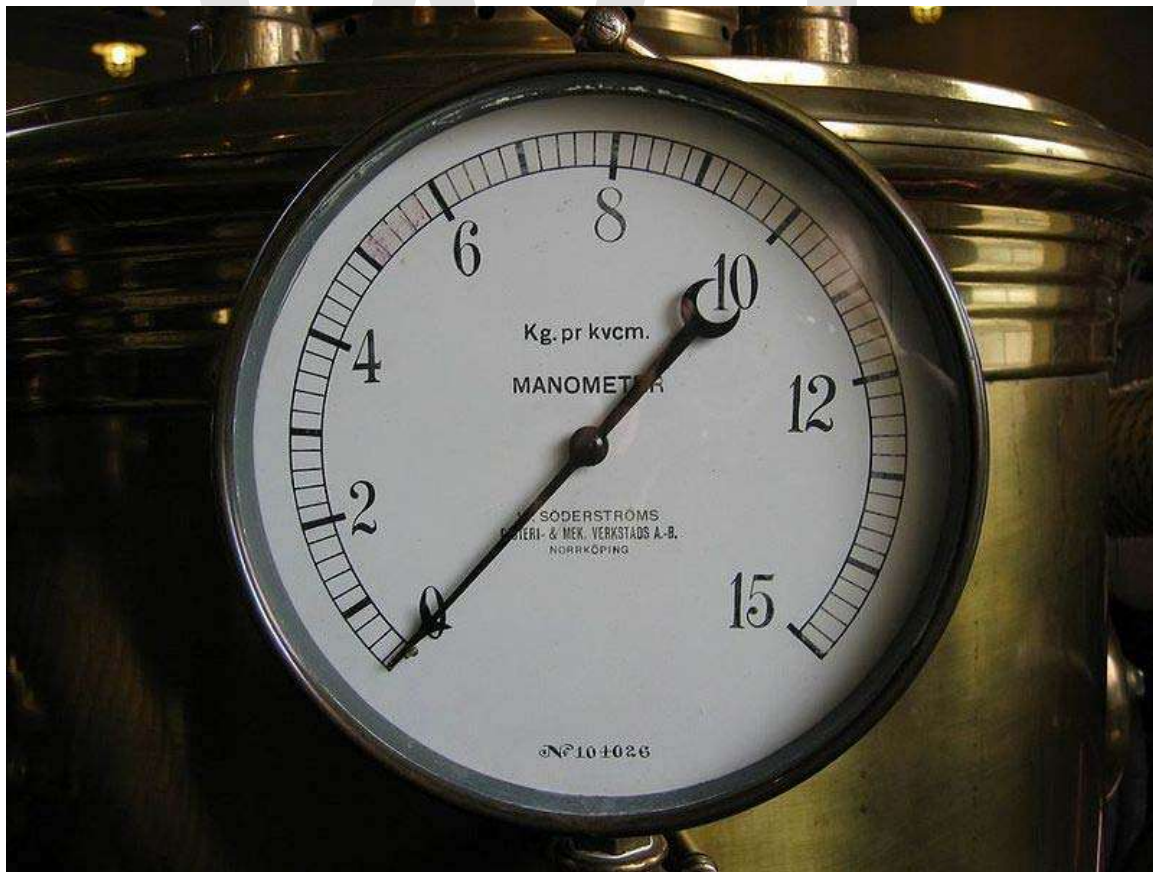
Useful range: above 10^{-4} torr (roughly 10^{-2} Pa) as high as 10^{-6} Torr (0.1 mPa),

0.1 mPa is the lowest direct measurement of pressure that is possible with current technology. Other vacuum gauges can measure lower pressures, but only indirectly by measurement of other pressure-controlled properties. These indirect measurements must be calibrated to SI units via a direct measurement, most commonly a McLeod gauge.

Aneroid

Aneroid gauges are based on a metallic pressure sensing element which flexes elastically under the effect of a pressure difference across the element. "Aneroid" means "without fluid," and the term originally distinguished these gauges from the hydrostatic gauges described above. However, aneroid gauges can be used to measure the pressure of a liquid as well as a gas, and they are not the only type of gauge that can operate without fluid. For this reason, they are often called **mechanical** gauges in modern language. Aneroid gauges are not dependent on the type of gas being measured, unlike thermal and ionization gauges, and are less likely to contaminate the system than hydrostatic gauges. The pressure sensing element may be a **Bourdon tube**, a diaphragm, a capsule, or a set of bellows, which will change shape in response to the pressure of the region in question. The deflection of the pressure sensing element may be read by a linkage connected to a needle, or it may be read by a secondary transducer. The most common secondary transducers in modern vacuum gauges measure a change in capacitance due to the mechanical deflection. Gauges that rely on a change in capacitances are often referred to as Baratron gauges.

Bourdon



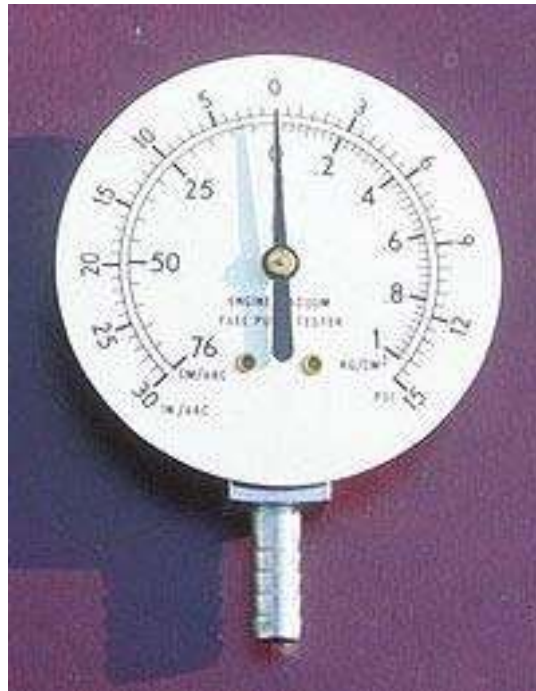
Membrane-type manometer

The Bourdon pressure gauge uses the principle that a flattened tube tends to change to a more circular cross-section when pressurized. Although this change in cross-section may be hardly noticeable, and thus involving moderate stresses within the elastic range of easily workable materials, the strain of the material of the tube is magnified by forming the tube into a C shape or even a helix, such that the entire tube tends to straighten out or uncoil, elastically, as it is pressurized. Eugene Bourdon patented his gauge in France in 1849, and it was widely adopted because of its superior sensitivity, linearity, and accuracy; Edward Ashcroft purchased Bourdon's American patent rights in 1852 and became a major manufacturer of gauges. Also in 1849, Bernard Schaeffer in Magdeburg, Germany patented a successful diaphragm (see below) pressure gauge, which together with the Bourdon gauge, revolutionized pressure measurement in industry. But in 1875 after Bourdon's patents expired, his company Schaeffer and Budenberg also manufactured Bourdon tube gauges.

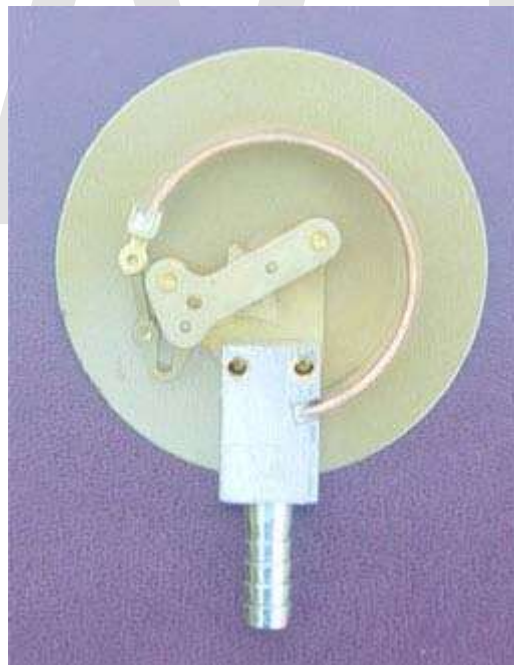
In practice, a flattened thin-wall, closed-end tube is connected at the hollow end to a fixed pipe containing the fluid pressure to be measured. As the pressure increases, the closed end moves in an arc, and this motion is converted into the rotation of a (segment of a) gear by a connecting link which is usually adjustable. A small diameter pinion gear is on the pointer shaft, so the motion is magnified further by the gear ratio. The positioning of the indicator card behind the pointer, the initial pointer shaft position, the linkage length and initial position, all provide means to calibrate the pointer to indicate the desired range of pressure for variations in the behaviour of the Bourdon tube itself. Differential pressure can be measured by gauges containing two different Bourdon tubes, with connecting linkages.

Bourdon tubes measure gauge pressure, relative to ambient atmospheric pressure, as opposed to absolute pressure; vacuum is sensed as a reverse motion. Some aneroid barometers use Bourdon tubes closed at both ends (but most use diaphragms or capsules, see below). When the measured pressure is rapidly pulsing, such as when the gauge is near a reciprocating pump, an orifice restriction in the connecting pipe is frequently used to avoid unnecessary wear on the gears and provide an average reading; when the whole gauge is subject to mechanical vibration, the entire case including the pointer and indicator card can be filled with an oil or glycerin. Typical high-quality modern gauges provide an accuracy of $\pm 2\%$ of span, and a special high-precision gauge can be as accurate as 0.1% of full scale.

In the following illustrations the transparent cover face of the pictured combination pressure and vacuum gauge has been removed and the mechanism removed from the case. This particular gauge is a combination vacuum and pressure gauge used for automotive diagnosis:



Indicator side with card and dial

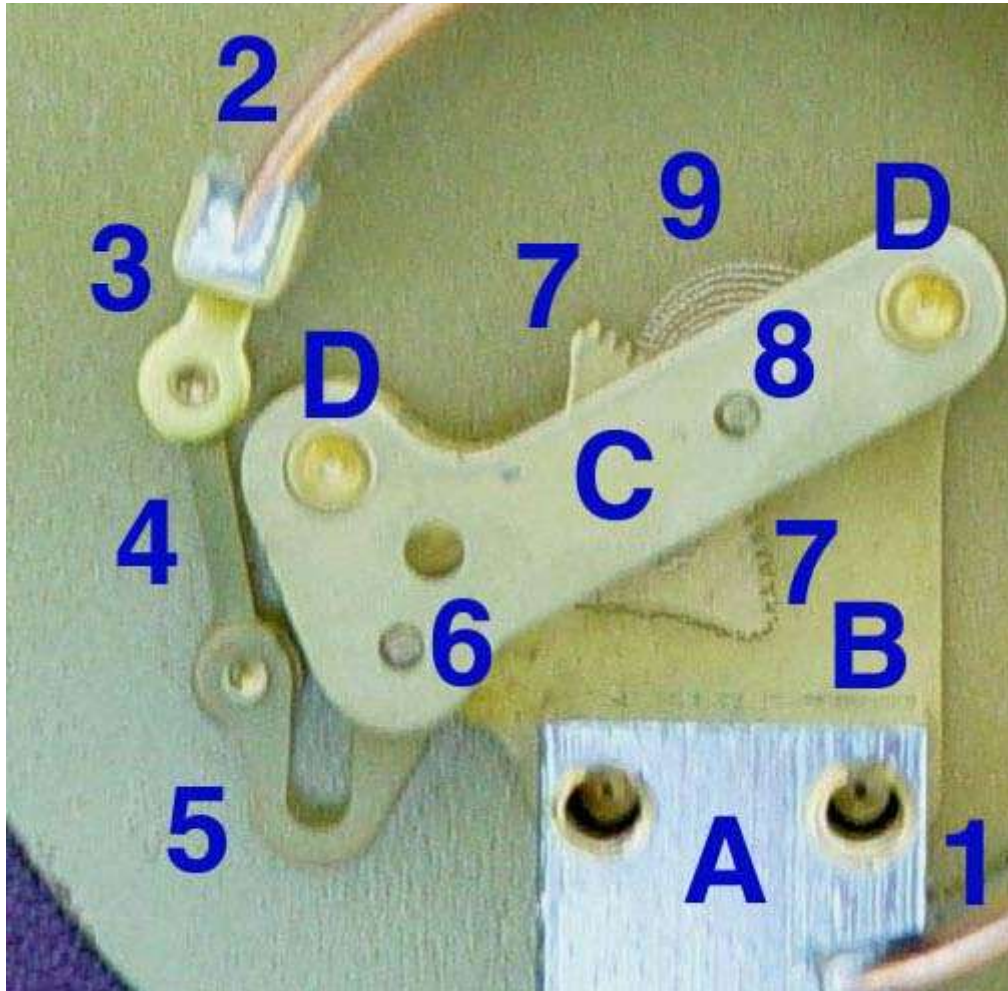


Mechanical side with Bourdon tube

- the left side of the face, used for measuring manifold vacuum, is calibrated in centimetres of mercury on its inner scale and inches of mercury on its outer scale.

- the right portion of the face is used to measure fuel pump pressure and is calibrated in fractions of 1 kgf/cm² on its inner scale and pounds per square inch on its outer scale.

Mechanical details



Mechanical details

Stationary parts:

- A: Receiver block. This joins the inlet pipe to the fixed end of the Bourdon tube (1) and secures the chassis plate (B). The two holes receive screws that secure the case.
- B: Chassis plate. The face card is attached to this. It contains bearing holes for the axles.
- C: Secondary chassis plate. It supports the outer ends of the axles.
- D: Posts to join and space the two chassis plates.

Moving Parts:

1. Stationary end of Bourdon tube. This communicates with the inlet pipe through the receiver block.
2. Moving end of Bourdon tube. This end is sealed.
3. Pivot and pivot pin.
4. Link joining pivot pin to lever (5) with pins to allow joint rotation.
5. Lever. This an extension of the sector gear (7).
6. Sector gear axle pin.
7. Sector gear.
8. Indicator needle axle. This has a spur gear that engages the sector gear (7) and extends through the face to drive the indicator needle. Due to the short distance between the lever arm link boss and the pivot pin and the difference between the effective radius of the sector gear and that of the spur gear, any motion of the Bourdon tube is greatly amplified. A small motion of the tube results in a large motion of the indicator needle.
9. Hair spring to preload the gear train to eliminate gear lash and hysteresis.

Diaphragm



A pile of pressure capsules with corrugated diaphragms in an aneroid barograph

A second type of aneroid gauge uses the deflection of a flexible membrane that separates regions of different pressure. The amount of deflection is repeatable for known pressures so the pressure can be determined by using calibration. The deformation of a thin diaphragm is dependent on the difference in pressure between its two faces. The reference face can be open to atmosphere to measure gauge pressure, open to a second port to measure differential pressure, or can be sealed against a vacuum or other fixed reference pressure to measure absolute pressure. The deformation can be measured using mechanical, optical or capacitive techniques. Ceramic and metallic diaphragms are used.

Useful range: above 10^{-2} Torr (roughly 1 Pa)

For absolute measurements, welded pressure capsules with diaphragms on either side are often used.

Shape:

- Flat
- corrugated
- flattened tube
- capsule

Bellows

In gauges intended to sense small pressures or pressure differences, or require that an absolute pressure be measured, the gear train and needle may be driven by an enclosed and sealed bellows chamber, called an **aneroid**, which means "without liquid". (Early barometers used a column of liquid such as water or the liquid metal mercury suspended by a vacuum.) This bellows configuration is used in aneroid barometers (barometers with an indicating needle and dial card), altimeters, altitude recording barographs, and the altitude telemetry instruments used in weather balloon radiosondes. These devices use the sealed chamber as a reference pressure and are driven by the external pressure. Other sensitive aircraft instruments such as air speed indicators and rate of climb indicators (variometers) have connections both to the internal part of the aneroid chamber and to an external enclosing chamber.

Electronic pressure sensors

- ***Piezoresistive Strain Gage***

Uses the piezoresistive effect of bonded or formed strain gauges to detect strain due to applied pressure.

- ***Capacitive***

Uses a diaphragm and pressure cavity to create a variable capacitor to detect strain due to applied pressure.

- ***Magnetic***

Measures the displacement of a diaphragm by means of changes in inductance (reluctance), LVDT, Hall Effect, or by eddy current principal.

- ***Piezoelectric***

Uses the piezoelectric effect in certain materials such as quartz to measure the strain upon the sensing mechanism due to pressure.

- ***Optical***

Uses the physical change of an optical fiber to detect strain due applied pressure.

- ***Potentiometric***

Uses the motion of a wiper along a resistive mechanism to detect the strain caused by applied pressure.

- ***Resonant***

Uses the changes in resonant frequency in a sensing mechanism to measure stress, or changes in gas density, caused by applied pressure.

Thermal conductivity

Generally, as a real gas increases in density -which may indicate an increase in pressure- its ability to conduct heat increases. In this type of gauge, a wire filament is heated by running current through it. A thermocouple or Resistance Temperature Detector (RTD) can then be used to measure the temperature of the filament. This temperature is dependent on the rate at which the filament loses heat to the surrounding gas, and therefore on the thermal conductivity. A common variant is the Pirani gauge which uses a single platinum filament as both the heated element and RTD. These gauges are accurate from 10 Torr to 10^{-3} Torr, but they are sensitive to the chemical composition of the gases being measured.

Two wire

One wire coil is used as a heater, and the other is used to measure nearby temperature due to convection.

Pirani (one wire)

A Pirani gauge consists of a metal wire open to the pressure being measured. The wire is heated by a current flowing through it and cooled by the gas surrounding it. If the gas pressure is reduced, the cooling effect will decrease, hence the equilibrium temperature of the wire will increase. The resistance of the wire is a function of its temperature: by measuring the voltage across the wire and the current flowing through it, the resistance (and so the gas pressure) can be determined. This type of gauge was invented by Marcello Pirani.

Thermocouple gauges and **thermistor gauges** work in a similar manner, except a thermocouple or thermistor is used to measure the temperature of the wire.

Useful range: 10^{-3} - 10 Torr (roughly 10^{-1} - 1000 Pa)

Ionization gauge

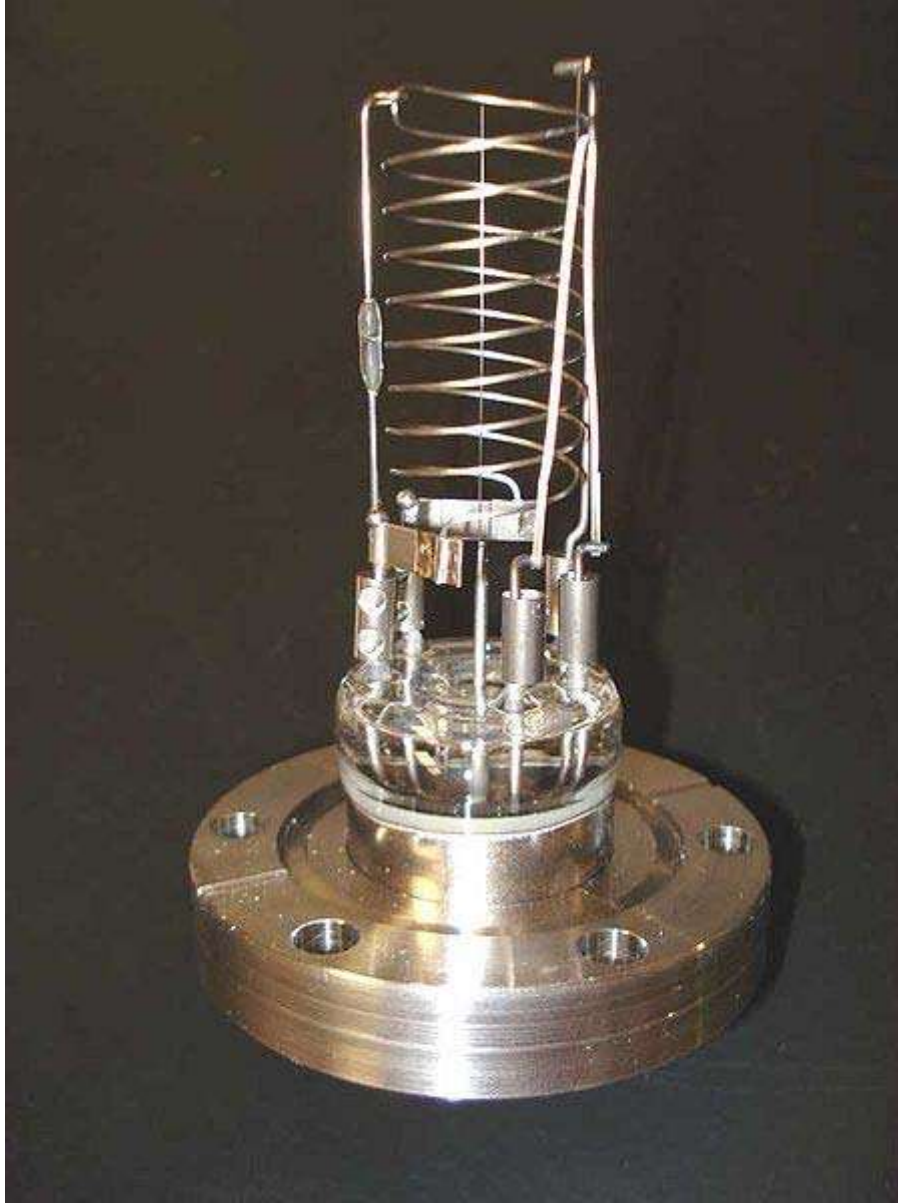
Ionization gauges are the most sensitive gauges for very low pressures (also referred to as hard or high vacuum). They sense pressure indirectly by measuring the electrical ions produced when the gas is bombarded with electrons. Fewer ions will be produced by lower density gases. The calibration of an ion gauge is unstable and dependent on the nature of the gases being measured, which is not always known. They can be calibrated against a McLeod gauge which is much more stable and independent of gas chemistry.

Thermionic emission generate electrons, which collide with gas atoms and generate positive ions. The ions are attracted to a suitably biased electrode known as the collector. The current in the collector is proportional to the rate of ionization, which is a function of the pressure in the system. Hence, measuring the collector current gives the gas pressure. There are several sub-types of ionization gauge.

Useful range: 10^{-10} - 10^{-3} torr (roughly 10^{-8} - 10^{-1} Pa)

Most ion gauges come in two types: hot cathode and cold cathode, a third type exists which is more sensitive and expensive known as a spinning rotor gauge, but is not discussed here. In the hot cathode version an electrically heated filament produces an electron beam. The electrons travel through the gauge and ionize gas molecules around them. The resulting ions are collected at a negative electrode. The current depends on the number of ions, which depends on the pressure in the gauge. Hot cathode gauges are accurate from 10^{-3} Torr to 10^{-10} Torr. The principle behind cold cathode version is the same, except that electrons are produced in the discharge of a high voltage. Cold Cathode gauges are accurate from 10^{-2} Torr to 10^{-9} Torr. Ionization gauge calibration is very sensitive to construction geometry, chemical composition of gases being measured, corrosion and surface deposits. Their calibration can be invalidated by activation at atmospheric pressure or low vacuum. The composition of gases at high vacuums will usually be unpredictable, so a mass spectrometer must be used in conjunction with the ionization gauge for accurate measurement.

Hot cathode



Bayard-Alpert hot cathode ionization gauge

A hot cathode ionization gauge is mainly composed of three electrodes acting together as a triode, where the cathode is the filament. The three electrodes are a collector or plate, a filament, and a grid. The collector current is measured in picoamps by an electrometer. The filament voltage to ground is usually at a potential of 30 volts while the grid voltage at 180–210 volts DC, unless there is an optional electron bombardment feature, by heating the grid which may have a high potential of approximately 565 volts. The most common ion gauge is the hot cathode **Bayard-Alpert gauge**, with a small ion collector inside the grid. A glass envelope with an opening to the vacuum can surround the electrodes, but usually the **Nude Gauge** is inserted in the vacuum chamber directly, the

pins being fed through a ceramic plate in the wall of the chamber. Hot cathode gauges can be damaged or lose their calibration if they are exposed to atmospheric pressure or even low vacuum while hot. The measurements of a hot cathode ionization gauge are always logarithmic.

Electrons emitted from the filament move several times in back and forth movements around the grid before finally entering the grid. During these movements, some electrons collide with a gaseous molecule to form a pair of an ion and an electron (Electron ionization). The number of these ions is proportional to the gaseous molecule density multiplied by the electron current emitted from the filament, and these ions pour into the collector to form an ion current. Since the gaseous molecule density is proportional to the pressure, the pressure is estimated by measuring the ion current.

The low pressure sensitivity of hot cathode gauges is limited by the photoelectric effect. Electrons hitting the grid produce x-rays that produce photoelectric noise in the ion collector. This limits the range of older hot cathode gauges to 10^{-8} Torr and the Bayard-Alpert to about 10^{-10} Torr. Additional wires at cathode potential in the line of sight between the ion collector and the grid prevent this effect. In the extraction type the ions are not attracted by a wire, but by an open cone. As the ions cannot decide which part of the cone to hit, they pass through the hole and form an ion beam. This ion beam can be passed on to a

- Faraday cup
- Microchannel plate detector with Faraday cup
- Quadrupole mass analyzer with Faraday cup
- Quadrupole mass analyzer with Microchannel plate detector Faraday cup
- ion lens and acceleration voltage and directed at a target to form a sputter gun. In this case a valve lets gas into the grid-cage.

Cold cathode

There are two subtypes of cold cathode ionization gauges: the **Penning gauge** (invented by Frans Michel Penning), and the **Inverted magnetron**, also called a **Redhead gauge**. The major difference between the two is the position of the anode with respect to the cathode. Neither has a filament, and each may require a DC potential of about 4 kV for operation. Inverted magnetrons can measure down to 1×10^{-12} Torr.

Such gauges cannot operate if the ions generated by the cathode recombine before reaching the anodes. If the mean-free path of the gas within the gauge is smaller than the gauge's dimensions, then the electrode current will essentially vanish. A practical upper-bound to the detectable pressure is, for a Penning gauge, of the order of 10^{-3} Torr.

Similarly, cold cathode gauges may be reluctant to start at very low pressures, in that the near-absence of a gas makes it difficult to establish an electrode current - particularly in Penning gauges which use an axially symmetric magnetic field to create path lengths for ions which are of the order of metres. In ambient air suitable ion-pairs are ubiquitously

formed by cosmic radiation; in a Penning gauge design features are used to ease the set-up of a discharge path. For example, the electrode of a Penning gauge is usually finely tapered to facilitate the field emission of electrons.

Maintenance cycles of cold cathode gauges are generally measured in years, depending on the gas type and pressure that they are operated in. Using a cold cathode gauge in gases with substantial organic components, such as pump oil fractions, can result in the growth of delicate carbon films and shards within the gauge which eventually either short-circuit the electrodes of the gauge, or impede the generation of a discharge path.

Calibration

Pressure gauges are either direct- or indirect-reading. Hydrostatic and elastic gauges measure pressure are directly influenced by force exerted on the surface by incident particle flux, and are called direct reading gauges. Thermal and ionization gauges read pressure indirectly by measuring a gas property that changes in a predictable manner with gas density. Indirect measurements are susceptible to more errors than direct measurements.

- Dead weight tester
- McLeod
- mass spec + ionization

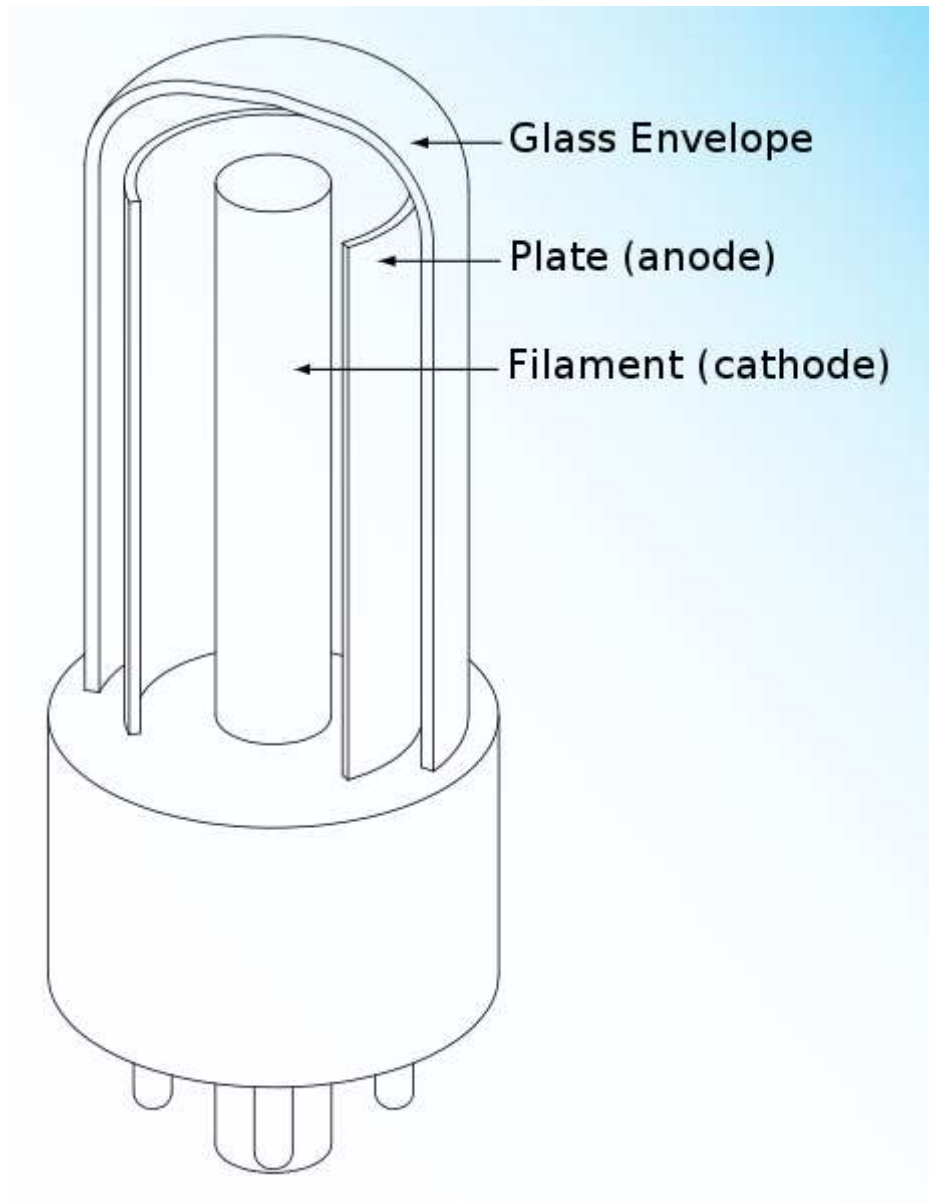
Dynamic transients

When fluid flows are not in equilibrium, local pressures may be higher or lower than the average pressure in a medium. These disturbances propagate from their source as longitudinal pressure variations along the path of propagation. This is also called sound. Sound pressure is the instantaneous local pressure deviation from the average pressure caused by a sound wave. Sound pressure can be measured using a microphone in air and a hydrophone in water. The effective sound pressure is the root mean square of the instantaneous sound pressure over a given interval of time. Sound pressures are normally small and are often expressed in units of microbar.

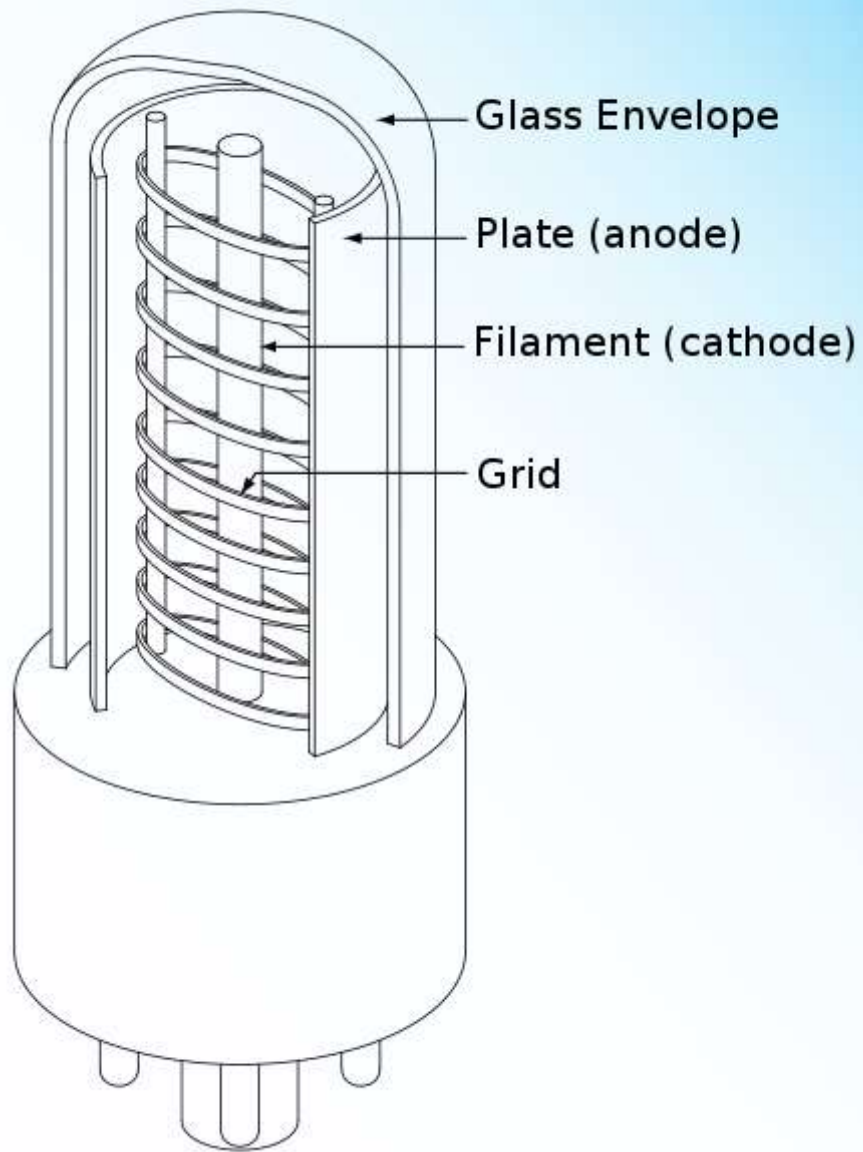
- frequency response of pressure sensors
- resonance

Chapter- 5

Vacuum Tube



Structure of a vacuum tube diode



Structure of a vacuum tube triode



An RCA triode vacuum tube, type 808

In electronics, a **vacuum tube**, **electron tube** (in North America), or **thermionic valve** (elsewhere, especially in Britain) is a device used to amplify, switch, otherwise modify, or create an electrical signal by controlling the movement of electrons in a low-pressure space. Some special function vacuum tubes are filled with low-pressure gas: these are so-called soft tubes as distinct from the hard vacuum type which have the internal gas pressure reduced as far as possible. Almost all tubes depend on the thermionic emission of electrons.

Description

A vacuum tube consists of electrodes in a vacuum in an insulating heat-resistant envelope which is usually tubular. Many tubes have glass envelopes, though some types such as power tubes may have ceramic or metal envelopes. The electrodes are attached to leads which pass through the envelope via an airtight seal. On most tubes, the leads are designed to plug into a tube socket for easy replacement.

The simplest vacuum tubes resemble incandescent light bulbs in that they have a filament sealed in a glass envelope which has been evacuated of all air. When hot, the filament releases electrons into the vacuum: a process called thermionic emission. The resulting negatively charged cloud of electrons is called a space charge. These electrons will be drawn to a metal plate inside the envelope, if the plate (also called the anode) is positively charged relative to the filament (or cathode). The result is a flow of electrons from filament to plate. This cannot work in the reverse direction because the plate is not heated and does not emit electrons. This very simple example described can thus be seen to operate as a diode: a device that conducts current only in one direction. The vacuum tube diode conducts conventional current from plate (anode) to the filament (cathode); this is the opposite direction to the flow of electrons (called electron current).

Vacuum tubes require a large temperature difference between the hot cathode and the cold anode. Because of this, vacuum tubes are inherently power-inefficient; enclosing the tube within a heat-retaining envelope of insulation would allow the entire tube to reach the same temperature, resulting in electron emission from the anode that would counter the normal one-way current. Because the tube requires a vacuum to operate, convection cooling of the anode is not generally possible unless the anode forms a part of the vacuum envelope (in which case the cooling is by conduction through the anode material and then convection outside the vacuum envelope). Thus anode cooling occurs in most tubes through black-body radiation and conduction of heat to the outer glass envelope via the anode mounting frame. Cold cathode tubes do not rely on thermionic emission at the cathode and usually have some form of gas discharge as the operating principle; such tubes are used for lighting (neon lamps) or as voltage regulators.

Sometimes another electrode, called a control grid, is added between the cathode and the anode. The vacuum tube is then known as a "triode." A triode is a *voltage-controlled device*, in that a voltage that is applied as an input to the grid can be used to modulate the rate of electron flow between anode and cathode. The relationship between this input voltage and the output current is determined by a transconductance function. Control grid current is practically negligible in most circuits. The solid-state device most closely analogous to the vacuum tube is the JFET, although the vacuum tube typically operates at far higher voltage (and power) levels than the JFET.

History and development



Inside of a vacuum tube with plate cut open

The 19th century saw increasing research with evacuated tubes, such as the Geissler and Crookes tubes. Scientists who experimented with such tubes included Eugen Goldstein, Nikola Tesla, Johann Wilhelm Hittorf, Thomas Edison, and many others. These tubes were mostly for specialized scientific applications, or were novelties, with the exception of the light bulb. The groundwork laid by these scientists and inventors, however, was critical to the development of vacuum tube technology.

Though the thermionic emission effect was originally reported in 1873 by Frederick Guthrie, it is Thomas Edison's 1884 investigation of the Edison Effect that is more often mentioned. Edison patented what he found, but he did not understand the underlying physics, or the potential value of the discovery. It wasn't until the early 20th century that this effect was put to use, in applications such as John Ambrose Fleming's diode used as a radio detector, and Lee De Forest's 1906 "audion" (soon improved by others as the triode in 1908) used in the first telephone amplifiers. These developments led to great improvements in telecommunications technology, particularly the first coast-to-coast telephone line in the US, and the birth of broadcast radio.

Diodes and triodes

The English physicist John Ambrose Fleming worked as an engineering consultant for firms, including Edison Telephone and the Marconi Company. In 1904, as a result of experiments conducted on Edison effect bulbs imported from the USA, he developed a device he called an "oscillation valve" (because it passes current in only one direction).

Later known as the Fleming valve, it could be used as a rectifier of alternating current and as a radio wave detector.

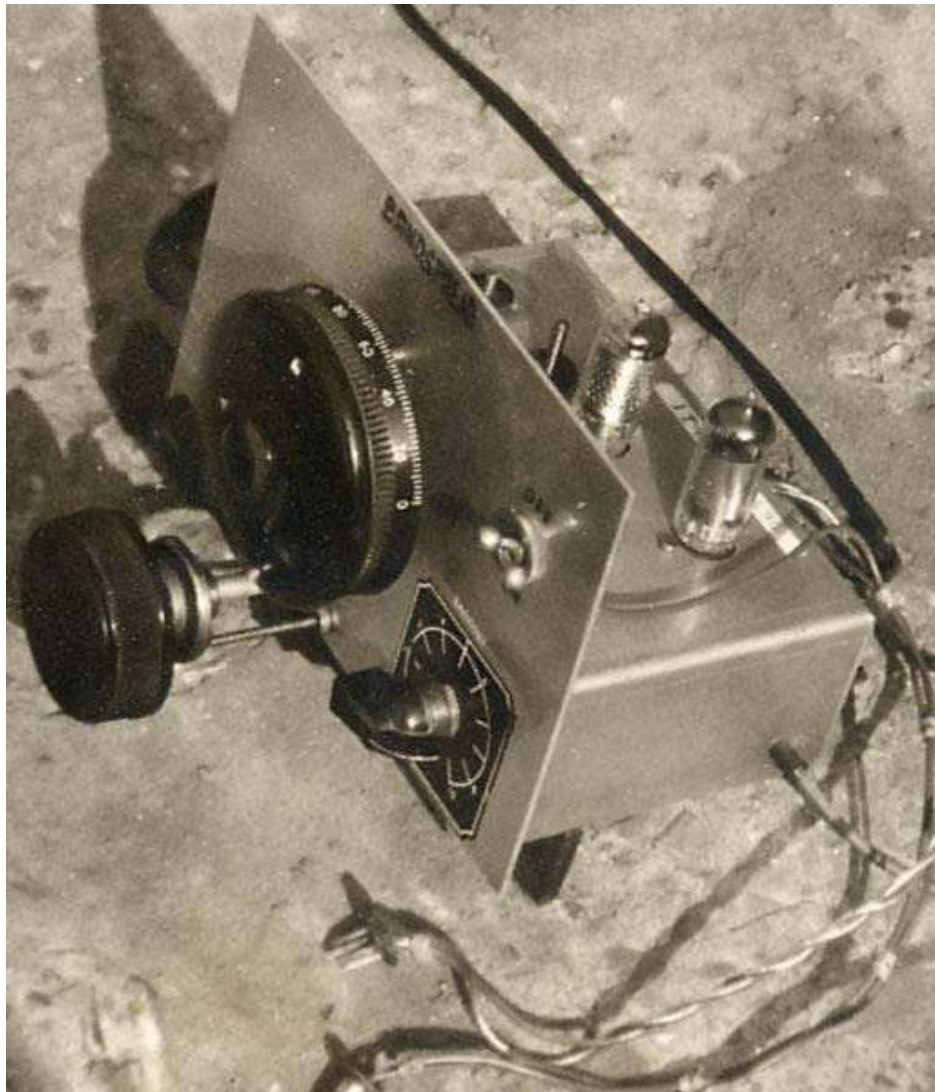
In 1906 Robert von Lieben filed for a three-electrode amplifying vacuum tube. His invention also included a beam-focusing electromagnet.

In 1907 Lee De Forest placed a bent wire serving as a screen, later known as the "grid" electrode, between the filament and plate electrode. As the voltage applied to the grid was varied from negative to positive, the number of electrons flowing from the filament to the plate would vary accordingly. Thus the grid was said to electrostatically "control" the plate current. The resulting three-electrode device was therefore an excellent and very sensitive amplifier of voltages. De Forest called his invention the "Audion". In 1907, De Forest filed for a three-electrode version of the Audion for use in radio communications. The device is now known as the triode. De Forest's device was not strictly a vacuum tube, but clearly depended for its action on ionisation of the relatively high levels of gas remaining after evacuation. The De Forest company, in its Audion leaflets, warned against operation which might cause the vacuum to become too hard. The Finnish inventor Eric Tigerstedt significantly improved on the original triode design in 1914, while working on his sound-on-film process in Berlin, Germany. The first true vacuum triodes were the Pliotrons developed by Irving Langmuir at the General Electric research laboratory (Schenectady, New York) in 1915. Langmuir was one of the first scientists to realize that a harder vacuum would improve the amplifying behaviour of the triode. Pliotrons were closely followed by the French 'R' Type which was in widespread use by the allied military by 1916. These two types were the first true *vacuum* tubes. Historically, vacuum levels in production vacuum tubes typically ranged between 10 μ Pa to 10 nPa.

The non-linear operating characteristic of the triode caused early tube audio amplifiers to exhibit harmonic distortions at low volumes. This is not to be confused with the

overdrive that tube amplifiers exhibit at high volume levels (known as the tube sound). To remedy the low-volume distortion problem, engineers plotted curves of the applied grid voltage and resulting plate currents, and discovered that there was a range of relatively linear operation. In order to use this range, a negative voltage had to be applied to the grid to place the tube in the "middle" of the linear area with no signal applied. This was called the idle condition, and the plate current at this point the "idle current". Today this current would be called the quiescent or standing current. The controlling voltage was superimposed onto this fixed voltage, resulting in linear swings of plate current for both positive and negative swings of the input voltage. This concept was called *grid bias*.

Tetrodes and pentodes



A two-tube homemade radio from 1958. The tubes are the two columns with the dark tops. The flying leads connect to the low-voltage filament and high-voltage anode supplies.

When triodes were first used in radio transmitters and receivers, it was found that they had a tendency to oscillate due to parasitic anode-to-grid capacitance. Many circuits were developed to reduce this problem (e.g. the Neutrodyne amplifier), but proved unsatisfactory over wide ranges of frequencies. It was discovered that the addition of a second grid, located between the control grid and the plate and called a *screen grid* could solve these problems. ("Screen" implies shielding, not physical construction.) A positive voltage slightly lower than the plate voltage was applied to it, and the screen grid was bypassed (for high frequencies) to ground with a capacitor. This arrangement decoupled the anode and the first grid, completely eliminating the oscillation problem. An additional side effect of this second grid is that the Miller capacitance is also reduced, which improves gain at high frequency. This two-grid tube is called a *tetrode*, meaning four active electrodes, and was common by 1926.



Radio transmitter high-power vacuum tube. The braided copper leads provide heater current for the cathode. The tube also has a heat sink. Dubendorf Museum of Military Aviation.

However, the tetrode has some new problems. In any tube, electrons strike the anode hard enough to knock out secondary electrons. In a triode these (less energetic) electrons cannot reach the grid or cathode, and are re-captured by the anode. But in a tetrode, they can be captured by the second grid, reducing the plate current and the amplification of the circuit. Since secondary electrons can outnumber the primary electrons, in the worst case, particularly when the plate voltage dips below the screen voltage, the plate current can actually go down with increasing plate voltage. This is negative-resistance behavior. This is the "tetrode kink". Another consequence of this effect is that under severe overload, the current collected by the screen grid can cause it to overheat and melt, destroying the tube.

Again, the solution was to add another grid, called a suppressor grid. This third grid was biased at either ground or cathode voltage and its negative voltage (relative to the anode) electrostatically suppressed the secondary electrons by repelling them back toward the

anode. This three-grid tube is called a pentode, meaning five electrodes. The pentode was invented in 1928 by Bernard D. H. Tellegen.

Improvements in construction and performance

The very earliest vacuum tubes strongly resembled incandescent light bulbs and were made by lamp manufacturers, who had the equipment for manufacture of glass envelopes and the powerful vacuum pumps required to evacuate the enclosures. After World War I, specialized manufacturers using more economical construction methods were set up to fill the growing demand for broadcast receivers. Bare tungsten filaments operated at a temperature of around 2200 °C. The development of oxide-coated filaments in the mid 1920s reduced filament operating temperature to a dull red heat (around 700 °C), which in turn reduced thermal distortion of the tube structure and allowed closer spacing of tube elements. This in turn improved tube gain, since the gain of a triode is inversely proportional to the spacing between grid and cathode. Development of the indirectly-heated cathode, with the filament inside a cylinder of oxide-coated nickel, further reduced distortion of the tube elements and also allowed the cathode heaters to be run from an AC supply.



Heat transfer and appearance of tubes



The anode of this tube has been designed to dissipate heat

Many types of vacuum tubes can be recognized from their appearance. A considerable amount of heat is produced when tubes operate. In most circuits the tube is about 30-60% efficient dependent on the class of operation (classes A, B, or C), which means that 40-70 % of input power to the stage is lost as heat. The requirements for heat removal significantly change the appearance of high-power vacuum tubes.

Most tubes contain two sources of heat when operating. The first one of these is the filament or heater. Some types contain a *directly heated cathode*. This is a filament

similar to an incandescent electric lamp and some types glow brightly like a lamp, but most glow dimly. (The "bright emitter" types possess a tungsten filament alloyed with 1-3 % thorium which reduces the work function of the metal, giving it the ability to emit sufficient electrons at about 2000 degrees Celsius. The "dull emitter" types also possess a tungsten filament but it is coated in a mixture of calcium, strontium and barium oxides, which emit electrons easily at much lower temperatures due to a monolayer of mixed alkali earth metals coating the tungsten when the cathode is heated to about 800-1000 degrees Celsius.)

The second form of cathode is the *indirectly heated* form which usually consists of a nickel tube, coated on the outside with the same strontium, calcium, barium oxide mix used in the "dull emitter" directly heated types, and fitted with a tungsten filament inside the tube to heat it. This tungsten filament is usually uncoiled and coated in a layer of alumina, (aluminium oxide), to insulate it from the nickel tube of the actual cathode. This form of construction allows for a much greater electron emitting area and, because the heater is insulated from the cathode, the cathode can be positioned in a circuit at up to 150 volts more positive than the heater or 50 volts more negative than the heater for most common types. It also allows all the heaters to be simply wired in series or parallel rather than some requiring special *isolated* power supplies such as specially insulated windings on power transformers or separate batteries. For small-signal tubes such as used in radio receivers, heaters are rated from 50 mW to 5 watts, (directly heated), and about 500 mW to 8 watts for indirectly heated types. Once filament/heater power is included in total power consumption, small tubes have very poor efficiencies. A 6BM8/ECL82 audio stage consumes a total power of some 15 watts for 3.5 watts of useful audio power, giving an efficiency of around 23%. Some signal amplifiers, particularly high-frequency amplifiers such as the 6BA6, consume some 5.9 watts of power in normal operation and deliver only 1.1 watts of power at the plate.

The second source of heat is generated at the anode, when electrons, accelerated by the voltage applied to the anode, strike the anode and impart a considerable fraction of their energy to it, raising its temperature. In tubes used in power amplifier or transmitting circuits, this source of heat will exceed the power dissipated in the cathode heater. (The plates or anodes of 6L6 devices used in guitar amplifiers can sometimes be seen to reach red heat if the *bias* is set too high, they should not emit any visible radiation when driven at maximum ratings.) No tubes in domestic, music, or studio equipment should operate with glowing anodes.

This heat usually escapes the device by black body radiation from the anode/plate as infra red light. Some is conducted away through the connecting wires going to the base but none is convected in most types of tube because of the vacuum and the absence of any gas inside the bulb to convect. It is the way tubes get rid of heat which most affects their overall appearance, next to the type of unit (triode, pentode, etc.) they contain, or whether they contain more than one of these basic units. For devices required to radiate more than 500 mW or so, usually indirectly heated cathode types, the anode or plate is often treated to make its surface less shiny, and to make it darker, either gray or black. This helps it radiate the generated heat and maintain the anode or plate at a temperature significantly

lower than the cathode, a requirement for proper operation. Types 6BQ5/EL84 and 6BM8/ECL82 are examples of indirectly heated types with gray anodes.

Other internal elements of high-power tubes, such as control grids and screen grids, may also dissipate heat if carrying large currents. Limits to grid dissipation are listed for such devices, to prevent distortion and failure of the grids.

Tubes used as power amplifier stages for radio transmitters may have additional heat exchangers, cooling fans, radiator fins, or other measures to improve heat transfer at the anode. Broadcast transmitters may use water-cooling or evaporative cooling for tube anodes. The water cooling system must withstand the high voltages present on the anode.

Low power rated tubes, such as the 1.4 volt filament, directly heated tubes, designed for use in battery powered equipment, often retain shiny metal anodes as they produce so little heat. 1T4, 1R5 and 1A7 are examples of devices with shiny untreated anodes. Gas filled tubes, such as thyratrons, although they possess a greater plate dissipation than a "1 volt battery type", still often possess a shiny metal anode finish as the gas filling conducts and convects the heat to the bulb wall. Types 884 and 2D21 are typical examples.

The outer electrode in most tubes is usually the anode. Some small signal types, such as sharp and remote cut-off R.F. and A.F. pentodes and some pentagrid converters have a shield fitted around all the electrodes enclosing the anode. This shield is sometimes a solid metal sheet, treated to make it dull and gray like an anode or plate, and sometimes it is fabricated from expanded metal mesh, acting as a Faraday cage but allowing sufficient heat from the anode beneath to escape. Types 6BX6/EF80 and 6BK8/EF86 are typical examples of this shielded type commonly using expanded mesh. Types 6AU6/EF94 and 6BE6/EK90 are examples using a gray sheet metal cylindrical shield giving them a very similar overall appearance.

Indicators such as some "magic eye" tubes and the type 6977 fluorescent-anode type have glowing electrodes.

Other variations



12SA7 Pentagrid converter tube

Frequency conversion can be accomplished by various methods in superheterodyne receivers. Tubes with 5 grids, called pentagrid converters, were generally used, although alternatives such as using a combination of a triode with a hexode were also used. Even octodes have been used for frequency conversion. The additional grids are either control grids, with different signals applied to each one, or screen grids. In many designs a special grid acted as a second 'leaky' plate to provide a built-in oscillator, which then coupled this signal with the incoming radio signal. These signals create a single, combined effect (equivalent to a crude analog multiplier) on the plate current (and thus

the signal output) of the tube circuit. The useful component of the output was the difference frequency between that of the incoming signal and that of the oscillator.

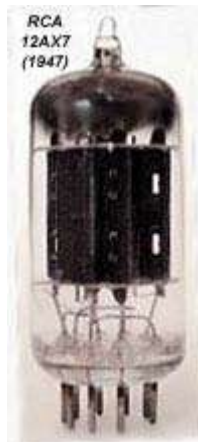
The heptode, or pentagrid converter, was the most common of these. 6BE6 is an example of a heptode (note that the first number in the tube ID indicates the filament voltage). Octodes were rare in the US, the 7A8 was one example, but much more common in Europe particularly in battery operated radios where the lower power consumption was an advantage.

Toward the end of the tube era, precision control and screen grids, called frame grids, offered enhanced performance. Instead of the typically elliptical fine-gauge wire supported by two larger wires, a frame grid was a metal stamping with rectangular openings that surrounded the cathode. The grid wires were in a plane defined by the stamping, and the control grid was placed much closer to the cathode surface than traditional construction would permit.

To reduce the cost and complexity of radio equipment, by 1940 it was common practice to combine more than one function, or more than one set of elements in the bulb of a single tube. The only constraint was where patents, and other licencing considerations required the use of multiple tubes.

For example, the RCA Type 55 was a double diode triode used as a detector, automatic gain control rectifier and audio preamp in early AC powered radios. The same set of tubes often included the 53 Dual Triode Audio Output.

Another early type of multi-section tube, the 6SN7, is a "dual triode" which, for most purposes, can perform the functions of two triode tubes, while taking up half as much space and costing less.



An RCA 12AX7 dual-triode tube (1947)

The 12AX7 is a dual high-gain triode widely used in guitar amplifiers, audio preamps, and instruments.

The invention of the 9-pin miniature tube base, besides allowing the 12AX7 family, also allowed many other multi section tubes, such as the 6GH8 triode pentode. Along with a host of similar tubes, the 6GH8 was quite popular in television receivers. Some color TV sets used exotic types like the 6JH8 which had two plates and beam deflection electrodes (it was known as the 'sheet beam' tube). Vacuum tubes used like this were designed for demodulation of synchronous signals, an example of which is color demodulation for television receivers.



Modern 12AX7 miniature vacuum tube (manufactured in Russia)

The desire to include many functions in one envelope resulted in the General Electric Compactron. A typical unit, the 6AG11 Compactron tube contained two triodes and two diodes, but many in the series had triple triodes.

An early example of multiple devices in one envelope was the Loewe 3NF. This 1920s device had 3 triodes in a single glass envelope together with all the fixed capacitors and resistors required to make a complete radio receiver. As the Loewe set had only one tube socket, it was able to substantially undercut the competition since, in Germany, state tax was levied by the number of sockets. However, reliability was compromised, and production costs for the tube were much greater. In a sense, these were akin to integrated circuits. In the US, Cleartron briefly produced the "Multivalve" triple triode for use in the Emerson Baby Grand receiver. This Emerson set also had a single tube socket, but

because it used a four-pin base, the additional element connections were made on a "mezzanine" platform at the top of the tube base.

Loewe were to also offer the 2NF (two tetrodes plus passive components) and the WG38 (two pentodes, a triode and the passive components).



Vacuum tubes in an Australian radio of the late 1930s

The beam power tube is usually a tetrode with the addition of beam-forming electrodes, which take the place of the suppressor grid. These angled plates focus the electron stream onto certain spots on the anode which can withstand the heat generated by the impact of massive numbers of electrons, while also providing pentode behavior. The positioning of the elements in a beam power tube uses a design called "critical-distance geometry", which minimizes the "tetrode kink", plate-grid capacitance, screen-grid current, and secondary emission effects from the anode, thus increasing power conversion efficiency. The control grid and screen grid are also wound with the same pitch, or number of wires per inch.

Aligning the grid wires also helps to reduce screen current, which represents wasted energy. This design helps to overcome some of the practical barriers to designing high-power, high-efficiency power tubes. 6L6 was the first popular beam power tube, introduced by RCA in 1936. Corresponding tubes in Europe were the KT66, KT77 and KT88 by GEC (the KT standing for "Kinkless Tetrode").

Variations of the 6L6 design are still widely used in guitar amplifiers, making it one of the longest lived electronic device families in history. Similar design strategies are used in the construction of large ceramic power tetrodes used in radio transmitters.

Special-purpose tubes

Some special-purpose tubes are constructed with particular gases in the envelope. For instance, voltage regulator tubes contain various inert gases such as argon, helium or neon, and take advantage of the fact that these gases will ionize at predictable voltages. The thyatron is a special-purpose tube filled with low-pressure gas or mercury, some of which vaporizes. Like other tubes, it contains a hot cathode and an anode, but also a control electrode, which behaves somewhat like the grid of a triode. When the control electrode starts conduction, the gas ionizes, and the control electrode no longer can stop the current; the tube "latches" into conduction. Removing plate (anode) voltage lets the gas de-ionize, restoring its non-conductive state. Some thyatrons can carry large currents for their physical size. One example is the miniature type 2D21, often seen in 1950s jukeboxes as control switches for relays. A cold-cathode version of the thyatron, which uses a pool of mercury for its cathode, is called an Ignitron (tm). It can switch thousands of amperes in its largest versions. Thyatrons containing hydrogen have a very consistent time delay between their turn-on pulse and full conduction, and have long been used in radar transmitters. Thyatrons behave much like silicon-controlled rectifiers, or to be more chronologically precise, silicon controlled rectifiers mimic some of the behaviours of Thyatrons.

Tubes usually have glass envelopes, but metal, fused quartz (silica), and ceramic are possible choices. The first version of the 6L6 used a metal envelope sealed with glass beads, while a glass disk fused to the metal was used in later versions. Metal and ceramic are used almost exclusively for power tubes above 2 kW dissipation. The nuvistor is a tiny tube made only of metal and ceramic. In some power tubes, the metal envelope is also the anode. The 4CX1000A is an external anode tube of this sort. Air is blown through an array of fins attached to the anode, thus cooling it. Power tubes using this cooling scheme are available up to 150 kW dissipation. Above that level, water or water-vapor cooling are used. The highest-power tube currently available is the Eimac 4CM2500KG, a forced water-cooled power tetrode capable of dissipating 2.5 megawatts. (By comparison, the largest power transistor can only dissipate about 1 kilowatt.) Another very high power tube is the Eimac 8974, a 1.25 megawatt tetrode used in military and commercial radio-frequency installations.

An extremely specialized tube is the Krytron, which is used for extremely precise, rapid high-voltage switching. Due to their intended purpose, the initiation of the precise sequence of detonations used to set off a nuclear weapon, they are heavily controlled at an international level.



X-ray tube insert

Medical imaging equipment, such as radiographic and nuclear imaging, use special vacuum tubes. Radiographic, fluoroscopic, and CT X-ray imaging equipment use a specially designed vacuum tube diode, which has a rotating anode to dissipate the large amounts of heat developed during operation, and a focused cathode. They are housed in an aluminum housing which is filled with a dielectric oil. Nuclear imaging equipment uses photomultiplier tube arrays to detect radiation.

Miniature tubes

The miniature vacuum tube made tubes smaller by eliminating the Bakelite base. It was invented in 1938. Instead of the separate base, the pins are fused in the glass base of the envelope. This forces the sealing tip to the top of the envelope. Making tubes smaller reduced the voltage that they could work at, and also the power of the filament, so the older style continued to be used for high power rectifiers, valve amplifier output stages and certain transmitting tubes. Miniature tubes with a size roughly that of half a cigarette were used in hearing-aid amplifiers.

Development continued, and led to the sub miniature tubes and the "acorn" valve (named due to its shape).

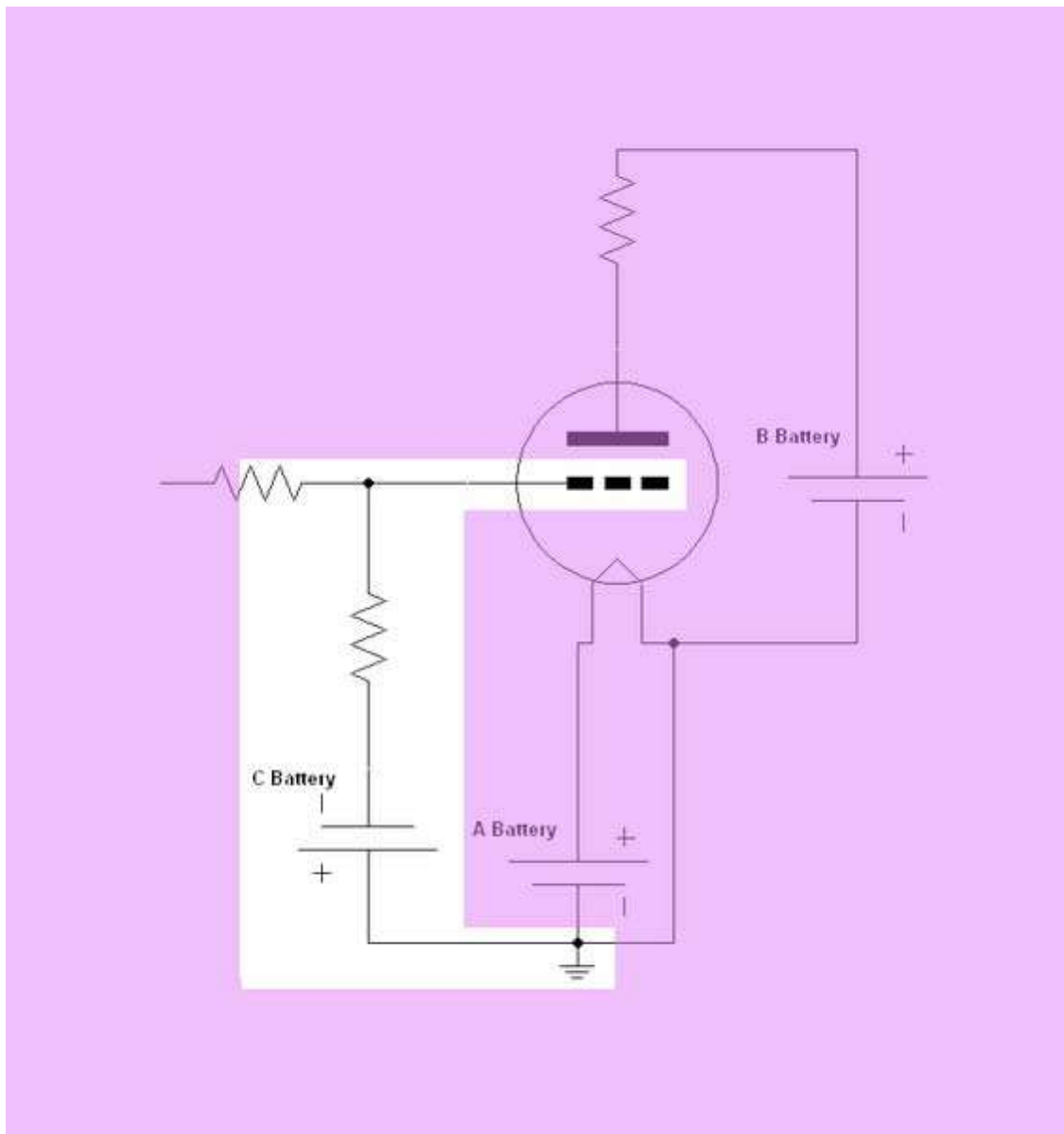
Powering the tube

Batteries

Batteries provided the voltages required by tubes in early radio sets. As many as three different voltages were required, using three different batteries. The "A" batteries or LT (low-tension) battery provided the filament voltage. Tube heaters were designed for single, double or triple-cell lead-acid batteries, giving nominal heater voltages of 2 V, 4 V or 6 V. In portable radios, dry batteries were sometimes used with 1.5 or 1 V heaters. Reducing filament consumption improved the life span of batteries. By 1955, receiving tubes using 50 mA down to as little as 10 mA for the heaters had been developed, but they were swept aside by development of the transistor.

The plate voltage was provided by "B" batteries or the HT (high-tension) supply or battery. These were generally of dry cell construction, containing many small 1.5 volt cells in series. They typically came in ratings of 22.5, 45, 67.5, 90 or 135 volts.

Some sets used a grid bias battery or "C" batteries, although many circuits used grid leak resistors, voltage dividers or cathode bias to provide proper tube bias. These batteries had very low drain.



Batteries for a vacuum tube circuit. The C battery is highlighted.

AC power

Replacement of batteries was a major cost of operation for early radio receiver users. The development of the battery eliminator, and, in 1925, batteryless receivers operated by household power, reduced operating costs and contributed to the growing popularity of radio. A power supply using a transformer with several windings, one or more rectifiers (which may themselves be vacuum tubes), and large filter capacitors provided the required direct current voltages from the alternating current source.

As a cost reduction measure, especially in high-volume consumer receivers, all the tube heaters could be connected in series across the AC supply, and the plate voltage derived from a half-wave rectifier directly connected to the AC input, eliminating the need for a heavy power transformer. As an additional feature, these radios could be operated on AC or DC mains. While this arrangement limited the plate voltage (and so, indirectly, the output power) that could be obtained, the resulting supply was adequate for many purposes. A filament tap on the rectifier tube provided the 6 volt, low current supply needed for a dial light. The so-called *series string* approach did have one safety defect: the chassis of the receiver was connected to one side of the power supply, presenting a shock hazard. Engineers reduced this hazard by enclosing the chassis in a plastic case, making the back out of particle board, and riveting the power cord chassis plug to the back so that consumers would not be able to power the radio while the chassis was accessible. (Technicians and tinkerers routinely bypassed this by using a separate cord, known colloquially as a "cheater cord" or "widowmaker.") Many consumer AM radio manufacturers of the era used a virtually identical circuit with the tube complement of 12BA6, 12BE6, 12AV6, 35W4, and 50C5, giving these radios the nickname All American Five or simply "Five Tube Radio." Although millions of such receivers were produced, they have now become collector's items.

Direct and indirect heating



Directly heated RS242 triodes

It became common to use the filament to heat a separate electrode called the cathode, and to use this cathode as the source of electron flow in the tube rather than the filament itself. This minimized the introduction of hum when the filament was energized with alternating current. In such tubes, the filament is called a heater to distinguish it as an inactive element. Development of vacuum tubes that could use alternating current for the heater supply allowed elimination of one rectifier element.

Reliability



Tube tester manufactured in 1930

One reliability problem of tubes with oxide cathodes is the possibility that the cathode may slowly become "poisoned" by gas molecules from other elements in the tube, which reduce its ability to emit electrons. Trapped gases or slow gas leaks can also damage the cathode or cause plate-current run away due to ionization of free gas molecules. Vacuum hardness and proper selection of construction materials are the major influences on tube lifetime. Depending on the material, temperature and construction, the surface material of the cathode may also diffuse onto other elements. The resistive heaters that heat the

cathodes may break in a manner similar to incandescent lamp filaments, but rarely do, since they operate at much lower temperatures than lamps.

The heater's failure mode is typically a stress-related fracture of the tungsten wire or at a weld point and generally occurs after accruing many thermal (power on-off) cycles. Tungsten wire has a very low resistance when at room temperature. A negative temperature coefficient device, such as a thermistor, may be incorporated in the equipment's heater supply or a ramp-up circuit may be employed to allow the heater or filaments to reach operating temperature more gradually than if powered-up in a step-function. Low-cost radios had tubes with heaters connected in series, with a total voltage equal to that of the line (mains). Following World War II, tubes intended to be used in series heater strings were redesigned to all have the same ("controlled") warm-up time. Earlier designs had quite-different thermal time constants. The audio output stage, for instance, had a larger cathode, and warmed up more slowly than lower-powered tubes. The result was that heaters that warmed up faster also temporarily had higher resistance, because of their positive temperature coefficient. This disproportionate resistance caused them to temporarily operate with heater voltages well above their ratings, and shortened their life.

Another important reliability problem is caused by air leakage into the tube. Usually oxygen in the air reacts chemically with the hot filament or cathode, quickly ruining it. Designers developed tube designs that sealed reliably. This was why most tubes were constructed of glass. Metal alloys (such as Cunife and Fernico) and glasses had been developed for light bulbs that expanded and contracted in similar amounts, as temperature changed. These made it easy to construct an insulating envelope of glass, while passing connection wires through the glass to the electrodes.

When a vacuum tube is overloaded or operated past its design dissipation, its anode (plate) may glow red. In consumer equipment, a glowing plate is universally a sign of an overloaded tube. However, some large transmitting tubes are designed to operate with their anodes at red, orange, or in rare cases, white heat.

Vacuum

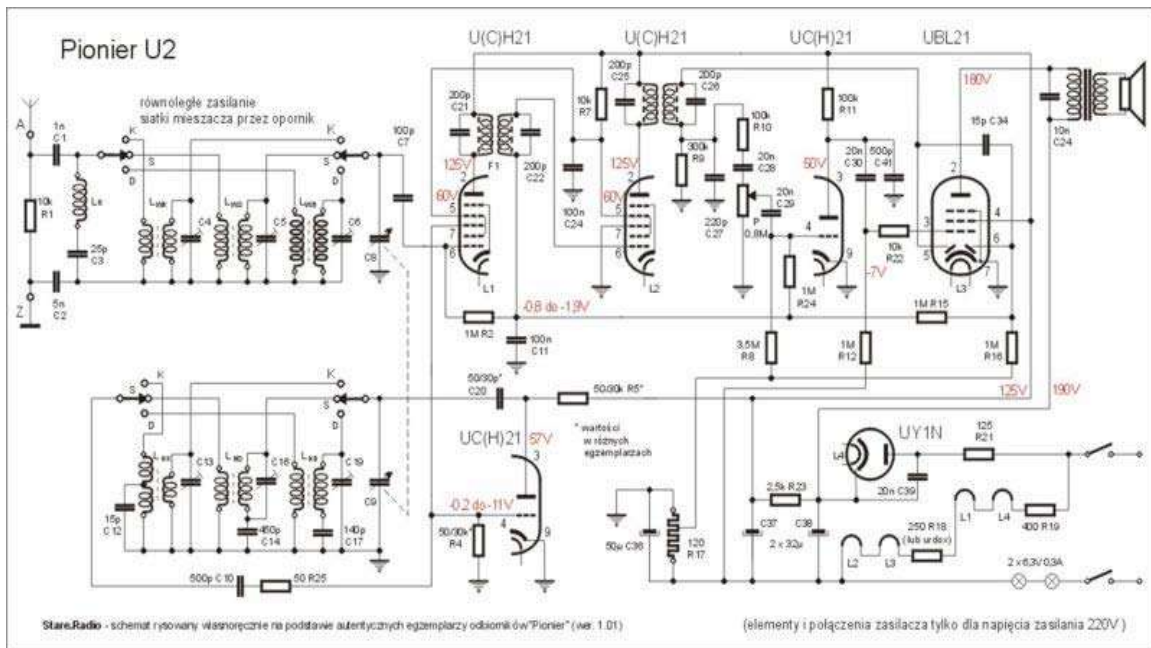


A valve in a modern guitar amplifier, glowing blue. This can indicate a "gassy" tube; however, this effect is commonly the result of electron bombardment of impurities in the glass envelope.

The vacuum inside the envelope must be as perfect, or "hard", as possible. Any gas atoms remaining might be ionized at operating voltages, and will conduct electricity between the elements in an uncontrolled manner. This can lead to erratic operation or even catastrophic destruction of the tube and associated circuitry. Unabsorbed free air sometimes ionizes and becomes visible as a pink-purple glow discharge between the tube elements.

To prevent any remaining gases from remaining in a free state in the tube, modern tubes are constructed with "getters", which are usually small, circular troughs filled with metals that oxidize quickly, with barium being the most common. While the tube envelope is being evacuated, the internal parts except the getter are heated by RF induction heating to extract any remaining gases from the metal. The tube is then sealed and the getter is heated to a high temperature, again by radio frequency induction heating. This causes the material to evaporate, absorbing/reacting with any residual gases and usually leaving a silver-colored metallic deposit on the inside of the envelope of the tube. The getter continues to absorb any gas molecules that leak into the tube during its working life. If a

tube develops a crack in the envelope, this deposit turns a white color when it reacts with atmospheric oxygen. Large transmitting and specialized tubes often use more exotic getter materials, such as zirconium. Early gettered tubes used phosphorus based getters and these tubes are easily identifiable, as the phosphorus leaves a characteristic orange or rainbow deposit on the glass. The use of phosphorus was short-lived and was quickly replaced by the superior barium getters. Unlike the barium getters, the phosphorus did not absorb any further gases once it had fired.



Schematic diagram of a 1948 vacuum tube radio from Poland

Transmitting tubes

Large transmitting tubes have carbonized tungsten filaments containing a small trace (1% to 2%) of thorium. An extremely thin (molecular) layer of thorium atoms forms on the outside of the wire's carbonized layer and, when heated, serve as an efficient source of electrons. The thorium slowly evaporates from the wire surface, while new thorium atoms diffuse to the surface to replace them. Such thoriated tungsten cathodes usually deliver lifetimes in the tens of thousands of hours. The end-of-life scenario for a thoriated-tungsten filament is when the carbonized layer has mostly been converted back into another form of tungsten carbide and emission begins to drop off rapidly; a complete loss of Thorium has never been found to be a factor in the end-of-life in a tube with this type of emitter. The highest reported tube life is held by an Eimac power tetrode used in a Los Angeles radio station's transmitter, which was removed from service after 80,000 hours (~9 years) of operation. It has been said that transmitters with vacuum tubes are better able to survive lightning strikes than transistor transmitters do. While it was commonly believed that at rf power levels above approx. 20 kilowatts, vacuum tubes were more efficient than solid state circuits, this is no longer the case especially in medium wave (AM broadcast) service where solid state transmitters at nearly all power

levels have measurably higher efficiency. FM broadcast transmitters with solid state PA's up to approx. 15 kW also show better overall mains-power efficiency than tube-based PA's and with the advent of newer, more efficient transistors this power level will certainly increase.

Receiving tubes

Cathodes in small "receiving" tubes are coated with a mixture of barium oxide and strontium oxide, sometimes with addition of calcium oxide or aluminium oxide. An electric heater is inserted into the cathode sleeve, and insulated from it electrically by a coating of aluminium oxide. This complex construction causes barium and strontium atoms to diffuse to the surface of the cathode when heated to about 780 degrees Celsius, thus emitting electrons.

Failure modes

Catastrophic failures

A catastrophic failure is one which suddenly makes the vacuum tube unusable. A crack in the glass envelope will allow air into the tube and destroy it. Cracks may result from stress in the glass, bent pins or impacts; tube sockets must allow for thermal expansion, to prevent stress in the glass at the pins. Stress may accumulate if a metal shield or other object presses on the tube envelope and causes differential heating of the glass. Glass may also be damaged by high-voltage arcing.

Tube heaters may also fail without warning, especially if exposed to over voltage or as a result of manufacturing defects. Tube heaters don't normally fail by evaporation like lamp filaments, since they operate at much lower temperature. The surge of inrush current when the heater is first energized causes stress in the heater, and can be avoided by slowly warming the heaters, gradually increasing current. Some tubes intended for series string operation of the heaters across the supply will have a definite controlled warm-up time to avoid excess voltage on some heaters as others warm up. Directly-heated filament-type cathodes as used in battery-operated tubes or some rectifiers may fail if the filament sags, causing internal arcing. Excess heater-to-cathode voltage in indirectly heated cathodes can break down the insulation between elements and destroy the heater.

Arcing between tube elements can destroy the tube. An arc can be caused by applying plate potential before the cathode has come up to operating temperature, or by drawing excess current through a rectifier which damages the emission coating. Arcs can also be initiated by any loose material inside the tube, or by excess screen voltage. An arc inside the tube allows gas to evolve from the tube materials, and may deposit conductive material on internal insulating spacers.

Degenerative failures

Degenerative failures cause the performance of the tube to slowly deteriorate with time.

Overheating of internal parts, such as control grids or mica spacer insulators, can result in trapped gas escaping into the tube; this can reduce performance. A getter is used to absorb gases evolved during tube operation, but has only a limited ability to combine with gas. Control of the envelope temperature prevents some types of gassing. A tube with very bad internal gas may have a visible blue glow when plate voltage is applied.

Gas and ions within the tube contribute to grid current which can disturb operation of a vacuum tube circuit. Another effect of overheating is the slow deposit of metallic vapors on internal spacers, resulting in inter-element leakage.

Tubes on standby for long periods, with heater voltage applied, may develop high cathode interface resistance and display poor emission characteristics. This effect occurred especially in pulse and digital circuits, where tubes had no plate current flowing for extended times.

Cathode depletion describes the loss of emission after thousands of hours of normal use. Sometimes emission can be restored for a time by raising heater voltage either for a short time or a permanent increase of a few percent. Cathode depletion was uncommon in signal tubes but was a frequent cause of failures of monochrome television cathode-ray tubes.

Other failures

Vacuum tubes may have or develop defects in operation that makes an individual tube useless in one device, but which may not prevent its satisfactory operation in another system. *Microphonics* refers to internal vibration of tube elements, which modulates the signal from the tube in an undesirable way; sound or vibration pick-up may affect the signals, or even cause uncontrolled howling if a feedback path develops between a microphonic tube and, for example, a loudspeaker. Leakage current between AC heaters and the cathode may couple into the circuit, or electrons emitted directly from the ends of the heater may also inject hum into the signal. Leakage current due to internal contamination may also inject noise.



Tubes from a 1950s computer

Computer vacuum tubes

Colossus

Colossus (and its successor Colossus Mk2) was built by the British during World War II to substantially speed up the task of breaking the German high level Lorenz encryption. Based on 1500 vacuum tubes, Colossus replaced an earlier machine based on relay and switch logic (the Heath Robinson). Colossus was able to break in a matter of hours messages that had previously taken several weeks. Colossus Mk2 used a total of around 2000 vacuum tubes. Colossus was the first ever use of vacuum tubes on such a large scale for a single machine. The largest project previously had used just 150 tubes and had proven to be extremely unreliable. The main design problem at Colossus's inception was how to make vacuum tube based equipment reliable when the tubes were used in large numbers.

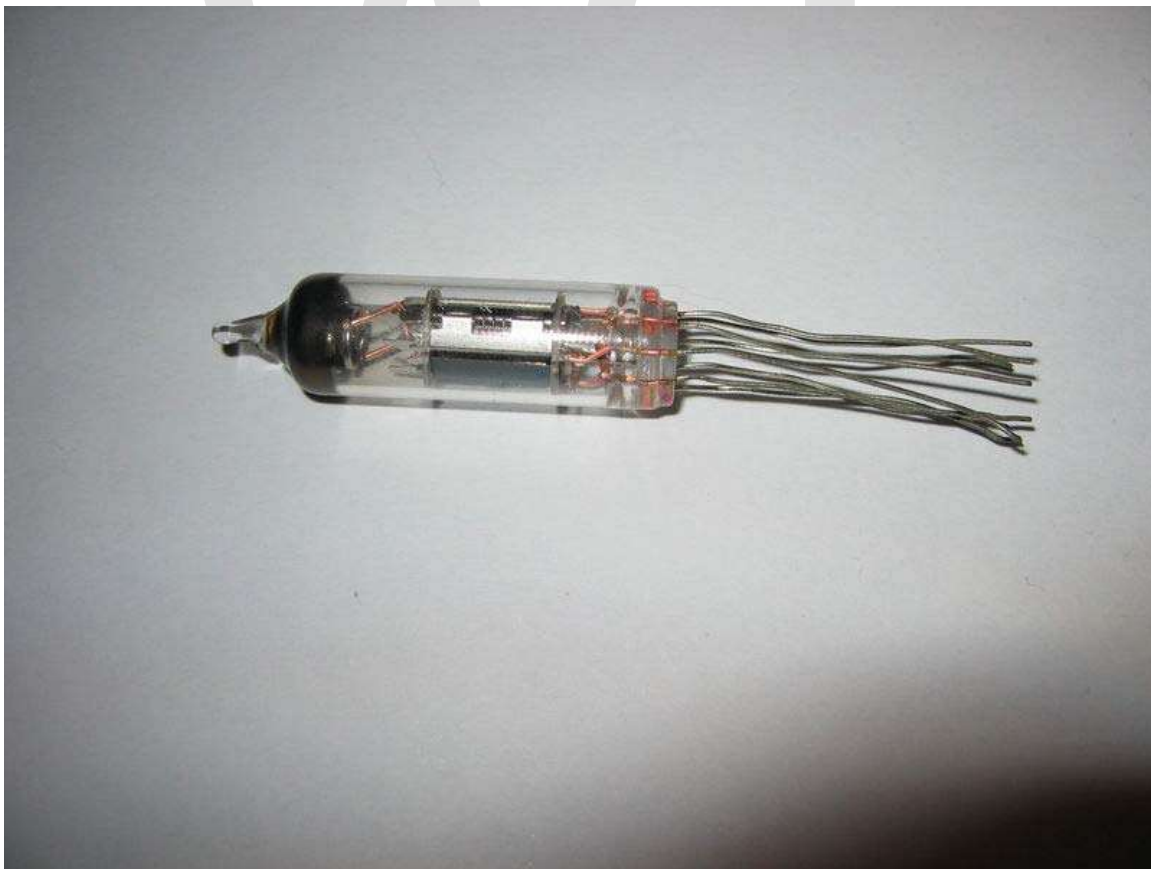
The Colossus computer's designer, Dr. Tommy Flowers, had a theory that most of the unreliability was caused during power down and (mainly) power up. Once Colossus was built and installed, it was switched on and left switched on running from dual redundant diesel generators (the wartime mains supply being considered too unreliable). The only time it was switched off was for conversion to the Colossus Mk2 and the addition of another 500 or so tubes. Another 9 Colossus Mk2s were built, and all 10 machines ran

with a surprising degree of reliability. The 10 Colossi consumed 15 kilowatts of power each, 24 hours a day, 365 days a year—nearly all of it for the tube heaters.

Whirlwind

To meet the reliability requirements of the early digital computer Whirlwind, it was necessary to build special "computer vacuum tubes" with extended cathode life. The problem of short lifetime was traced to evaporation of silicon, used in the tungsten alloy to make the heater wire easier to draw. Elimination of the silicon from the heater wire alloy (and paying extra for more frequent replacement of the wire drawing dies) allowed production of tubes that were reliable enough for the Whirlwind project. The tubes developed for Whirlwind later found their way into the giant SAGE air-defense computer system. High-purity nickel tubing and cathode coatings free of materials that can poison emission (such as silicates and aluminium) also contribute to long cathode life. The first such "computer tube" was Sylvania's 7AK7 of 1948. By the late 1950s it was routine for special-quality small-signal tubes to last for hundreds of thousands of hours, if operated conservatively. This reliability made mid-cable amplifiers in submarine cables possible.

World War II



A CV4501 subminiature tube for use in a military radio set. The tube is a special quality type based on the EF72, 35 mm long and 10 mm in diameter (excluding leads).

Near the end of World War II, to make radios more rugged, some aircraft and army radios began to integrate the tube envelopes into the radio's cast aluminium or zinc chassis. The radio became just a printed circuit with non-tube components, soldered to the chassis that contained all the tubes. During WWII in 1942, rugged metal vacuum tubes were mounted in anti-aircraft shells. These proximity fuzes made anti-aircraft shells 6 times more effective. In the fall of 1944, artillery shells with proximity fuses were used. The tiny tubes were later known as "subminiature" types. They were widely used in 1950s military and aviation electronics.

Applications

Vacuum tubes were critical to the development of electronic technology, which drove the expansion and commercialization of radio broadcasting, television, radar, sound reproduction, large telephone networks, analog and digital computers, and industrial process control. Some of these applications pre-dated electronics, but it was the vacuum tube that made them widespread and practical.

Tubes were heavily used in the early generations of electronic devices, such as radios, televisions, and early computers such as the Colossus which used 2000 tubes, the ENIAC which used nearly 18,000 tubes, and the IBM 700 series.

For most purposes, the vacuum tube has been replaced by solid-state devices such as transistors and solid-state diodes. Solid-state devices last much longer, are smaller, more efficient, more reliable, and cheaper than equivalent vacuum tube devices. However, tubes are still used in specialized applications: for engineering reasons, as in high-power radio frequency transmitters; or for their aesthetic appeal and distinct sound signature, as in audio amplification. Cathode ray tubes are still used as display devices in television sets, video monitors, and oscilloscopes, although they are being replaced by LCDs and other flat-panel displays. A specialized form of the electron tube, the magnetron, is the source of microwave energy in microwave ovens and some radar systems. The klystron, a powerful but narrow-band radio-frequency amplifier, is commonly deployed by broadcasters as a high-power UHF television transmitter.

Vacuum tubes are less susceptible than corresponding solid-state components to the electromagnetic pulse effect of nuclear explosions. This property kept them in use for certain military applications long after transistors had replaced them elsewhere. Vacuum tubes are still used for very high-powered applications such as industrial radio-frequency heating, generating large amounts of RF energy for particle accelerators, and power amplification for broadcasting. In microwave ovens, cost-engineered magnetrons efficiently generate microwave power on the order of hundreds of watts.



Russian 12AX7 tubes inside a modern guitar amplifier

Many audiophiles, professional audio engineers, and musicians prefer the tube sound of audio equipment based on vacuum tubes over electronics based on transistors. There are companies which still make specialized audio hardware featuring tube technology. A common usage is in the high-end microphone preamplifiers preferred by professional music recording studios, and in electric guitar amplification. The sound produced by a tube based amplifier with the tubes overloaded (overdriven) has defined the texture of some genres of music such as classic rock and blues. Guitarists often prefer tube amplifiers for the warmth of their tone and the natural compression effect they can apply to an input signal.

Cathode-ray tubes (CRTs) are a highly-evolved type of vacuum tube, described elsewhere.

In 2002, computer motherboard maker AOpen brought back the vacuum tube for modern computer use by releasing the AX4GE Tube-G motherboard. This motherboard uses a Sovtek 6922 vacuum tube (a version of the 6DJ8) as part of AOpen's TubeSound Technology. AOpen claims that the vacuum tube brings superior sound.

Cooling



A 12AX7 preamp tube, EL84 power tube (with dampener rings) and a GZ34 rectifier tube in a guitar amplifier.

Like any electronic device, vacuum tubes produce heat while operating. This waste heat is one of the principal factors that affect tube life. The majority of this waste heat originates in the anode though some grids may also require cooling to remove excess heat. For example, the cooling of the screen grid in an EL34 is facilitated by the addition of two small radiators or "wings," located near the top of the tube. The heater (filament) also contributes to the total waste heat. A tube's data sheet will normally identify the maximum amount of heat each element may dissipate.

The method of anode cooling is dependent on the construction of the tube itself. For tubes with internal anodes such as the 12AX7 or EL34, the cooling occurs by radiating the heat by black body radiation from the anode to the glass envelope. Natural air circulation, convection, then removes the heat from the envelope. Tube shields that aided heat dispersal could be retrofitted on certain select types of tubes. These shields act by improving heat conduction from the surface of the tube to the shield itself by means of tens of copper tongues in contact with the glass tube, and have an opaque, black outside finish for improved heat radiation. The ability to remove heat may be further increased by implementing forced air cooling, adding fins to the anode, and operating the anode at red hot temperatures. All of these measures are implemented in the 4-1000A transmitting tube.

The amount of heat that may be removed from a tube with an internal anode is limited. Tubes with external anodes may be cooled using forced air, water, vapor, and multiphase. The 3CX10,000A7 is an example of a tube with an external anode cooled by forced air. The water, vapor, and multiphase cooling techniques all depend on the high specific heat and latent heat of water. The 8974 is an example of a water cooled tube and is among the largest commercial tube available today.

In a water cooled tube, the anode voltage appears directly on the cooling water surface, thus requiring the water to be an electrical insulator. Otherwise the high voltage can be conducted through the cooling water to the radiator system; hence the need for deionized water. Such systems usually have a built-in water-conductance monitor which will shut down the high-tension supply (often tens of kilovolts) if the conductance becomes too high.

Other vacuum tube devices

Many devices were built during the 1920–1960 period using vacuum-tube techniques. Most such tubes were rendered obsolete by semiconductors; some techniques for integrating multiple devices in a single module, sharing the same glass envelope have been discussed above, such as the Loewe 3NF. Vacuum-tube electronic devices still in common use include the magnetron, klystron, photomultiplier, x-ray tube, traveling-wave tube and cathode ray tube. The magnetron is the type of tube used in all microwave ovens. In spite of the advancing state of the art in power semiconductor technology, the vacuum tube still has reliability and cost advantages for high-frequency RF power generation. Photomultipliers are still the most sensitive detectors of light. Many televisions, oscilloscopes and computer monitors still use cathode ray tubes, though flat panel displays are becoming more popular as prices drop.

Many of the better tube radios had so-called "tuning eye" indicator tubes behind their front panels, with just the top of the tube showing. These tubes were used as a visual indication of received signal strength, and an aid to properly tuning in a station.

Secondary emission is the term for what happens when electrons in a vacuum strike certain materials, and the impacts cause electrons to be emitted. For some materials, more electrons are emitted than originally hit the surface. Such devices, called electron multipliers, amplify the current represented by the incoming electrons. Several stages (as many as 15 or so) can be cascaded for high gain, and are essential parts of very sensitive phototubes, usually called photomultipliers or multiplier phototubes. The image orthicon TV studio camera tubes also used multistage photomultipliers.

For decades, electron-tube designers tried to use secondary emission to obtain more amplification in vacuum tubes with hot cathodes, but they suffered from short life because the material used for the secondary-emission electrode (called a dynode) "poisoned" the tube's hot cathode. (For instance, the interesting RCA 1630 secondary-emission tube was marketed, but did not last.) However, eventually, Philips of The Netherlands developed the EFP60 tube that had a satisfactory lifetime, and was used in at

least one product, a laboratory pulse generator. However, transistors were rapidly improving, and eclipsed tubes in general.

A variant, called a channel electron multiplier, is a curved tube, such as a helix, coated on the inside with material with good secondary emission. One type had a little funnel to capture incoming electrons. The tube was resistive, and its ends were connected to enough voltage to create repeated cascades of electrons.

Tektronix made a high-performance wideband oscilloscope CRT with a channel electron multiplier plate behind the phosphor layer. This plate was a bundled array of a huge number of short individual c.e.m. tubes that accepted a low-current beam and intensified it to provide a display of practical brightness. (The electron optics of the wideband electron gun could not provide enough current to directly excite the phosphor.)

The fluorescent displays commonly used on videocassette recorders, some microwave oven control panels, and automotive dashboards are vacuum tubes, using phosphor-coated anodes to form the display characters, and a heated filamentary cathode as an electron source. These are referred to as "VFDs", or vacuum fluorescent displays. Because the filaments are in view, they must be operated at temperatures where the filament does not glow visibly. Often found in automotive applications, their high brightness allows reading the display in daylight. VFD tubes are flat and rectangular, as well as relatively thin. Typical VFD phosphors emit a broad-spectrum greenish-white light, permitting use of color filters. This type of phosphor provides a bright glow with only modest operating voltage, low tens of volts.

Some tubes, such as magnetrons, traveling-wave tubes, carcinotrons, and klystrons, combine magnetic and electrostatic effects. These are efficient (usually narrow-band) RF producers and still find use in radar, microwave ovens and industrial heating. Traveling-wave tubes (TWTs) are very good amplifiers; they are used in some communications satellites. High-powered klystron amplifier tubes can provide hundreds of kilowatts in the UHF range.

Gyrotrons or vacuum masers, used to generate high-power millimetre band waves, are magnetic vacuum tubes in which a small relativistic effect, due to the high voltage, is used for bunching the electrons. Gyrotrons can generate very high powers (hundreds of kilowatts). Free electron lasers, used to generate high-power coherent light and perhaps even X rays, are highly relativistic vacuum tubes driven by high-energy particle accelerators.

Particle accelerators can be considered vacuum tubes that work backward, the electric fields driving the electrons, or other charged particles. In this respect, a cathode ray tube is a particle accelerator.

A tube in which electrons move through a vacuum (or gaseous medium) within a gas-tight envelope is generically called an *electron tube*.

Some condenser microphone designs use built-in vacuum tube preamplifiers.

As of 2008, scores of small companies are manufacturing audiophile amplifiers and preamps that use vacuum tubes.

Vacuum tubes using field electron emitters

In the early years of the 21st century there has been renewed interest in vacuum tubes, this time with the electron emitter formed on a flat silicon substrate, as in integrated circuit technology. This subject is now called vacuum nanoelectronics. The most common design uses a cold cathode in the form of a large-area field electron source (for example a field emitter array). With these devices, electrons are field-emitted from a large number of closely spaced individual emission sites.

Their claimed advantages include greatly enhanced robustness combined with the ability to provide high power outputs at low power consumptions. Operating on the same principles as traditional tubes, prototype device cathodes have been fabricated in several different ways. Although a common approach is to use a field emitter array, one interesting idea is to etch electrodes to form hinged flaps – similar to the technology used to create the microscopic mirrors used in Digital Light Processing) that are stood upright by an electrostatic charge.

Such integrated microtubes may find application in microwave devices including mobile phones, for Bluetooth and Wi-Fi transmission, in radar and for satellite communication. Presently they are being studied for possible applications in field emission display technology, but significant production problems seem to exist.

Chapter- 6

Vacuum State

In quantum field theory, the **vacuum state** (also called the **vacuum**) is the quantum state with the lowest possible energy. Generally, it contains no physical particles. **Zero-point field** is sometimes used as a synonym for the vacuum state of an individual quantized field.

According to present-day understanding of what is called the vacuum state or the quantum vacuum, it is "by no means a simple empty space", and again: "it is a mistake to think of any physical vacuum as some absolutely empty void." According to quantum mechanics, the vacuum state is not truly empty but instead contains fleeting electromagnetic waves and particles that pop into and out of existence.

The QCD vacuum of quantum chromodynamics is the object of study in the Relativistic Heavy Ion Collider and the Large Hadron Collider, and is related to the so-called *vacuum structure of strong interactions*.

Non-vanishing vacuum state

If the quantum field theory can be accurately described through perturbation theory, then the properties of the vacuum are analogous to the properties of the ground state of a quantum mechanical harmonic oscillator (or more accurately, the ground state of a QM problem). In this case the vacuum expectation value (VEV) of any field operator vanishes. For quantum field theories in which perturbation theory breaks down at low energies (for example, Quantum chromodynamics or the BCS theory of superconductivity) field operators may have non-vanishing vacuum expectation values called condensates. In the Standard Model, the non-zero vacuum expectation value of the Higgs field, arising from spontaneous symmetry breaking, is the mechanism by which the other fields in the theory acquire mass.

The energy of the vacuum state

Vacuum energy is an underlying background energy that exists in space even when the space is devoid of matter (free space). The concept of vacuum energy has been deduced from the concept of virtual particles, which is itself derived from the energy-time uncertainty principle. The effects of vacuum energy can be observed in various

phenomena such as spontaneous emission, the Casimir effect, the van der Waals bonds and the Lamb shift, and are thought to influence the behavior of the Universe on cosmological scales. Using the upper limit of the cosmological constant, the vacuum energy in a cubic centimeter of free space has been estimated to be 10^{-15} Joules. However, in both Quantum Electrodynamics (QED) and Stochastic Electrodynamics (SED), consistency with the principle of Lorentz invariance and with the magnitude of the Planck Constant requires it to have a much larger value of 10^{107} Joules per cubic centimeter.

Origin

Quantum field theory states that all fundamental fields, such as the electromagnetic field, must be quantized at each and every point in space. A field in physics may be envisioned as if space were filled with interconnected vibrating balls and springs, and the strength of the field were like the displacement of a ball from its rest position. The theory requires "vibrations" in, or more accurately changes in the strength of, such a field to propagate as per the appropriate wave equation for the particular field in question. The second quantization of quantum field theory requires that each such ball-spring combination be quantized, that is, that the strength of the field be quantized at each point in space. Canonically, if the field at each point in space is a simple harmonic oscillator, its quantization places a quantum harmonic oscillator at each point. Excitations of the field correspond to the elementary particles of particle physics. Thus, according to the theory, even the vacuum has a vastly complex structure and all calculations of quantum field theory must be made in relation to this model of the vacuum.

The theory considers vacuum to implicitly have the same properties as a particle, such as spin or polarization in the case of light, energy, and so on. According to the theory, most of these properties cancel out on average leaving the vacuum empty in the literal sense of the word. One important exception, however, is the vacuum energy or the vacuum expectation value of the energy. The quantization of a simple harmonic oscillator requires the lowest possible energy, or zero-point energy of such an oscillator to be:

$$E = \frac{1}{2}\hbar\omega$$

Summing over all possible oscillators at all points in space gives an infinite quantity. To remove this infinity, one may argue that only differences in energy are physically measurable, much as the concept of potential energy has been treated in classical mechanics for centuries. This argument is the underpinning of the theory of renormalization. In all practical calculations, this is how the infinity is handled.

Vacuum energy can also be thought of in terms of virtual particles (also known as vacuum fluctuations) which are created and destroyed out of the vacuum. These particles are always created out of the vacuum in particle-antiparticle pairs, which shortly annihilate each other and disappear. However, these particles and antiparticles may interact with others before disappearing, a process which can be mapped using Feynman diagrams. Note that this method of computing vacuum energy is mathematically

equivalent to having a quantum harmonic oscillator at each point and, therefore, suffers the same renormalization problems.

Additional contributions to the vacuum energy come from spontaneous symmetry breaking in quantum field theory.

Implications

Vacuum energy has a number of consequences. In 1948, Dutch physicists Hendrik B. G. Casimir and Dirk Polder predicted the existence of a tiny attractive force between closely placed metal plates due to resonances in the vacuum energy in the space between them. This is now known as the Casimir effect and has since been extensively experimentally verified. It is therefore believed that the vacuum energy is "real" in the same sense that more familiar conceptual objects such as electrons, magnetic fields, etc., are real.

Other predictions are more esoteric and harder to verify. Vacuum fluctuations are always created as particle/antiparticle pairs. The creation of these virtual particles near the event horizon of a black hole has been hypothesized by physicist Stephen Hawking to be a mechanism for the eventual "evaporation" of black holes. The net energy of the Universe remains zero so long as the particle pairs annihilate each other within Planck time. If one of the pair is pulled into the black hole before this, then the other particle becomes "real" and energy/mass is essentially radiated into space from the black hole. This loss is cumulative and could result in the black hole's disappearance over time. The time required is dependent on the mass of the black hole but could be on the order of 10^{100} years for large solar-mass black holes.

The vacuum energy also has important consequences for physical cosmology. Special relativity predicts that energy is equivalent to mass, and therefore, if the vacuum energy is "really there", it should exert a gravitational force. Essentially, a non-zero vacuum energy is expected to contribute to the cosmological constant, which affects the expansion of the universe. In the special case of vacuum energy, general relativity stipulates that the gravitational field is proportional to $\rho - 3p$ (where ρ is the mass-energy density, and p is the pressure). Quantum theory of the vacuum further stipulates that the pressure of the zero-state vacuum energy is always negative and equal to ρ . Thus, the total of $\rho - 3p$ becomes -2ρ : A negative value. This calculation implies a repulsive gravitational field, giving rise to expansion, if indeed the vacuum ground state has non-zero energy. However, the vacuum energy is mathematically infinite without renormalization, which is based on the assumption that we can only measure energy in a relative sense, which is not true if we can observe it indirectly via the cosmological constant.

The existence of vacuum energy is also sometimes used as theoretical justification for the possibility of free energy machines. It has been argued that due to the broken symmetry (in QED), free energy does not violate conservation of energy, since the laws of thermodynamics only apply to equilibrium systems. However, consensus amongst physicists is that this is incorrect and that vacuum energy cannot be harnessed to generate

free energy. In particular, the second law of thermodynamics is unaffected by the existence of vacuum energy. Moreover, in Stochastic Electrodynamics, the energy density is taken to be a classical random noise wave field which consists of real electromagnetic waves propagating isotropically in all directions. The energy in such a wave field would seem to be accessible e.g. with nothing more complicated than a directional coupler. The most obvious difficulty appears to be the spectral distribution of the energy, which compatibility with Lorentz invariance requires to take the form Kf^3 , where K is a constant and f denotes frequency. It follows that the energy and momentum flux in this wave field only becomes significant at extremely short wavelengths where directional coupler technology is currently lacking.

History

In 1934, Georges Lemaître used an unusual perfect-fluid equation of state to interpret the cosmological constant as due to vacuum energy. In 1948, the Casimir effect provided the first experimental verification of the existence of vacuum energy. In 1957, Lee and Yang proved the concepts of broken symmetry and parity violation, for which they won the Nobel prize. In 1973, Edward Tryon proposed that the Universe may be a large-scale quantum-mechanical vacuum fluctuation where positive mass-energy is balanced by negative gravitational potential energy. During the 1980s, there were many attempts to relate the fields that generate the vacuum energy to specific fields that were predicted by Grand unification theory and to use observations of the Universe to confirm that theory. However, the exact nature of the particles or fields that generate vacuum energy, with a density such as that required by inflation theory, remains a mystery.

The symmetry of the vacuum state

For a relativistic field theory, the vacuum is Poincaré invariant. Poincaré invariance implies that only scalar combinations of field operators have non-vanishing VEV's. The VEV may break some of the internal symmetries of the Lagrangian of the field theory. In this case the vacuum has less symmetry than the theory allows, and one says that *spontaneous symmetry breaking* has occurred.

Electrical permittivity of vacuum state

In principle, quantum corrections to Maxwell's equations can cause the experimental electrical permittivity ϵ of the vacuum state to deviate from the defined scalar value ϵ_0 of the electric constant. These theoretical developments are described, for example, in Dittrich and Gies. In particular, the theory of quantum electrodynamics predicts that vacuum should exhibit nonlinear effects that will make it behave like a birefringent material with ϵ slightly greater than ϵ_0 for extremely strong electric fields. Explanations for dichroism from particle physics, outside quantum electrodynamics, also have been proposed. Active attempts to measure such effects have been unsuccessful so far.

Notations

The vacuum state is written as $|0\rangle$ or $|\rangle$. The VEV of a field ϕ , which should be written as $\langle 0|\phi|0\rangle$, is usually condensed to $\langle\phi\rangle$.

Virtual particles

The uncertainty principle in the form $\Delta E \Delta t \geq \hbar$ implies that in the vacuum one or more particles with energy ΔE above the vacuum may be created for a short time Δt . These *virtual particles* are included in the definition of the vacuum.

QCD vacuum

The **QCD vacuum** is the vacuum state of quantum chromodynamics (QCD). It is an example of a *non-perturbative* vacuum state, characterized by many non-vanishing condensates such as the gluon condensate or the quark condensate. These condensates characterize the **normal phase** or the **confined phase** of quark matter.

Symmetries and symmetry breaking

Symmetries of the QCD Lagrangian

Like any relativistic quantum field theory, QCD enjoys Poincaré symmetry including the discrete symmetries CPT (each of which is realized). Apart from these space-time symmetries, it also has internal symmetries. Since QCD is an SU(3) gauge theory, it has local SU(3) gauge symmetry.

Since it has many flavours of quarks, it has approximate flavour and chiral symmetry. This approximation is said to involve the **chiral limit** of QCD. Of these chiral symmetries, the baryon number symmetry is exact. Some of the broken symmetries include the axial U(1) symmetry of the flavour group. This is broken by the chiral anomaly. The presence of instantons implied by this anomaly, also breaks CP symmetry.

In summary, the QCD Lagrangian has the following symmetries:

- Poincare symmetry and CPT invariance
- SU(3) local gauge symmetry
- approximate global SU(N_f)XSU(N_f) flavour chiral symmetry and the U(1) baryon number symmetry

The following classical symmetries are broken in the QCD Lagrangian:

- scale, i.e., conformal symmetry (through the scale anomaly), giving rise to asymptotic freedom
- the axial part of the U(1) flavour chiral symmetry (through the chiral anomaly), giving rise to the strong CP problem.

Spontaneous symmetry breaking

When the Hamiltonian of a system (or the Lagrangian) has a certain symmetry, but the ground state (i.e., the vacuum) does not, then one says that **spontaneous symmetry breaking** (SSB) has taken place.

A familiar example of SSB is in magnetic materials. Microscopically, the material consists of atoms with a non-vanishing spin, each of which acts like a tiny bar magnet, i.e., a magnetic dipole. The Hamiltonian of the material, describing the interaction of neighbouring dipoles, is invariant under rotations. At high temperature, there is no magnetization of a large sample of the material. Then one says that the symmetry of the Hamiltonian is realized by the system. However, at low temperature, there could be an overall magnetization. This magnetization has a *preferred direction*, since one can tell the north magnetic pole of the sample from the south magnetic pole. In this case, there is spontaneous symmetry breaking of the rotational symmetry of the Hamiltonian.

When a continuous symmetry is spontaneously broken, massless bosons appear, corresponding to the remaining symmetry. This is called the **Goldstone phenomenon** and the bosons are called Goldstone bosons.

Symmetries of the QCD vacuum

The $SU(N_f) \times SU(N_f)$ chiral flavour symmetry of the QCD Lagrangian is broken in the vacuum state of the theory. The symmetry of the vacuum state is the diagonal $SU(N_f)$ part of the chiral group. The diagnostic for this is the formation of a non-vanishing chiral condensate $\langle \bar{\psi}_i \psi_i \rangle$, where ψ_i is the quark field operator, and the flavour index i is summed. The Goldstone bosons of the symmetry breaking are the pseudoscalar mesons.

When $N_f=2$, i.e., only the **u** and **d** quarks are treated as massless, the three pions are the Goldstone bosons. When the **s** quark is also treated as massless, i.e., $N_f=3$, all eight pseudoscalar mesons of the quark model become Goldstone bosons. The actual masses of these mesons are obtained in chiral perturbation theory through an expansion in the (small) actual masses of the quarks.

In other phases of quark matter the full chiral flavour symmetry may be recovered, or broken in completely different ways.

Evidence: experimental consequences

The evidence for QCD condensates comes from two eras, the pre-QCD era 1950-1973 and the post-QCD era, after 1974. The pre-QCD results established that the strong interactions vacuum contains a quark chiral condensate, while the post-QCD results established that the vacuum also contains a gluon condensate.

Pre-QCD: gradient coupling

In the 1950s, there were many attempts to produce a field theory to describe the interactions of pions and nucleons. The obvious renormalizable interaction between the two objects is the Yukawa coupling to a pseudoscalar:

$$L_I = \bar{N}\gamma_5\pi N$$

And this is clearly theoretically correct, since it is leading order and it takes all the symmetries into account. But it doesn't match experiment. The interaction that does couples the nucleons to the *gradient* of the pion field.

$$g\bar{N}\gamma^\mu\partial_\mu\pi N$$

This is the **gradient-coupling model**. This interaction has a very different dependence on the energy of the pion—it vanishes at zero momentum.

This type of coupling means that a coherent state of low momentum pions barely interacts at all. This is a manifestation of an approximate symmetry, a **shift symmetry** of the pion field. The replacement

$$\pi \rightarrow \pi + C$$

leaves the gradient coupling alone, but not the pseudoscalar coupling.

The modern explanation for the shift symmetry was first proposed by Yoichiro Nambu. The pion field is a Goldstone boson, and the shift symmetry is the lowest order approximation to moving along the flat directions.

Pre-QCD: Goldberger-Treiman relation

There is a mysterious relationship between the strong interaction coupling of the pions to the nucleons, the coefficient g in the gradient coupling model, and the axial vector current coefficient of the nucleon which determines the weak decay rate of the neutron. The relation is

$$g_{\pi NN}F_\pi = G_A M_N$$

and it is obeyed to 1% accuracy.

The constant G_A is the coefficient that determines the neutron decay rate. It gives the normalization of the weak interaction matrix elements for the nucleon. On the other hand, the pion-nucleon coupling is a phenomenological constant describing the scattering of bound states of quarks and gluons.

The weak interactions are current-current interactions ultimately because they come from a nonabelian gauge theory. The Goldberger Treiman relation suggests that the pions for some reason interact as if they are related to the same symmetry current.

PCAC

The phenomena which gives rise to the Goldberger Treiman relation was called the "Partially Conserved Axial Current" hypothesis, or PCAC. Partially conserved is an archaic term for spontaneously broken, and the axial current is now called the chiral symmetry current.

The idea is that the symmetry current which performs axial rotations on the fundamental fields does not preserve the vacuum. This means that the current J applied to the vacuum produces particles. The particles must be scalars, otherwise the vacuum wouldn't be Lorentz invariant. By index matching, the matrix element is:

$$J_\mu |0\rangle = k_\mu |\pi\rangle,$$

where k_μ is the momentum carried by the created pion. Since the divergence of the axial current operator is zero, we must have

$$\partial_\mu J^\mu |0\rangle = k^\mu k_\mu |\pi\rangle = m_\pi^2 |\pi\rangle = 0.$$

Hence the pions are massless, $m_\pi^2 = 0$, in accordance with Goldstone's theorem.

Now if the scattering matrix element is considered, we have

$$k_\mu \langle N(p) | \pi(k) N(p') \rangle = \langle N(p) | J_\mu | N(p') \rangle.$$

Up to a momentum factor, which is the gradient in the coupling, it takes the same form as the axial current turning a neutron into a proton in the current-current form of the weak interaction.

$$\langle N | J^\mu | N \rangle \langle e | J_\mu | \nu \rangle$$

Pre-QCD: soft pion emission

Extensions of the PCAC ideas allowed Steven Weinberg to calculate the amplitudes for collisions which emit low energy pions from the amplitude for the same process with no pions. The amplitudes are those given by acting with symmetry currents on the external particles of the collision.

These successes established the basic properties of the strong interaction vacuum well before QCD.

Pseudo-Goldstone bosons

Experimentally it is seen that the masses of the octet of pseudoscalar mesons is very much lighter than the next lightest states; i.e., the octet of vector mesons (such as the rho). The most convincing evidence for SSB of the chiral flavour symmetry of QCD is the appearance of these pseudo-Goldstone bosons. These would have been strictly massless in the chiral limit. There is convincing demonstration that the observed masses are compatible with chiral perturbation theory. The internal consistency of this argument is further checked by lattice QCD computations allow one to vary the quark mass and check that the variation of the pseudoscalar masses with the quark mass is as required by chiral perturbation theory.

The η'

This pattern of SSB solves one of the mysteries of the quark model where all the pseudoscalar mesons should have been of nearly the same mass. Since $N_f=3$, there should have been nine of these. However, one (the SU(3) singlet η') has quite a different mass from the SU(3) octet. In the quark model this has no natural explanation— a mystery named the **η - η' mass splitting** (the η is one member of the octet which should have been degenerate in mass with the η'). In QCD one realizes that the η' is associated with the axial U(1) which is broken through the chiral anomaly and not by SSB. One says therefore, that instantons cause the η - η' mass splitting.

Current algebra and QCD sum rules

PCAC and current algebra also provide evidence for this pattern of SSB. Direct estimates of the chiral condensate also comes from such analysis.

Another method of analysis of correlation functions in QCD is through an operator product expansion (OPE). This writes the vacuum expectation value of a non-local operator as a sum over VEVs of local operators, i.e., condensates. The value of the correlation function then dictates the values of the condensates. Analysis of many separate correlation functions gives consistent results for several condensates, including the gluon condensate, the quark condensate, and many mixed and higher order condensates. In particular one obtains

$$\begin{aligned} \langle (gG)^2 \rangle &\stackrel{\text{def}}{=} \langle g^2 G_{\mu\nu} G^{\mu\nu} \rangle \simeq 0.5 \text{ GeV}^4 \\ \langle \bar{\psi}\psi \rangle &\simeq (-0.23)^3 \text{ GeV}^3 \\ \langle (gG)^4 \rangle &\simeq 5 : 10 \langle (gG)^2 \rangle^2 \end{aligned}$$

Here G refers to the gluon field tensor, ψ to the quark field, and g to the QCD coupling.

These analyses are being refined further through improved sum rule estimates and direct estimates in lattice QCD. They provide the *raw data* which must be explained by models of the QCD vacuum.

Models of the QCD vacuum

A full solution of QCD would automatically give a full description of the vacuum, confinement and the hadron spectrum. Lattice QCD is making rapid progress towards providing the solution as a systematically improvable numerical computation. However, approximate models of the QCD vacuum remain useful in more restricted domains. The purpose of these models is to make quantitative sense of some set of condensates and hadron properties such as masses and form factors.

This section is devoted to models. Opposed to these are systematically improvable computational procedures such as large N QCD and lattice QCD, which are described in their own articles.

The Savvidy vacuum, instabilities and structure

This is a model of the QCD vacuum which at a basic level is a statement that it cannot be the conventional Fock vacuum empty of particles and fields. In 1977, George Savvidy showed that the QCD vacuum with zero field strength is unstable, and decays into a state with a calculable non vanishing value of the field. Since condensates are scalar, it seems like a good first approximation that the vacuum contains some non-zero but homogeneous field which gives rise to these condensates. This would then be a more complicated version of the Higgs mechanism. However, Stanley Mandelstam showed that a homogeneous vacuum field is also unstable. The instability of a homogeneous gluon field was argued by Niels Kjær Nielsen and Poul Olesen in their 1978 paper. These arguments suggest that the scalar condensates are an effective long-distance description of the vacuum, and at short distances, below the QCD scale, the vacuum may have structure.

The dual superconducting model

In a type II superconductor, electric charges condense into Cooper pairs. As a result magnetic flux is squeezed into tubes. In the dual superconductor picture of the QCD vacuum, chromomagnetic monopoles condense into dual Cooper pairs, causing chromoelectric flux to be squeezed into tubes. As a result, confinement and the *string*

picture of hadrons follows. This **dual superconductor** picture is due to Gerard 't Hooft and Stanley Mandelstam. 't Hooft showed further that an Abelian projection of a non-Abelian gauge theory contains magnetic monopoles.

While the vortices in a type II superconductor are neatly arranged into a hexagonal or occasionally square lattice, as is reviewed in Olesen's 1980 seminar one may expect a much more complicated and possibly dynamical structure in QCD. For example, nonabelian Abrikosov-Nielsen-Olesen vortices may vibrate wildly or be knotted.

String models

String models of confinement and hadrons have a long history. They were first invented to explain certain aspects of crossing symmetry in the scattering of two mesons. They were also found to be useful in the description of certain properties of the Regge trajectory of the hadrons. These early developments took on a life of their own called the dual resonance model (later renamed string theory). However, even after the development of QCD string models continued to play a role in the physics of strong interactions. These models are called *non-fundamental strings* or QCD strings, since they should be derived from QCD, as they are, in certain approximations such as the strong coupling limit of lattice QCD.

The model states that the colour electric flux between a quark and an antiquark collapses into a string, rather than spreading out into a Coulomb field as the normal electric flux does. This string also obeys a different force law. It behaves as if the string had constant tension, so that separating out the ends (quarks) would give a potential energy increasing linearly with the separation. When the energy is higher than that of a meson, the string breaks and the two new ends become a quark-antiquark pair, thus describing the creation of a meson. Thus confinement is incorporated naturally into the model.

In the form of the Lund model Monte Carlo program, this picture has had remarkable success in explaining experimental data collected in electron-electron and hadron-hadron collisions.

Bag models

Strictly, these models are not models of the QCD vacuum, but of physical single particle quantum states — the hadrons. The model consists of putting some version of a quark model in a *perturbative vacuum* inside a volume of space called a **bag**. Outside this bag is the real QCD vacuum, whose effect is taken into account through boundary conditions on the quark wave functions. The hadron spectrum is obtained by solving the Dirac equation for quarks with the **bag boundary conditions**.

The **chiral bag model** couples the axial vector current $\bar{\psi}\gamma_5\gamma_\mu\psi$ of the quarks at the bag boundary to a pionic field outside of the bag. In the most common formulation, the chiral bag model basically replaces the interior of the skyrmion with the bag of quarks. Very

curiously, most physical properties of the nucleon become mostly insensitive to the bag radius. Prototypically, the baryon number of the chiral bag remains an integer, independent of bag radius: the exterior baryon number is identified with the topological winding number density of the Skyrme soliton, while the interior baryon number consists of the valence quarks (totaling to one) plus the spectral asymmetry of the quark eigenstates in the bag. The spectral asymmetry is just the vacuum expectation value $\langle \bar{\psi} \gamma_0 \psi \rangle$ summed over all of the quark eigenstates in the bag. Other values, such as the total mass and the axial coupling constant g_A , are not precisely invariant like the baryon number, but are mostly insensitive to the bag radius, as long as the bag radius is kept below the nucleon diameter. Because the quarks are treated as free quarks inside the bag, the radius-independence in a sense validates the idea of asymptotic freedom.

Instanton ensemble

Another view states that BPST-like instantons play an important role in the vacuum structure of QCD. These instantons were discovered in 1975 by Belavin, Polyakov, Schwartz and Tyupkin as topologically stable solutions to the Yang-Mills field equations. They represent tunnelling transitions from one vacuum state to another. These instantons are indeed found in lattice calculations. The first computations performed with instantons used the dilute gas approximation. The results obtained did not solve the infrared problem of QCD, making many physicists turn away from instanton physics. Later, though, an instanton liquid model was proposed, turning out to be more promising an approach.

The **dilute instanton gas model** departs from the supposition that the QCD vacuum consists of a gas of BPST-like instantons. Although only the solutions with one or few instantons (or anti-instantons) are known exactly, a dilute gas of instantons and anti-instantons can be approximated by considering a superposition of one-instanton solutions at great distances from one another. 't Hooft calculated the effective action for such an ensemble, and he found an infrared divergence for big instantons, meaning that an infinite amount of infinitely big instantons would populate the vacuum.

Later, an **instanton liquid model** was studied. This model starts from the assumption that an ensemble of instantons cannot be described by a mere sum of separate instantons. Various models have been proposed, introducing interactions between instantons or using variational methods (like the "valley approximation") endeavoring to approximate the exact multi-instanton solution as closely as possible. Many phenomenological successes have been reached. Whether an instanton liquid can explain confinement in 3+1 dimensional QCD is not known, but many physicists think that it is unlikely.

Center vortex picture

A more recent picture of the QCD vacuum is one in which center vortices play an important role. These vortices are topological defects carrying a center element as charge. These vortices are usually studied using lattice simulations, and it has been found that the behavior of the vortices is closely linked with the confinement-deconfinement phase transition: in the confining phase vortices percolate and fill the space-time volume, in the

deconfining phase they are much suppressed. Also it has been shown that the string tension vanished upon removal of center vortices from the simulations, hinting at an important role for center vortices.

WWT