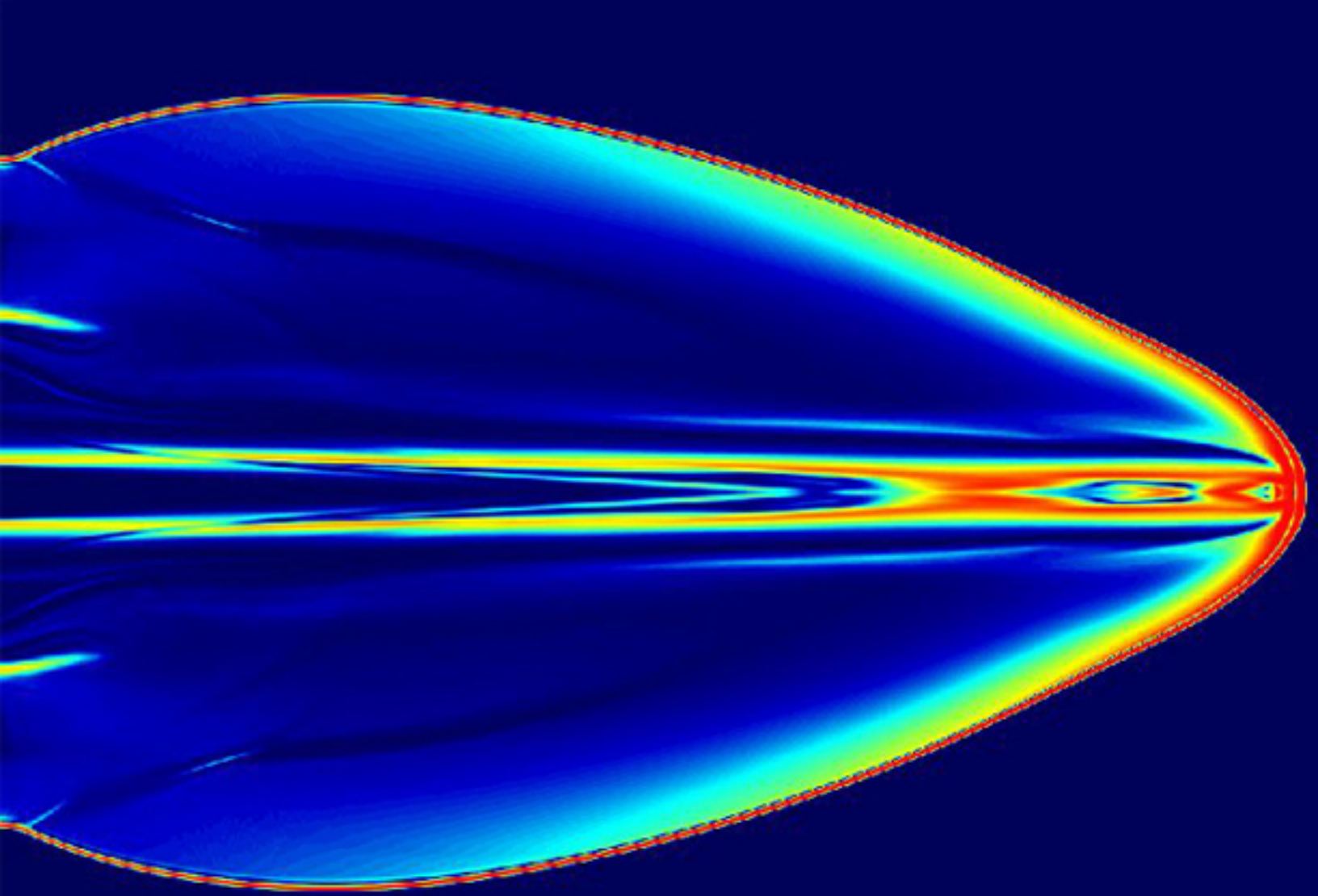


Advanced Engineering Physics

Una Hancock

First Edition, 2012

ISBN 978-81-323-4064-5

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

Table of Contents

Chapter 1

Geophysics

Geophysics is the physics of the Earth and its environment in space. Its subjects include the shape of the Earth, its gravitational and magnetic fields, the dynamics of the Earth as a whole and of its component parts, the Earth's internal structure, composition and tectonics, the generation of magmas, volcanism and rock formation, the hydrological cycle including snow and ice, all aspects of the oceans, the atmosphere, ionosphere, magnetosphere and solar-terrestrial relations, and analogous problems associated with the Moon and other planets.

Geophysics is also applied to societal needs, such as mineral resources, mitigation of natural hazards and environmental protection. Geophysical survey data are used to analyze potential petroleum reservoirs and mineral deposits, to locate groundwater, to locate archaeological finds, to find the thicknesses of glaciers and soils, and for environmental remediation.

History

Replica of Zhang Heng's seismoscope.

Ancient and classical eras

The magnetic compass existed in China back as far as the fourth century BC. It was used as much for feng shui as for navigation on land. It was not until good steel needles could be forged that compasses were used for navigation at sea; before that, they could not retain their magnetism for long. The first mention of a compass in Europe was in 1190.

In circa 240 BC, Erastothenes of Cyrene deduced that the Earth was round and measured the circumference of the Earth, using trigonometry and the angle of the Sun at more than

one latitude in Egypt. He developed a system of latitude and longitude and measured the tilt of the Earth's axis.

Perhaps the earliest contribution to seismology was the invention of a seismoscope by the prolific inventor Zhang Heng in 132 CE. This instrument was designed to drop a bronze ball from the mouth of a dragon into the mouth of a toad. By looking at which of eight toads had the ball, one could determine the direction of the earthquake. It was 1571 years before the first design for a seismoscope was published in Europe, by Jean de la Hautefeuille. It was never built.

Beginnings of modern science

One of the publications that marked the beginning of modern science was William Gilbert's *De Magnete* (1600), a report of a series of meticulous experiments in magnetism. Gilbert deduced that compasses point north because the Earth itself is magnetic.

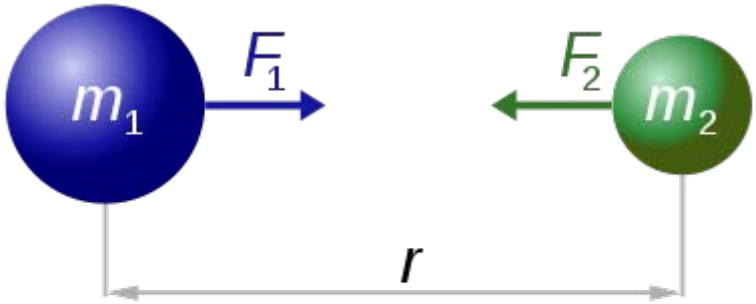
In 1687 Isaac Newton published his *Principia*, which not only laid the foundations for classical mechanics and gravitation but also explained a variety of geophysical phenomena such as the tides and the precession of the equinox.

The first seismometer, an instrument capable of keeping a continuous record of seismic activity, was built by James Forbes in 1844.

Physical phenomena

Geophysics is a highly interdisciplinary subject and geophysicists contribute to every area of the Earth sciences. To provide a clearer idea of what constitutes geophysics, here we, describes phenomena that are studied in physics and how they relate to the Earth and its surroundings.

Gravity

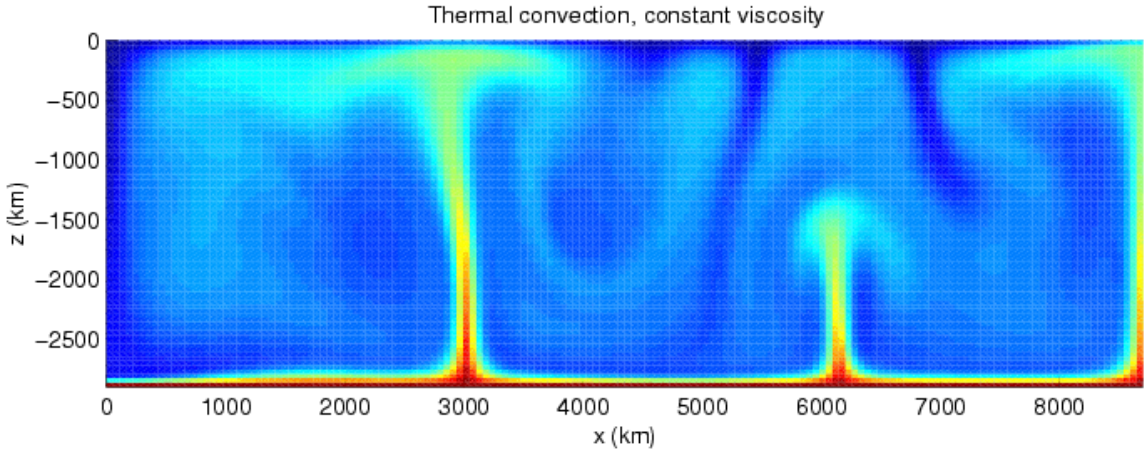


$$F_1 = F_2 = G\frac{m_1 \times m_2}{r^2}$$

The mechanism of Newton's law of universal gravitation.

The gravitational pull of the Moon and Sun give rise to two high tides and two low tides every lunar day, or every 24 hours and 50 minutes. Therefore, there is a gap of 12 hours and 25 minutes between every high tide and between every low tide. Gravitational forces make rocks press down on deeper rocks, increasing their density as the depth increases. Measurements of gravitational acceleration and gravitational potential at the Earth's surface and above it can be used to look for mineral deposits. They also reflect the dynamics of tectonic plates. The geopotential surface called the geoid is one definition of the shape of the Earth. The geoid would be the global mean sea level if the oceans were in equilibrium and could be extended through the continents (such as with very narrow canals).

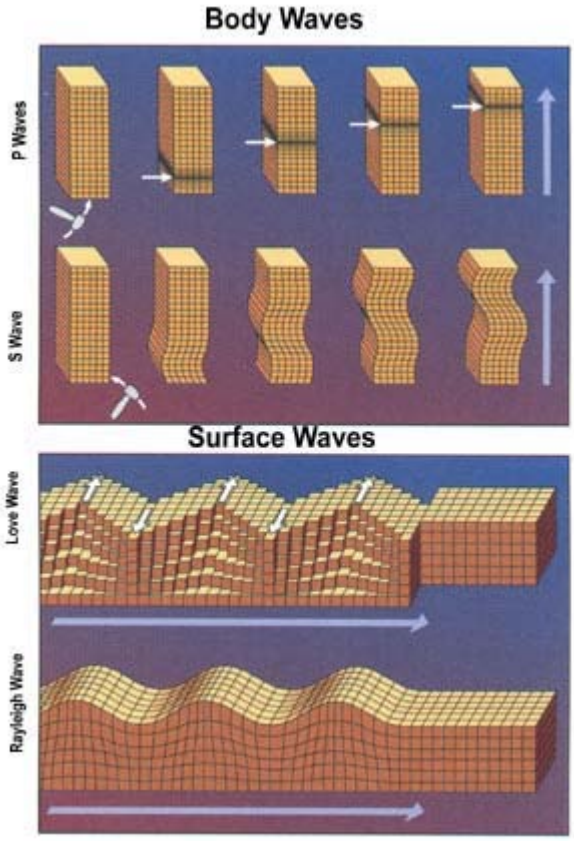
Heat flow



A model of thermal convection in the Earth's mantle.

The Earth is cooling, and the resulting heat flow generates the Earth's magnetic field through the geodynamo and plate tectonics through mantle convection. The main sources of heat are the primordial heat and radioactivity, although there are also contributions from phase transitions. Heat is mostly carried to the surface by thermal convection, although there are two thermal boundary layers - the core-mantle boundary and the lithosphere - in which heat is transported by conduction. Some heat is carried up from the bottom of the mantle by mantle plumes. The heat flow at the Earth's surface is about 4.2×10^{13} W, and it is a potential source of geothermal energy.

Vibrations



Body waves and surface waves.

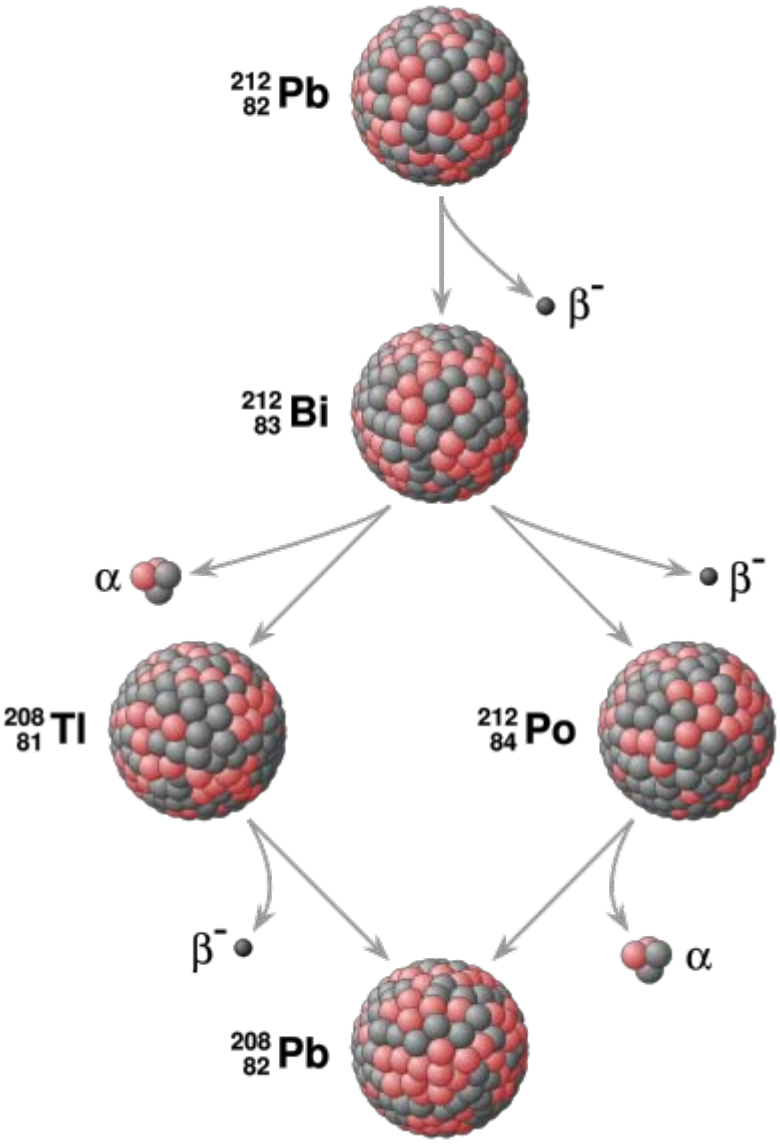
Seismic waves are vibrations that travel through the Earth's interior or along its surface. The entire Earth can also oscillate in forms that are called normal modes. One such mode is the "breathing mode", a uniform expansion and contraction of the Earth.

Ground motions from waves or normal modes are measured using seismographs. If the waves come from a localized source such as an earthquake or explosion, measurements at more than one location can be used to locate the source. The locations of earthquakes provide information on plate tectonics and mantle convection.

Seismic waves can also provide information on the region that the waves travel through. If the density or composition of the rock changes suddenly, some of the waves are reflected. Reflections can provide information on near-surface structure. Changes in the travel direction, called refraction, can be used to infer the deep structure of the Earth.

Earthquakes pose a risk to humans. Understanding their mechanisms, which depend on the type of earthquake (e.g., intraplate or deep focus), can lead to better estimates of earthquake risk and improvements in earthquake engineering.

Radioactivity



Example of a radioactive decay chain

Radioactive decay, in addition to being the main source of heat in the Earth, is an invaluable tool for geochronology. Unstable isotopes decay at predictable rates, and the decay rates of different isotopes cover several orders of magnitude, so radioactive decay can be used to accurately date both recent events and events in past geologic eras.

Electricity

Although we mainly notice electricity during thunderstorms, there is always a downward electric field near the surface that averages 120 V m^{-1}. Relative to the solid Earth, the atmosphere has a net positive charge due to bombardment by cosmic rays. A current of about 1800 A flows in the global circuit. It flows downward from the ionosphere over most of the Earth and back upwards through thunderstorms. The flow is manifested by lightning below the clouds and sprites above.

A variety of electric methods are used in geophysical survey. Some measure spontaneous potential, a potential that arises in the ground because of man-made or natural disturbances. Telluric currents flow in Earth and the Oceans. They have two causes: electromagnetic induction by the time-varying, external-origin geomagnetic field and motion of conducting bodies (such as seawater) across the Earth's permanent magnetic field. The distribution of telluric current density can be used to detect variations in electrical resistivity of underground structures. Geophysicists can also provide the electric current themselves.

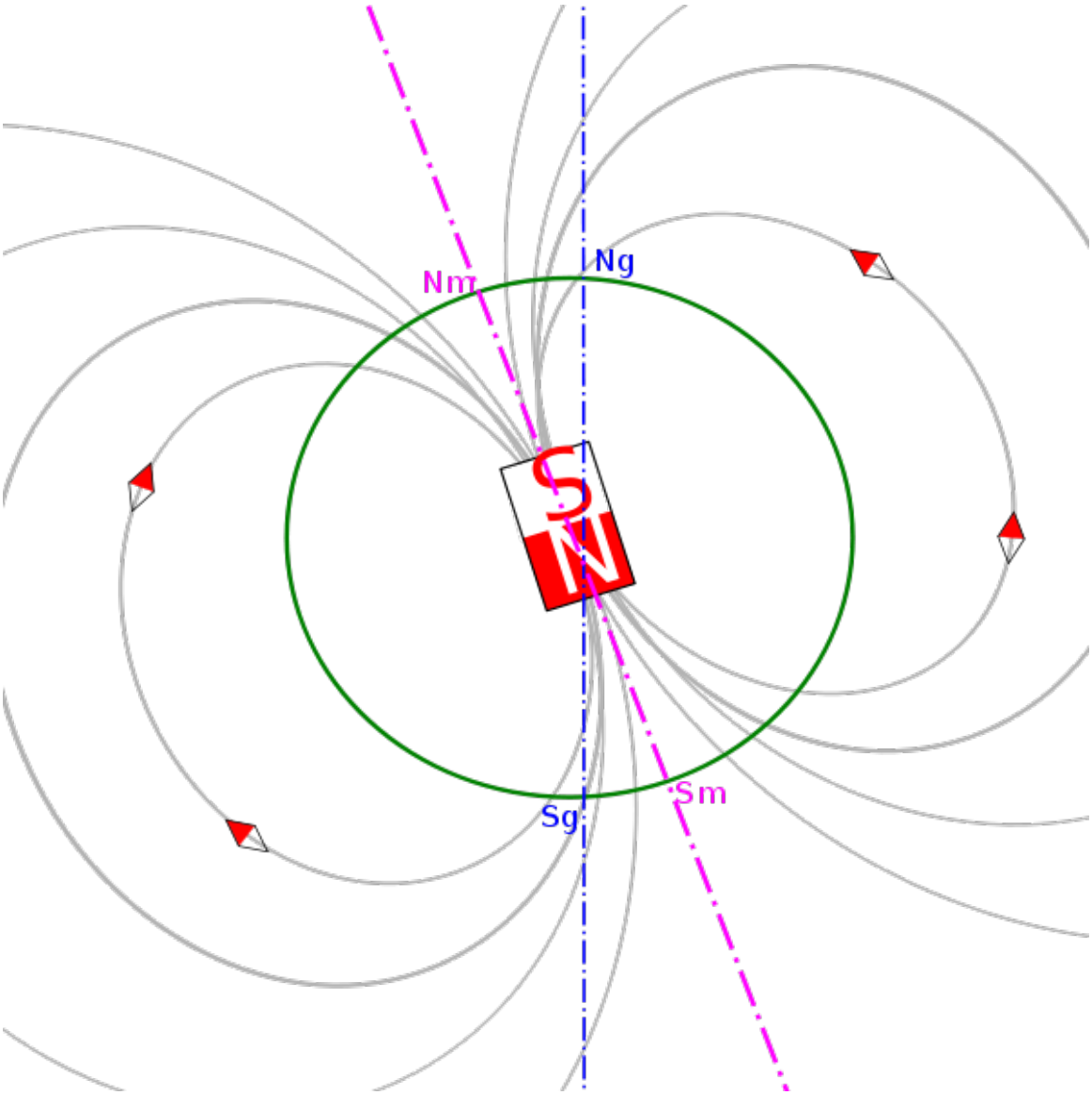
Electromagnetic waves

Electromagnetic waves occur in the ionosphere and magnetosphere as well as the Earth's outer core. dawn chorus is caused by high-energy electrons that get caught in the Van Allen radiation belt. Whistlers are produced by lightning strikes. Hiss may be generated by both. Electromagnetic waves may also be generated by earthquakes.

In the Earth's outer core, electric currents in the highly conductive liquid iron create magnetic fields by electromagnetic induction. Alfvén waves are magnetohydrodynamic waves in the magnetosphere or the Earth's core. In the core, they probably have little observable effect on the geomagnetic field, but slower waves such as magnetic Rossby waves may be one source of geomagnetic secular variation.

Electromagnetic methods that are used for geophysical survey include transient electromagnetics and magnetotellurics.

Magnetism



The variation between magnetic north and "true" north.

The Earth's magnetic field protects the Earth from the deadly Solar wind and has long been used for navigation. It originates in the fluid motions of the Earth's outer core. The magnetic field in the upper atmosphere gives rise to the auroras.

The Earth's field is roughly like a tilted dipole, but it changes over time (a phenomenon called geomagnetic secular variation). Mostly the geomagnetic pole stays near the geographic pole, but at random intervals averaging a million years or so, the polarity of the Earth's field reverses. These geomagnetic reversals are recorded in rocks and their signature can be seen in striped magnetic anomalies on the seafloor. These stripes provide quantitative information on seafloor spreading, a part of plate tectonics. In addition, the magnetization in rocks can be used to measure the motion of continents.

Fluid dynamics

Fluid motions occur in the magnetosphere, atmosphere, ocean, mantle and core. Even the mantle, though it has an enormous viscosity, flows like a fluid over long time intervals. This flow is reflected in phenomena such as isostasy and post-glacial rebound. The mantle flow drives plate tectonics and the flow in the Earth's core drives the geodynamo.

Geophysical fluid dynamics is a primary tool in physical oceanography and meteorology. The rotation of the Earth has profound effects on the Earth's fluid dynamics, often due to the Coriolis effect. In the atmosphere it gives rise to large-scale patterns like Rossby waves and determines the basic circulation patterns of storms. In the ocean they drive large-scale circulation patterns as well as Kelvin waves and Ekman spirals at the ocean surface. In the Earth's core, the circulation of the molten iron is structured by Taylor columns.

Waves and other phenomena in the magnetosphere can be modeled using magnetohydrodynamics.

Condensed matter physics

The physical properties of minerals must be understood to infer the composition of the Earths' interior from seismology, the geothermal gradient and other sources of information. Mineral physicists study the elastic properties of minerals as well as their high-pressure phase diagrams, melting points and equations of state at high pressure. Studies of creep determine how rocks that are brittle at the surface can flow deep down. These properties determine the rheology that determines the geodynamics.

Water is a very complex substance and its unique properties are essential for life. Its physical properties shape the hydrosphere and are an essential part of the water cycle and climate. Its thermodynamic properties determine evaporation and the thermal gradient in the atmosphere. The many types of precipitation involve a complex mixture of processes such as coalescence, supercooling and supersaturation. Some of the precipitated water becomes groundwater, and groundwater flow includes phenomena such as percolation, while the conductivity of water makes electrical and electromagnetic methods useful for tracking groundwater flow. Physical properties of water such as salinity have a large effect on its motion in the oceans.

The many phases of ice form the cryosphere and come in forms like ice sheets, glaciers, sea ice, freshwater ice, snow, and frozen ground (or permafrost).

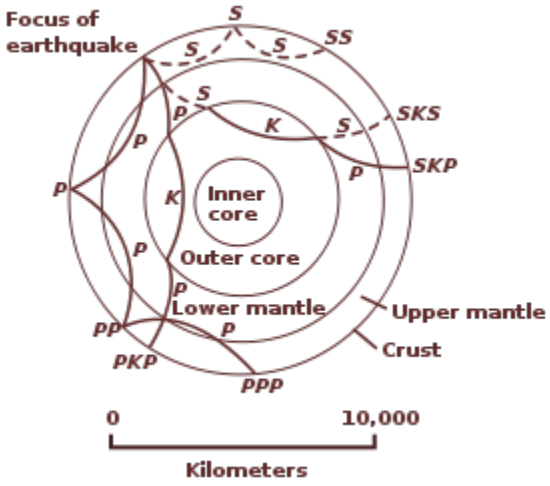
Regions of the Earth

Size and form of the Earth

The Earth is roughly spherical, but it bulges towards the Equator, so it is roughly in the shape of an ellipsoid. This bulge is due to its rotation and is nearly consistent with an

Earth in hydrostatic equilibrium. The detailed shape of the Earth, however, is also affected by the distribution of continents and ocean basins, and to some extent by the dynamics of the plates.

Structure of the Earth



Mapping the interior of the Earth with earthquake waves.

Evidence from seismology, heat flow at the surface, and mineral physics is combined with the Earth's mass and moment of inertia to infer models of the Earth's interior - its composition, density, temperature, pressure. The Earth's mass is $M = 5.975 \times 10^{24}$ kg and its mean radius is $R = 6371$ km , so its mean specific gravity is $< \rho > = 5.515$. This is substantially higher than the typical specific gravity (2.7–3.3) of rocks at the surface. Its moment of inertia is $0.33\ M\ R^2$, whereas it would be $0.4\ M\ R^2$ if the earth was a sphere of constant density. Both lines of evidence point to a concentration of mass near the center. However, the density of the rock will increase with depth because of the increasing pressure. To determine how large this effect is, the Adams–Williamson equation is used to determine how density increases with pressure. The conclusion is that pressure alone cannot account for the increase in density. Instead, we know that the Earth's core is composed of an alloy of iron and other minerals.

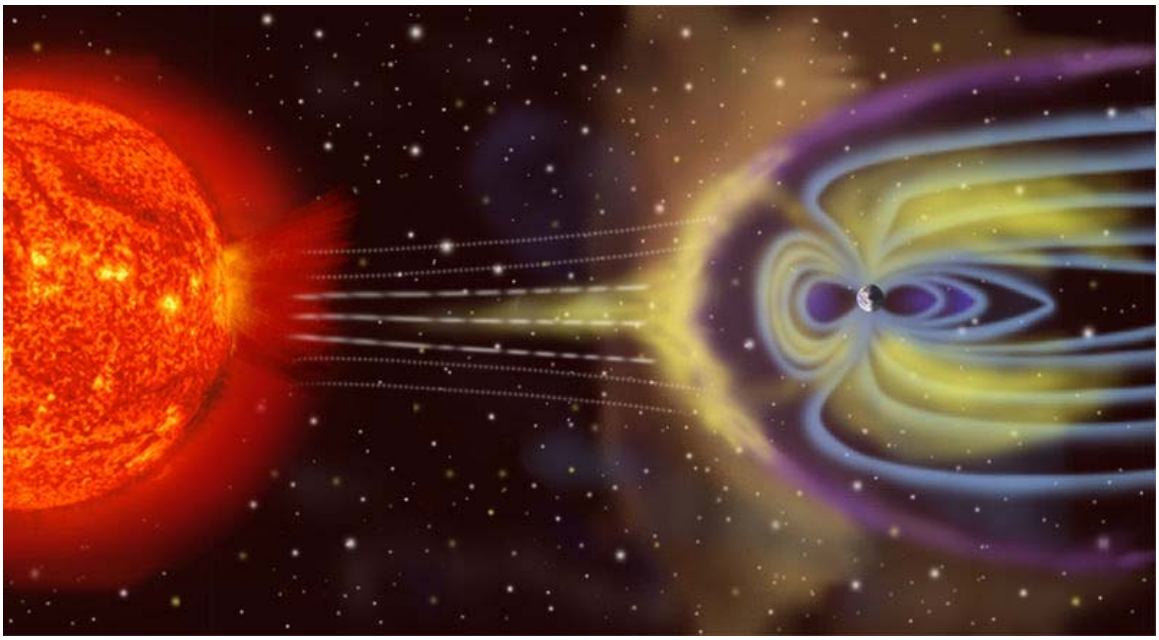
Reconstructions of seismic waves in the deep interior of the Earth show that there are no S-waves in the outer core. This indicates that the outer core is liquid, because liquids cannot support shear. The outer core is liquid, and the motion of this highly conductive fluid generates the Earth's field. The inner core, however, is solid because of the enormous pressure.

Reconstruction of seismic reflections in the deep interior indicate some major discontinuities in seismic velocities that demarcate the major zones of the Earth: inner core, outer core, mantle, lithosphere and crust. The mantle itself is divided into the upper mantle, transition zone, lower mantle and D'' layer. Between the crust and the mantle is the Mohorovičić discontinuity.

The seismic model of the Earth does not by itself determine the composition of the layers. For a complete model of the Earth, mineral physics is needed to interpret seismic velocities in terms of composition. The mineral properties are temperature-dependent, so the geotherm must also be determined. This requires physical theory for thermal conduction and convection and the heat contribution of radioactive elements. The main model for the radial structure of the interior of the Earth is the Preliminary Reference Earth Model (PREM). Some parts of this model have been updated by recent findings in mineral physics and supplemented by seismic tomography. The mantle is mainly composed of silicates, and the boundaries between layers of the mantle are probably due to phase transitions.

The mantle acts as a solid for seismic waves, but under high pressures and temperatures it deforms so that over millions of years it acts like a liquid. This makes plate tectonics possible. Geodynamics is the study of the fluid flow in the mantle and core.

The magnetosphere



The solar wind is deflected by the magnetosphere (not to scale)

If a planet's magnetic field is strong enough, its interaction with the solar wind forms a magnetosphere around a planet. Early space probes mapped out the gross dimensions of the terrestrial magnetic field, which extends about 10 Earth radii towards the Sun. The solar wind, a stream of charged particles, streams out and around the terrestrial magnetic field, and continues behind the magnetic tail, hundreds of Earth radii downstream. Inside the magnetosphere, there are relatively dense regions of solar wind particles, the Van Allen radiation belts.

Other fields and related disciplines

Fields

- Geodesy, measurement of the Earth: GPS, vertical and horizontal motions of the Earth's surface, navigation, the study of the Earth's gravitational field, and the size and form of the Earth
- The study of large-scale motions of the Earth's surface and interior, including:
    - Tectonophysics, the study of the physical processes that cause and result from plate tectonics
    - Geodynamics, the study of modes of transport deformation within the Earth: rock deformation, mantle flow and convection, heat flow, lithosphere dynamics
- Geomagnetism, the study of the Earth's magnetic field, including its origin, telluric currents driven by the magnetic field, the Van Allen belts, and the interaction between the magnetosphere and the solar wind. This field is associated with paleomagnetism, or the measurement of the orientation of the Earth's magnetic field over the geologic past.
- Seismology, the study of the structure and composition of the Earth through seismic waves, and of surface deformations during earthquakes and seismic hazards
- Mathematical geophysics, The development and applications of mathematical methods and techniques for the solution of geophysical problems.
- Geophysical surveying:
    - Exploration and engineering geophysics, using surface methods to detect or infer the presence and position of concentrations of ore minerals and hydrocarbons
    - Archaeological geophysics, for archaeological imaging or mapping
    - Environmental and Engineering Geophysics, for locating underground storage tanks (USTs) or utilities, Unexploded ordnance (UXO), delineating landfills, locating voids or potential subsidence, finding depth to, P-wave or S-wave velocity of, or rippability of bedrock, or the pathway of groundwater movement
    - Shallow seismology is used in exploration geophysics (to find oil and gas) and for environmental characterization of the subsurface

Related disciplines

- Volcanology, the study of volcanoes, volcanic features (hot springs, geysers, fumaroles), volcanic rock, and heat flow related to volcanoes
- Atmospheric sciences, which includes:
    - Atmospheric electricity and the ionosphere

    - Aeronomy, the study of the physical structure and chemistry of the atmosphere.
    - Meteorology and Climatology, which both involve studies of the weather.

- The study of water on the Earth, hydrology, physical oceanography and glaciology
- Geological and geophysical engineering and Engineering geology, applying geophysics to the engineering design of facilities including roads, tunnels, and mines
- The study of the rocks and minerals, including petrophysics and aspects of mineralogy such as physical mineralogy and crystal structure

Methods of geophysics

Space probes

Space probes made it possible to collect data from not only the visible light region, but in other areas of the electromagnetic spectrum. The planets can be characterized by their force fields: gravity and their magnetic fields, which are studied through geophysics and space physics.

Measuring the changes in acceleration experienced by spacecraft as they orbit has allowed fine details of the gravity fields of the planets to be mapped. For example, in the 1970s, the gravity field disturbances above lunar maria were measured through lunar orbiters, which lead to the discovery of concentrations of mass, mascons, beneath the Imbrium, Serenitatis, Crisium, Nectaris and Humorum basins.

In 2002, NASA launched the Gravity Recovery and Climate Experiment, wherein two twin satellites map variations in Earth's gravity field by making measurements of the distance between the two satellites using GPS and a microwave ranging system. Gravity variations detected by GRACE include those caused by changes in ocean currents; runoff and ground water depletion; melting ice sheets and glaciers.

Chapter 2

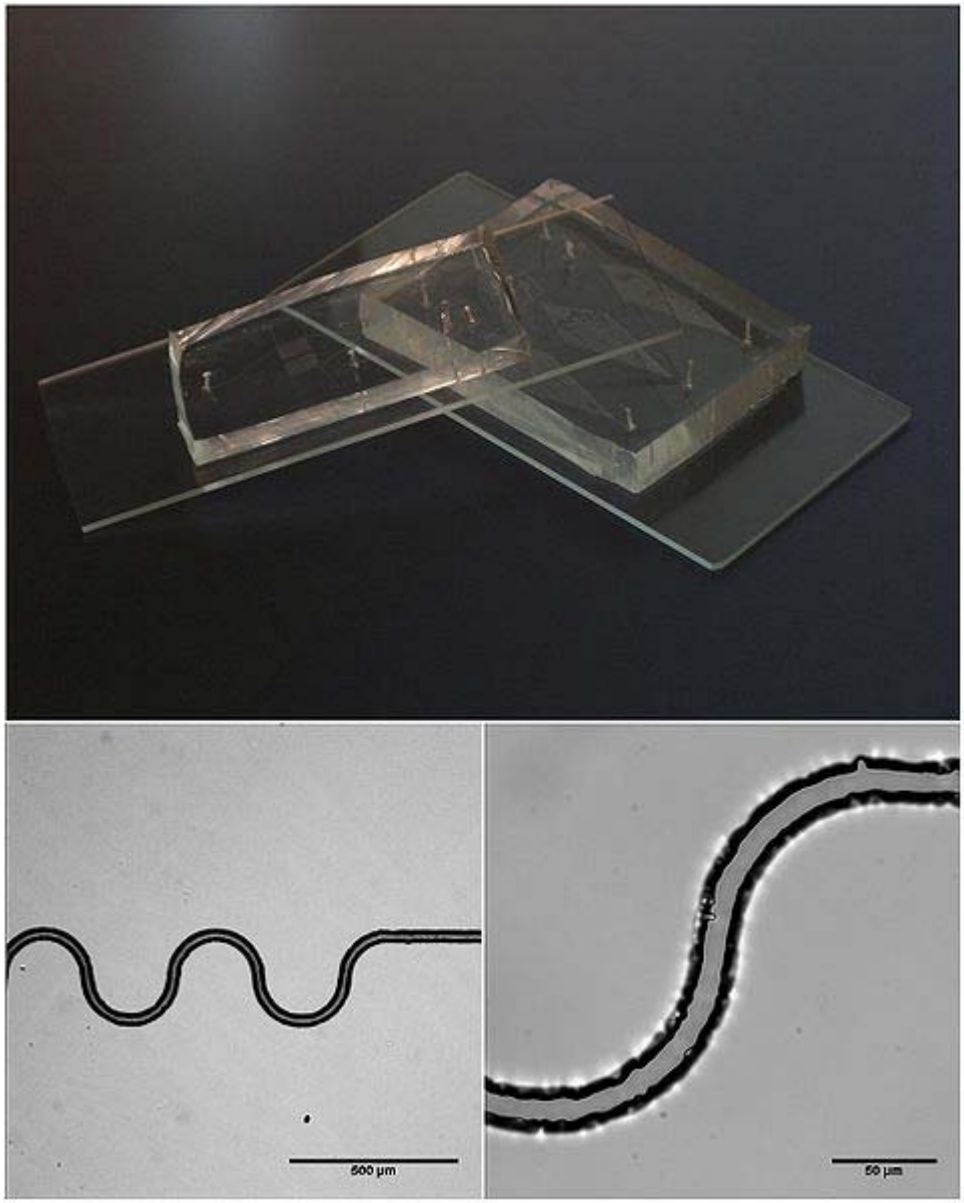
Microfluidics

Microfluidics deals with the behavior, precise control and manipulation of fluids that are geometrically constrained to a small, typically sub-millimeter, scale. Typically, **micro** means one of the following features:

- small volumes (nl, pl, fl)
- small size
- low energy consumption
- effects of the micro domain

It is a multidisciplinary field intersecting engineering, physics, chemistry, microtechnology and biotechnology, with practical applications to the design of systems in which such small volumes of fluids will be used. Microfluidics emerged in the beginning of the 1980s and is used in the development of inkjet printheads, DNA chips, lab-on-a-chip technology, micro-propulsion, and micro-thermal technologies.

Microscale behavior of fluids



Silicone rubber and glass microfluidic devices. Top: a photograph of the devices. Bottom: DIC micrographs of a serpentine channel ~15 µm wide.

The behavior of fluids at the microscale can differ from 'macrofluidic' behavior in that factors such as surface tension, energy dissipation, and fluidic resistance start to dominate the system. Microfluidics studies how these behaviors change, and how they can be worked around, or exploited for new uses.

At small scales (channel diameters of around 100 nanometers to several hundred micrometers) some interesting and sometimes unintuitive properties appear. In particular, the Reynolds number (which compares the effect of momentum of a fluid to the effect of

viscosity) can become very low. A key consequence of this is that fluids, when side-by-side, do not necessarily mix in the traditional sense; molecular transport between them must often be through diffusion.

High specificity of chemical and physical properties (concentration, pH, temperature, shear force, etc.) can also be ensured resulting in more uniform reaction conditions and higher grade products in single and multi-step reactions.

Effects of micro domain

- laminar flow
- surface tension
- electrowetting
- fast thermal relaxation
- electrical surface charges
- diffusion

Key application areas

Microfluidic structures include micropneumatic systems, i.e. microsystems for the handling of off-chip fluids (liquid pumps, gas valves, etc.), and microfluidic structures for the on-chip handling of nano- and picolitre volumes. To date, the most successful commercial application of microfluidics is the inkjet printhead. Significant research has been applied to the application of microfluidics for the production of industrially relevant quantities of material.

Advances in microfluidics technology are revolutionizing molecular biology procedures for enzymatic analysis (e.g., glucose and lactate assays), DNA analysis (e.g., polymerase chain reaction and high-throughput sequencing), and proteomics. The basic idea of microfluidic biochips is to integrate assay operations such as detection, as well as sample pre-treatment and sample preparation on one chip.

An emerging application area for biochips is clinical pathology, especially the immediate point-of-care diagnosis of diseases. In addition, microfluidics-based devices, capable of continuous sampling and real-time testing of air/water samples for biochemical toxins and other dangerous pathogens, can serve as an always-on "bio-smoke alarm" for early warning.

Continuous-flow microfluidics

These technologies are based on the manipulation of continuous liquid flow through microfabricated channels. Actuation of liquid flow is implemented either by external pressure sources, external mechanical pumps, integrated mechanical micropumps, or by combinations of capillary forces and electrokinetic mechanisms. Continuous-flow microfluidic operation is the mainstream approach because it is easy to implement and less sensitive to protein fouling problems. Continuous-flow devices are adequate for

many well-defined and simple biochemical applications, and for certain tasks such as chemical separation, but they are less suitable for tasks requiring a high degree of flexibility or ineffect fluid manipulations. These closed-channel systems are inherently difficult to integrate and scale because the parameters that govern flow field vary along the flow path making the fluid flow at any one location dependent on the properties of the entire system. Permanently-etched microstructures also lead to limited reconfigurability and poor fault tolerance capability.

Process monitoring capabilities in continuous-flow systems can be achieved with highly sensitive microfluidic flow sensors based on MEMS technology which offer resolutions down to the nanoliter range.

Digital (droplet-based) microfluidics

Alternatives to the above closed-channel continuous-flow systems include novel open structures, where discrete, independently controllable droplets are manipulated on a substrate using electrowetting. Following the analogy of digital microelectronics, this approach is referred to as digital microfluidics. Le Pesant et al. pioneered the use of electrocapillary forces to move droplets on a digital track. The "fluid transistor" pioneered by Cytonix also played a role. The technology was subsequently commercialized by Duke University. By using discrete unit-volume droplets, a microfluidic function can be reduced to a set of repeated basic operations, i.e., moving one unit of fluid over one unit of distance. This "digitization" method facilitates the use of a hierarchical and cell-based approach for microfluidic biochip design. Therefore, digital microfluidics offers a flexible and scalable system architecture as well as high fault-tolerance capability. Moreover, because each droplet can be controlled independently, these systems also have dynamic reconfigurability, whereby groups of unit cells in a microfluidic array can be reconfigured to change their functionality during the concurrent execution of a set of bioassays. Although droplets are manipulated in confined microfluidic channels, since the control on droplets is not independent, it should not be confused as "digital microfluidics". One common actuation method for digital microfluidics is electrowetting-on-dielectric (EWOD). Many lab-on-a-chip applications have been demonstrated within the digital microfluidics paradigm using electrowetting. However, recently other techniques for droplet manipulation have also been demonstrated using Surface Acoustic Waves, optoelectrowetting, mechanical actuation, etc.

DNA chips (microarrays)

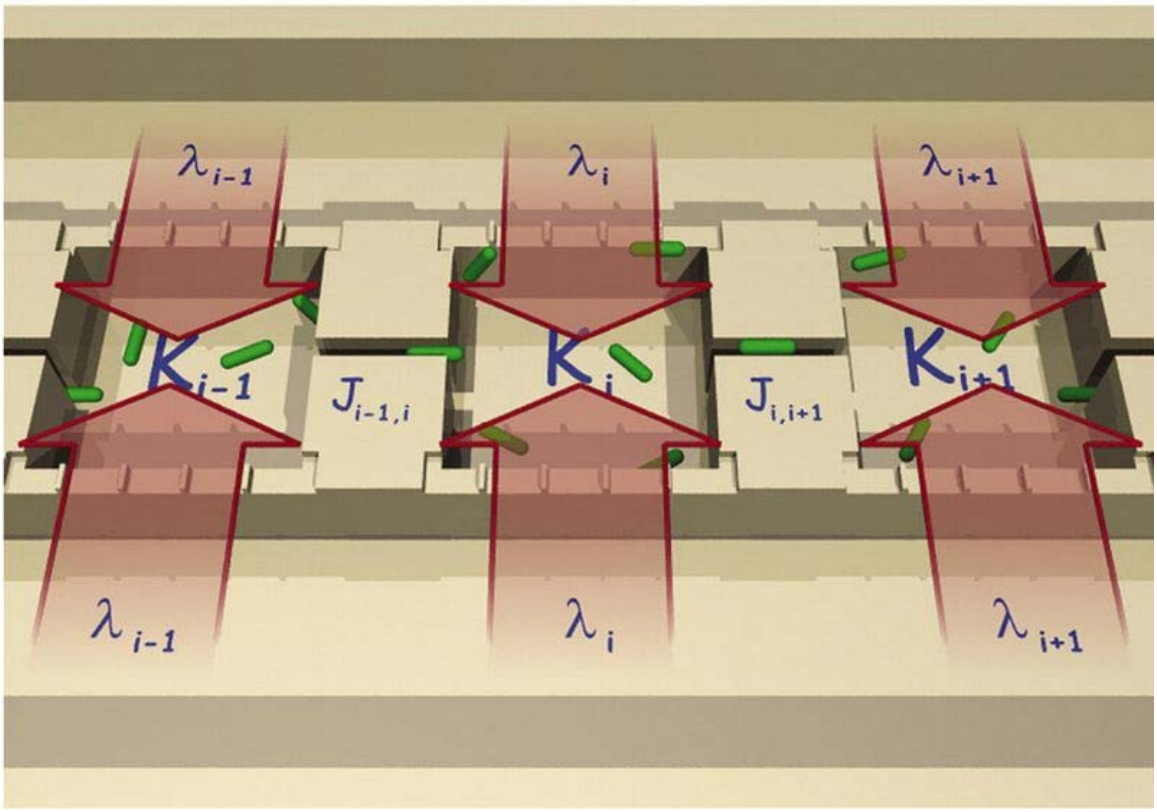
Early biochips were based on the idea of a DNA microarray, e.g., the GeneChip DNAarray from Affymetrix, which is a piece of glass, plastic or silicon substrate on which pieces of DNA (probes) are affixed in a microscopic array. Similar to a DNA microarray, a protein array is a miniature array where a multitude of different capture agents, most frequently monoclonal antibodies, are deposited on a chip surface; they are used to determine the presence and/or amount of proteins in biological samples, e.g., blood. A drawback of DNA and protein arrays is that they are neither reconfigurable nor

scalable after manufacture. Digital microfluidics has been described as a means for carrying out Digital PCR.

Molecular biology

In addition to microarrays biochips have been designed for two-dimensional electrophoresis, transcriptome analysis, and PCR amplification. Other applications include various electrophoresis and liquid chromatography applications for proteins and DNA, cell separation, in particular blood cell separation, protein analysis, cell manipulation and analysis including cell viability analysis and microorganism capturing.

Evolutionary biology



Three Micro Habitat Patches MHPs connected by dispersal corridors (indicated here as $J_{i,j}$) into a 1D lattice. The ecosystem service (of habitat renewal) to each MHP represented here as λ_i (red arrows). Each MHP can also hold different carrying capacity K_i for its supporting local population of bacterial cells (depicted in green).

By combining microfluidics with landscape ecology and nanofluidics, a nano/micro fabricated fluidic landscape can be constructed by building local patches of bacterial habitat and connecting them by dispersal corridors. The resulting landscapes can be used as physical implementations of an adaptive landscape, by generating a spatial mosaic of patches of opportunity distributed in space and time. The patchy nature of these fluidic landscapes allows for the study of adapting bacterial cells in a metapopulation system.

The evolutionary ecology of these bacterial systems in these synthetic ecosystems allows for using biophysics to address questions in evolutionary biology.

Cellular biophysics

By rectifying the motion of individual swimming bacteria, microfluidic structures can be use to extract mechanical motion from a population of motile bacterial cells. This way, bacteria-powered rotors can be built.

Optics

The merger of microfluidics and optics is typical known as Optofluidics. An example of an optofluidic device is a Tuneable Microlens Array

Acoustic droplet ejection (ADE)

Acoustic droplet ejection uses a pulse of ultrasound to move low volumes of fluids (typically nanoliters or picoliters) without any physical contact. This technology focuses acoustic energy into a fluid sample in order to eject droplets as small as a millionth of a millionth of a liter (picoliter = 10^{-12} liter). ADE technology is a very gentle process, and it can be used to transfer proteins, high molecular weight DNA and live cells without damage or loss of viability. This feature makes the technology suitable for a wide variety of applications including proteomics and cell-based assays.

Fuel cells

Microfluidic fuel cells can use laminar flow to separate the fuel and its oxidant to control the interaction of the two fluids without a physical barrier as would be required in conventional fuel cells.

Ceramic Pot Water Filters

In recent times, developing nations are adopting clay based ceramic water filters for cost effective water filtration. These water filters made from molds manufactured by mixing clay and waste plant materials such as rice husk, sawdust,dried plant biomass etc. in some volumetric or weight proportions. These vessels are in use from very early ages in Africa and Asia. The initial water percolation through there clay pots are alkaline. Recently it was found that this alkaline nature can be easily predicted by use of simple models incorporating micro/nano fluid transport processes such as Capillary Osmosis, Thermo Osmosis, Electro-osmosis and Discharge from these clay ceramic devices.

A tool for cell biological research

Microfluidic technology is creating powerful tools for cell biologists to control the complete cellular environement, leading to new questions and new discoveries . We can list here diverse advantages for microbiology of these tools:

- Microenvironemental control
- Precise spatiotemporal concentration gradients
- Mechanical deformation
- Force measurements of adherent cells
- Confining cells
- Exerting a controlled force
- Fast and precise temperature control
- Electric field integration
- Cell culture

Chapter 3

Accelerator Physics and Quantum Optics

Accelerator physics

Accelerator physics deals with the problems of building and operating particle accelerators.

The experiments conducted with particle accelerators are not regarded as part of **accelerator physics**. These belong (according to the objectives of the experiments) to particle physics, nuclear physics, condensed matter physics, materials physics, etc. as well as to other sciences and technical fields. The types of experiments done at a particular accelerator and/or its other uses are largely constrained by the characteristics of the accelerator itself, such as energy (per particle), types of particles, beam intensity, beam quality, etc.

Accelerator physics itself is the study of the motion of the particle beam through the machine, control and manipulation of the beam, interaction with the machine itself, and measurements of the various parameters associated with particle beams.

Equations of motion

The motion of charged particles through an accelerator is controlled using applied electro-magnetic fields, and the equations of motion may be derived from (or, since in many cases a general solution is not possible, approximated from) relativistic Hamiltonian mechanics. Typically, a separate Hamiltonian is written down for each element (e.g. for a single quadrupole magnet, or accelerating structure) to allow the equations of motion to be solved for this one element. Once this has been done for each element encountered in the machine, the full trajectory of each particle may be calculated for the entire machine.

In many cases a general solution of the full Hamiltonian is not possible, so it is necessary to make approximations. This may take the form of the Paraxial approximation (a Taylor series in the dynamical variables, truncated to low order), however, even in the cases of

strongly non-linear magnetic fields, a Lie transform may be used to construct an integrator with a high degree of accuracy, and the paraxial approximation is not necessary.

Diagnostics

A vital component of any accelerator are the diagnostic devices that allow various properties of the particle bunches to be measured.

A typical machine may use many different types of measurement device in order to measure different properties. These include (but are not limited to) Beam Position Monitors (BPMs) to measure the position of the bunch, screens (fluorescent screens, Optical Transition Radiation (OTR) devices) to image the profile of the bunch, wire-scanners to measure its cross-section, and toroids or ICTs to measure the bunch charge (i.e. the number of particles per bunch).

While many of these devices rely on well understood technology, designing a device capable of measuring a beam for a particular machine is a complex task requiring much expertise. Not only is a full understanding of the physics of the operation of the device necessary, but it is also necessary to ensure that the device is capable of measuring the expected parameters of the machine under consideration.

Success of the full range of beam diagnostics often underpins the success of the machine as a whole.

Machine tolerances

Errors in the alignment of components, field strength, etc., are inevitable in machines of this scale, so it is important to consider the tolerances under which a machine may operate.

Engineers will provide the physicists with expected tolerances for the alignment and manufacture of each component to allow full physics simulations of the expected behaviour of the machine under these conditions. In many cases it will be found that the performance is degraded to an unacceptable level, requiring either re-engineering of the components, or the invention of algorithms that allow the machine performance to be 'tuned' back to the design level.

This may require many simulations of different error conditions in order to determine the relative success of each tuning algorithm, and to allow recommendations for the collection of algorithms to be deployed on the real machine.

Interactions between the beam and the machine

Due to the strong electro-magnetic fields that follow the beam, it is possible for it to interact with any electrical impedance in the walls of the beam pipe. This may be in the

form of a resistive impedance (i.e. the finite resistivity of the beam pipe material) or an inductive/capacitive impedance (due to the geometric changes in the beam pipe's cross section).

These impedances will induce so called 'wake-fields' (a strong warping of the electromagnetic field of the beam) that can interact with later particles. Since this interaction may have a negative effect, it must be studied to determine its magnitude, and to determine any actions that may be taken to mitigate it.

Quantum optics

Quantum optics is a field of research in physics, dealing with the application of quantum mechanics to phenomena involving light and its interactions with matter.

History of quantum optics

Light is made up of particles called photons and hence inherently is "grainy" (quantized). Quantum optics is the study of the nature and effects of light as quantized photons. The first indication that light might be quantized came from Max Planck in 1899 when he correctly modeled blackbody radiation. By assuming blackbody radiation is quantized, Bohr showed that the atoms were also quantized, in the sense that they could only emit discrete amounts of energy. The understanding of the interaction between light and matter following these developments not only formed the basis of quantum optics but were also crucial for the development of quantum mechanics as a whole. However, the subfields of quantum mechanics dealing with matter-light interaction were principally regarded as research into matter rather than into light; hence one rather spoke of atom physics and quantum electronics in 1960. Laser science—i.e., research into principles, design and application of these devices—became an important field, and the quantum mechanics underlying the laser's principles was studied now with more emphasis on the properties of light, and the name *quantum optics* became customary.

As laser science needed good theoretical foundations, and also because research into these soon proved very fruitful, interest in quantum optics rose. Following the work of Dirac in quantum field theory, George Sudarshan, Roy J. Glauber, and Leonard Mandel applied quantum theory to the electromagnetic field in the 1950s and 1960s to gain a more detailed understanding of photodetection and the statistics of light. This led to the introduction of the coherent state as a quantum description of laser light and the realization that some states of light could not be described with classical waves. In 1977, Kimble et al. demonstrated the first source of light which required a quantum description: a single atom that emitted one photon at a time. This was the first conclusive evidence that light was made up of photons. Another quantum state of light with certain advantages over any classical state, squeezed light, was soon proposed. At the same time, development of short and ultrashort laser pulses—created by Q switching and modelocking techniques—opened the way to the study of unimaginably fast ("ultrafast") processes. Applications for solid state research (e.g. Raman spectroscopy) were found,

and mechanical forces of light on matter were studied. The latter led to levitating and positioning clouds of atoms or even small biological samples in an optical trap or optical tweezers by laser beam. This, along with Doppler cooling was the crucial technology needed to achieve the celebrated Bose-Einstein condensation.

Other remarkable results are the demonstration of quantum entanglement, quantum teleportation, and (recently, in 1995) quantum logic gates. The latter are of much interest in quantum information theory, a subject which partly emerged from quantum optics, partly from theoretical computer science.

Today's fields of interest among quantum optics researchers include parametric down-conversion, parametric oscillation, even shorter (attosecond) light pulses, use of quantum optics for quantum information, manipulation of single atoms, Bose-Einstein condensates, their application, and how to manipulate them (a sub-field often called atom optics), coherent perfect absorbers, and much more.

Research into quantum optics that aims to bring photons into use for information transfer and computation is now often called photonics to emphasize the claim that photons and photonics will take the role that electrons and electronics now have.

Concepts of quantum optics

According to quantum theory, light may be considered not only as an electro-magnetic wave but also as a "stream" of particles called photons which travel with *c*, the vacuum speed of light. These particles should not be considered to be classical billiard balls, but as quantum mechanical particles described by a wavefunction spread over a finite region.

Each particle carries one quantum of energy equal to *hf*, where h is Planck's constant and f is the frequency of the light. The postulation of the quantization of light by Max Planck in 1899 and the discovery of the general validity of this idea in Albert Einstein's 1905 explanation of the photoelectric effect soon led physicists to realize the possibility of population inversion and the possibility of the laser.

This kind of use of statistical mechanics is the fundament of most concepts of quantum optics: Light is described in terms of field operators for creation and annihilation of photons—i.e. in the language of quantum electrodynamics.

A frequently encountered state of the light field is the coherent state as introduced by George Sudarshan in 1963. This state, which can be used to approximately describe the output of a single-frequency laser well above the laser threshold, exhibits Poissonian photon number statistics. Via certain nonlinear interactions, a coherent state can be transformed into a squeezed coherent state, which can exhibit super- or sub- Poissonean photon statistics. Such light is called squeezed light. Other important quantum aspects are related to correlations of photon statistics between different beams. For example, parametric nonlinear processes can generate so-called twin beams, where ideally each photon of one beam is associated with a photon in the other beam.

Atoms are considered as quantum mechanical oscillators with a discrete energy spectrum with the transitions between the energy eigenstates being driven by the absorption or emission of light according to Einstein's theory with the oscillator strength depending on the quantum numbers of the states.

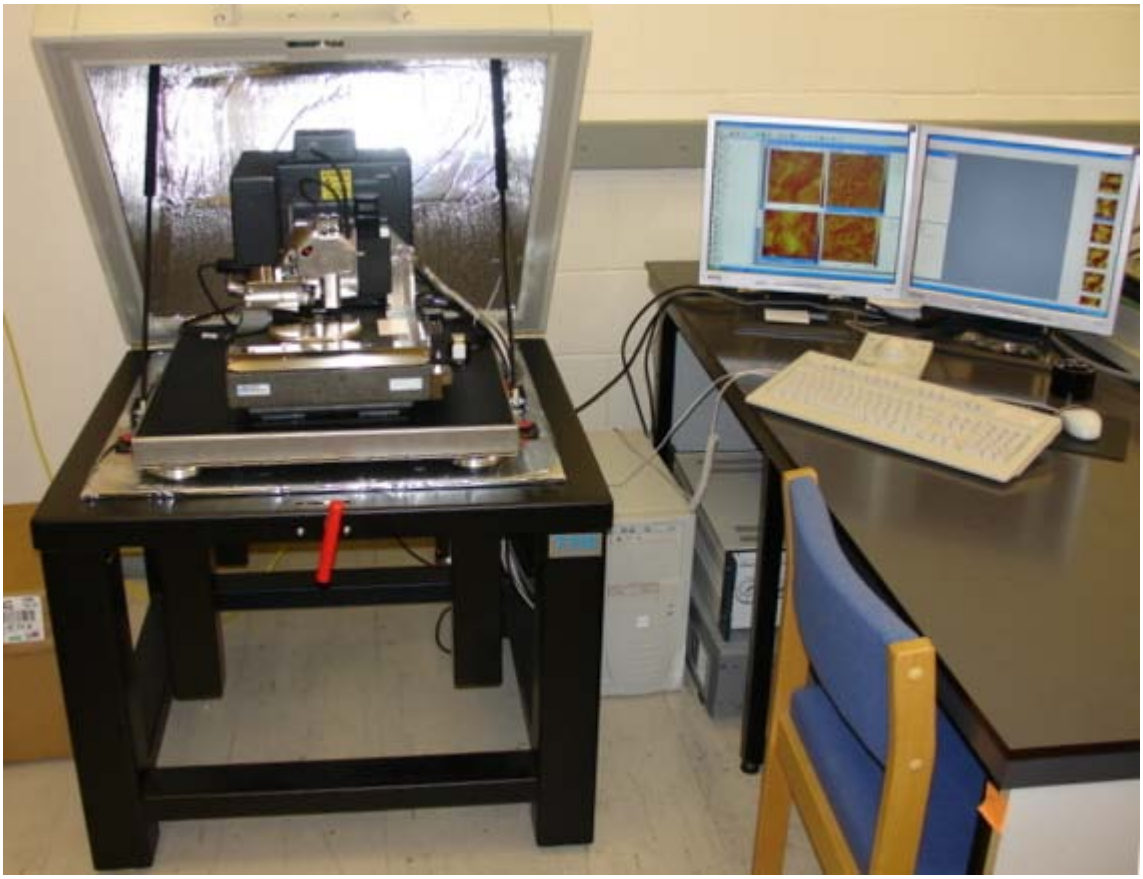
For solid state matter one uses the energy band models of solid state physics. This is important as understanding how light is detected (typically by a solid-state device that absorbs it) is crucial for understanding experiments.

Quantum electronics

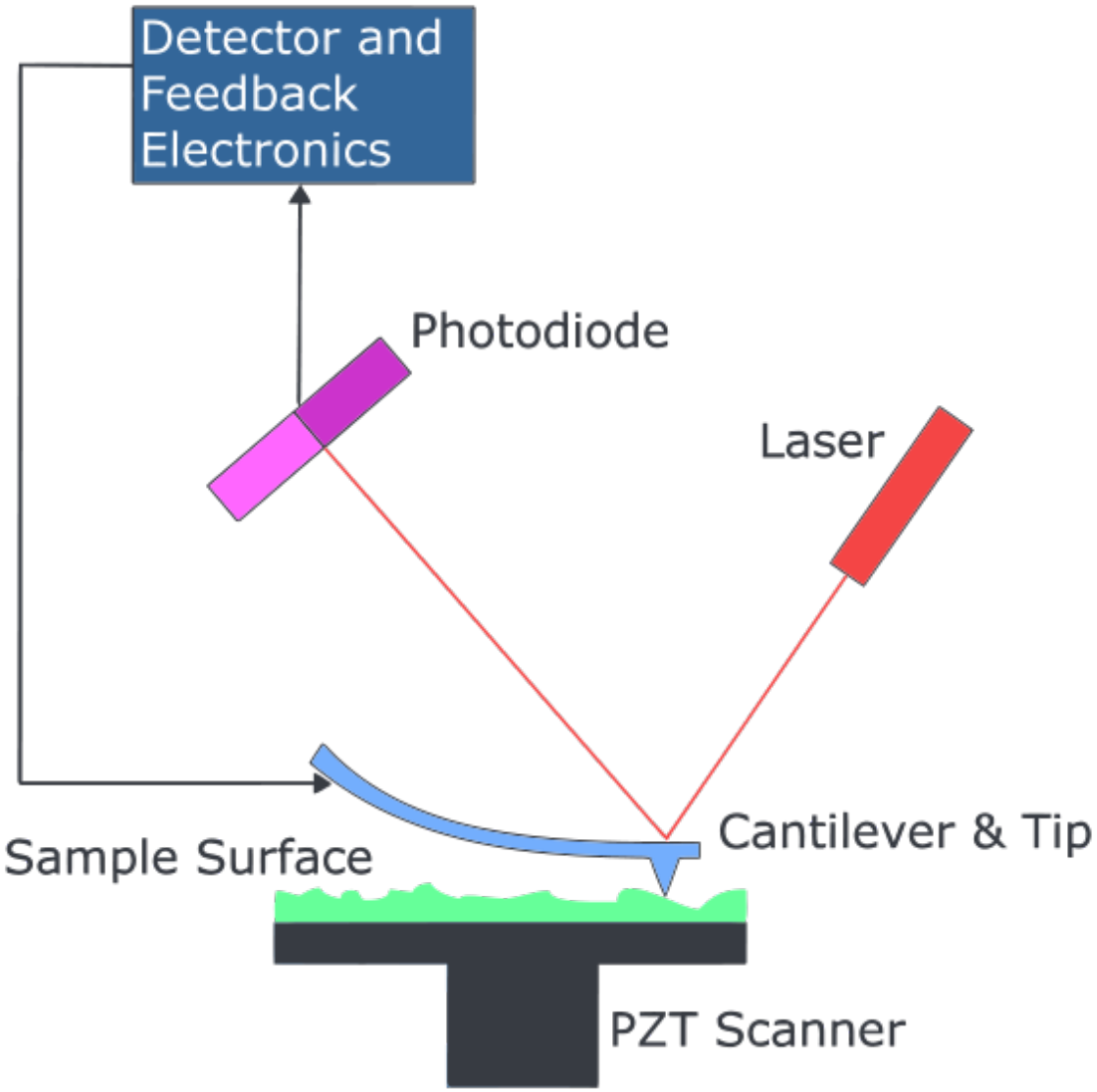
This term was used for the area of physics dealing with the effects of quantum mechanics on the behavior of electrons in matter, and their interactions with photons. It is today rarely considered a sub-field in its own right, as it has been absorbed by other fields. Solid state physics regularly takes quantum mechanics into account, and is usually concerned with electrons. Specific application to electronics is researched within semiconductor physics. The term also encompassed the basic processes of laser operation where photons are interacting with electrons: absorption, spontaneous emission, and stimulated emission. The term was mainly used between the 1950s and the 1970s. Today, the research output of this field is mainly used in quantum optics, especially for the part of it that draws not from atomic physics but from solid-state physics. Its usage overlapped Quantum Hall effect and Quantum cellular automata.

Chapter 4

Atomic Force Microscopy



A commercial AFM setup

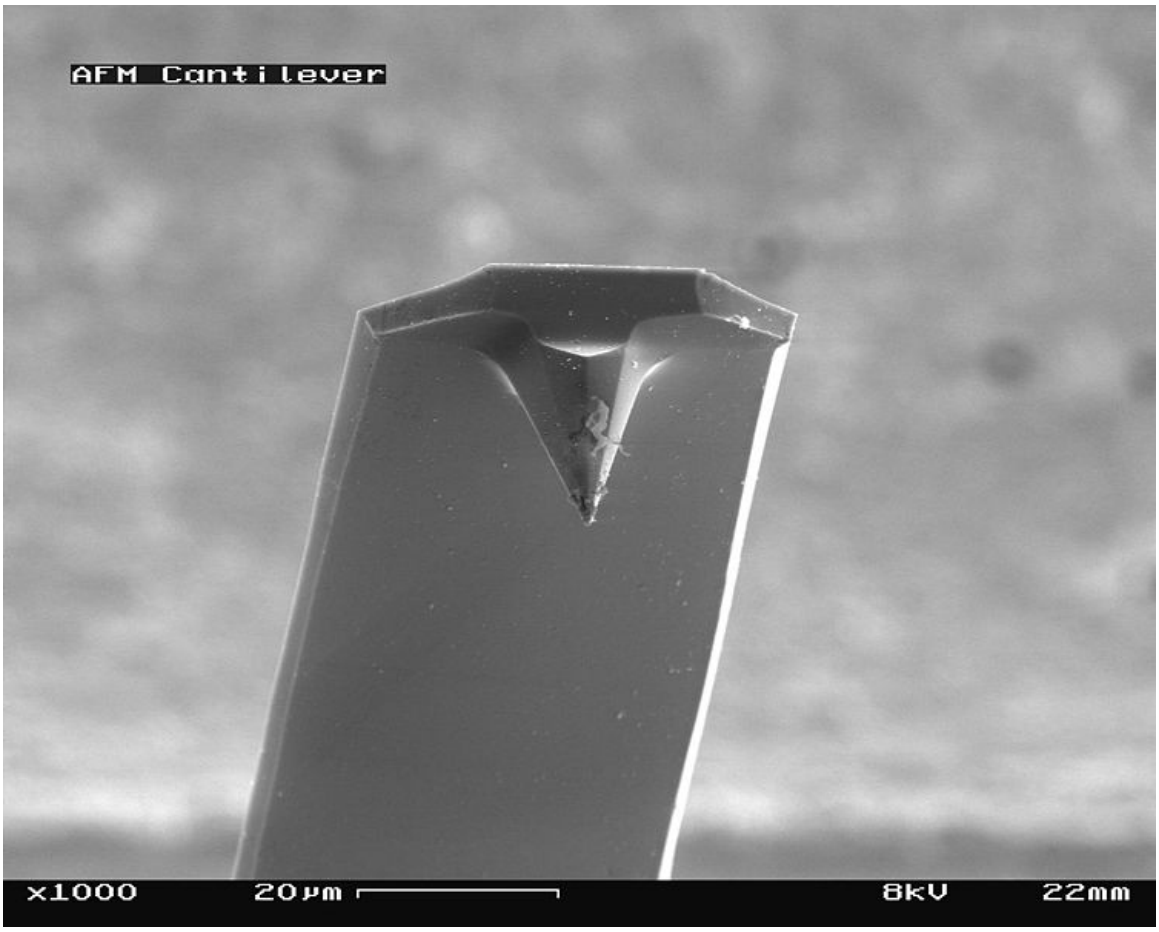


Block diagram of atomic force microscope

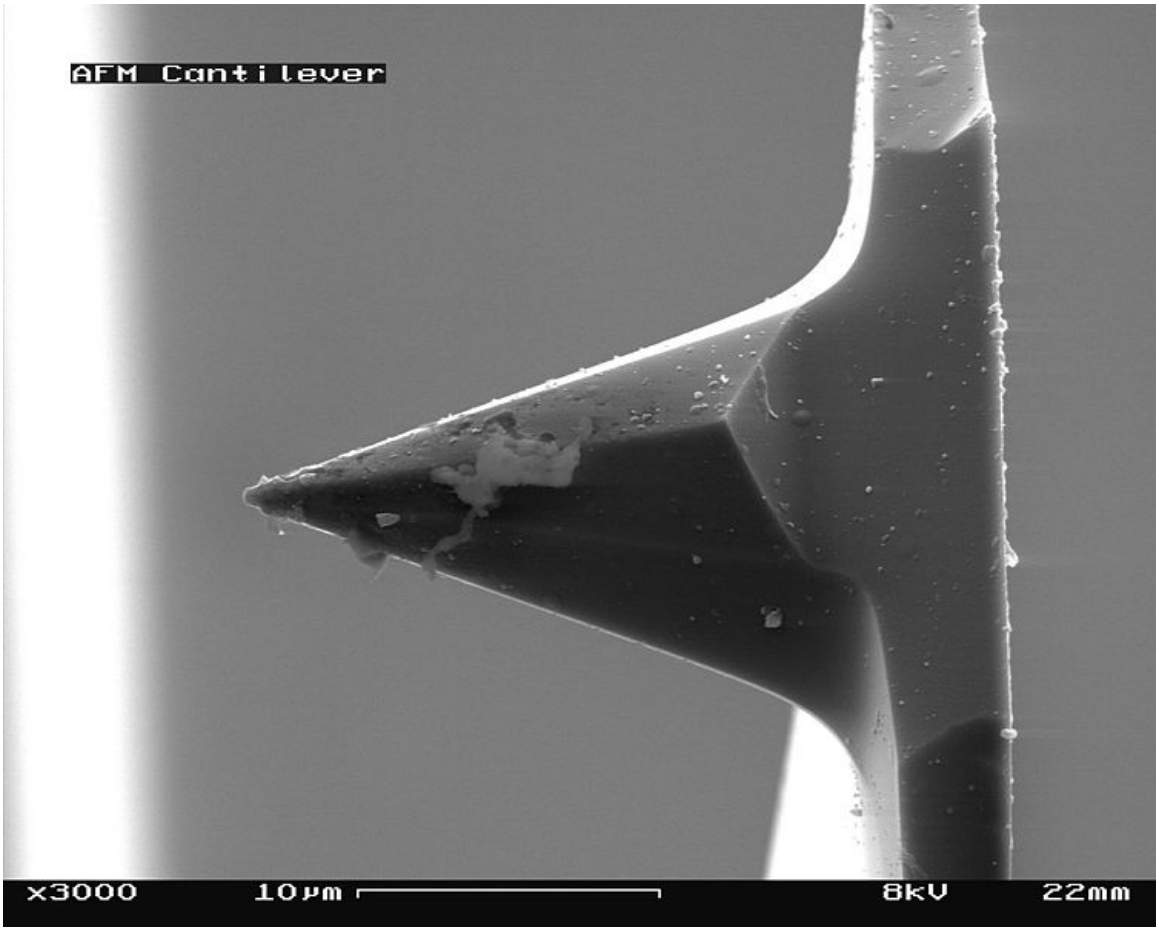
Atomic force microscopy (AFM) or scanning force microscopy (SFM) is a very high-resolution type of scanning probe microscopy, with demonstrated resolution on the order of fractions of a nanometer, more than 1000 times better than the optical diffraction limit. The precursor to the AFM, the scanning tunneling microscope, was developed by Gerd Binnig and Heinrich Rohrer in the early 1980s at IBM Research - Zurich, a development that earned them the Nobel Prize for Physics in 1986. Binnig, Quate and Gerber invented the first atomic force microscope (also abbreviated as AFM) in 1986. The first commercially available atomic force microscope was introduced in 1989. The AFM is one of the foremost tools for imaging, measuring, and manipulating matter at the nanoscale. The information is gathered by "feeling" the surface with a mechanical probe. Piezoelectric elements that facilitate tiny but accurate and precise movements on (electronic) command enable the very precise scanning. In some variations, electric

potentials can also be scanned using conducting cantilevers. In newer more advanced versions, currents can even be passed through the tip to probe the electrical conductivity or transport of the underlying surface, but this is much more challenging with very few research groups reporting reliable data.

Basic principles

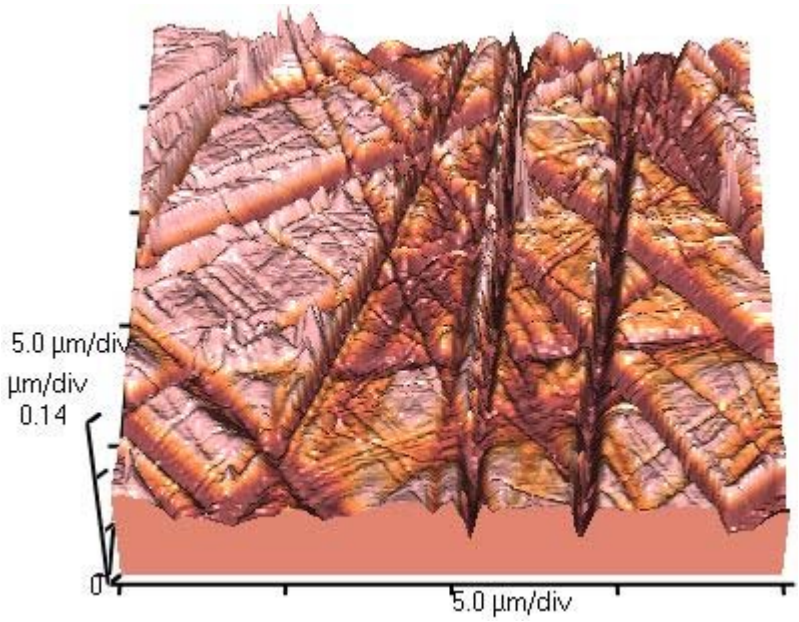


Electron micrograph of a used AFM cantilever image width ~100 micrometers...



and ~30 micrometers

The AFM consists of a cantilever with a sharp tip (probe) at its end that is used to scan the specimen surface. The cantilever is typically silicon or silicon nitride with a tip radius of curvature on the order of nanometers. When the tip is brought into proximity of a sample surface, forces between the tip and the sample lead to a deflection of the cantilever according to Hooke's law. Depending on the situation, forces that are measured in AFM include mechanical contact force, van der Waals forces, capillary forces, chemical bonding, electrostatic forces, magnetic forces, Casimir forces, solvation forces, etc. Along with force, additional quantities may simultaneously be measured through the use of specialized types of probe. Typically, the deflection is measured using a laser spot reflected from the top surface of the cantilever into an array of photodiodes. Other methods that are used include optical interferometry, capacitive sensing or piezoresistive AFM cantilevers. These cantilevers are fabricated with piezoresistive elements that act as a strain gauge. Using a Wheatstone bridge, strain in the AFM cantilever due to deflection can be measured, but this method is not as sensitive as laser deflection or interferometry.



Atomic force microscope topographical scan of a glass surface. The micro and nano-scale features of the glass can be observed, portraying the roughness of the material. The image space is (x,y,z) = (20um x 20um x 420nm).

If the tip was scanned at a constant height, a risk would exist that the tip collides with the surface, causing damage. Hence, in most cases a feedback mechanism is employed to adjust the tip-to-sample distance to maintain a constant force between the tip and the sample. Traditionally, the sample is mounted on a piezoelectric tube, that can move the sample in the *z* direction for maintaining a constant force, and the *x* and *y* directions for scanning the sample. Alternatively a 'tripod' configuration of three piezo crystals may be employed, with each responsible for scanning in the x,y and z directions. This eliminates some of the distortion effects seen with a tube scanner. In newer designs, the tip is mounted on a vertical piezo scanner while the sample is being scanned in X and Y using another piezo block. The resulting map of the area $z = f(x,y)$ represents the topography of the sample.

The AFM can be operated in a number of modes, depending on the application. In general, possible imaging modes are divided into static (also called *contact*) modes and a variety of dynamic (or non-contact) modes where the cantilever is vibrated.

Imaging modes

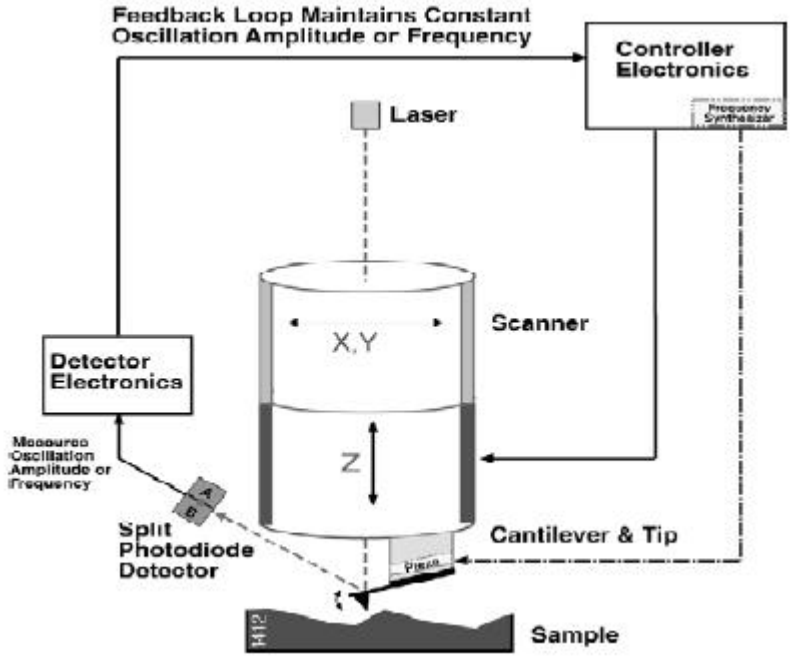
The primary modes of operation for an AFM are static mode and dynamic mode. In static mode, the cantilever is "dragged" across the surface of the sample and the contours of the surface are measured directly using the deflection of the cantilever. In the dynamic mode, the cantilever is externally oscillated at or close to its fundamental resonance frequency or a harmonic. The oscillation amplitude, phase and resonance frequency are modified by

tip-sample interaction forces. These changes in oscillation with respect to the external reference oscillation provide information about the sample's characteristics.

Contact mode

In the static mode operation, the static tip deflection is used as a feedback signal. Because the measurement of a static signal is prone to noise and drift, low stiffness cantilevers are used to boost the deflection signal. However, close to the surface of the sample, attractive forces can be quite strong, causing the tip to "snap-in" to the surface. Thus static mode AFM is almost always done in contact where the overall force is repulsive. Consequently, this technique is typically called "contact mode". In contact mode, the force between the tip and the surface is kept constant during scanning by maintaining a constant deflection.

Non-contact mode



AFM - non-contact mode

In this mode, the tip of the cantilever does not contact the sample surface. The cantilever is instead oscillated at a frequency slightly above its resonant frequency where the amplitude of oscillation is typically a few nanometers (<10 nm). The van der Waals forces, which are strongest from 1 nm to 10 nm above the surface, or any other long range force which extends above the surface acts to decrease the resonance frequency of the cantilever. This decrease in resonant frequency combined with the feedback loop system maintains a constant oscillation amplitude or frequency by adjusting the average tip-to-sample distance. Measuring the tip-to-sample distance at each (x,y) data point allows the scanning software to construct a topographic image of the sample surface.

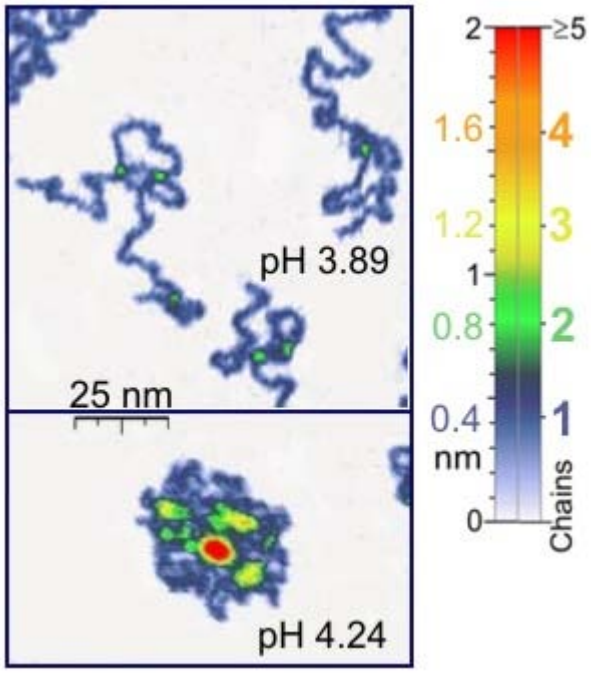
Non-contact mode AFM does not suffer from tip or sample degradation effects that are sometimes observed after taking numerous scans with contact AFM. This makes non-contact AFM preferable to contact AFM for measuring soft samples. In the case of rigid samples, contact and non-contact images may look the same. However, if a few monolayers of adsorbed fluid are lying on the surface of a rigid sample, the images may look quite different. An AFM operating in contact mode will penetrate the liquid layer to image the underlying surface, whereas in non-contact mode an AFM will oscillate above the adsorbed fluid layer to image both the liquid and surface.

Schemes for dynamic mode operation include frequency modulation and the more common amplitude modulation. In frequency modulation, changes in the oscillation frequency provide information about tip-sample interactions. Frequency can be measured with very high sensitivity and thus the frequency modulation mode allows for the use of very stiff cantilevers. Stiff cantilevers provide stability very close to the surface and, as a result, this technique was the first AFM technique to provide true atomic resolution in ultra-high vacuum conditions.

In amplitude modulation, changes in the oscillation amplitude or phase provide the feedback signal for imaging. In amplitude modulation, changes in the phase of oscillation can be used to discriminate between different types of materials on the surface. Amplitude modulation can be operated either in the non-contact or in the intermittent contact regime. In dynamic contact mode, the cantilever is oscillated such that the separation distance between the cantilever tip and the sample surface is modulated.

Amplitude modulation has also been used in the non-contact regime to image with atomic resolution by using very stiff cantilevers and small amplitudes in an ultra-high vacuum environment.

Tapping mode



Single polymer chains (0.4 nm thick) recorded in a tapping mode under aqueous media with different pH.

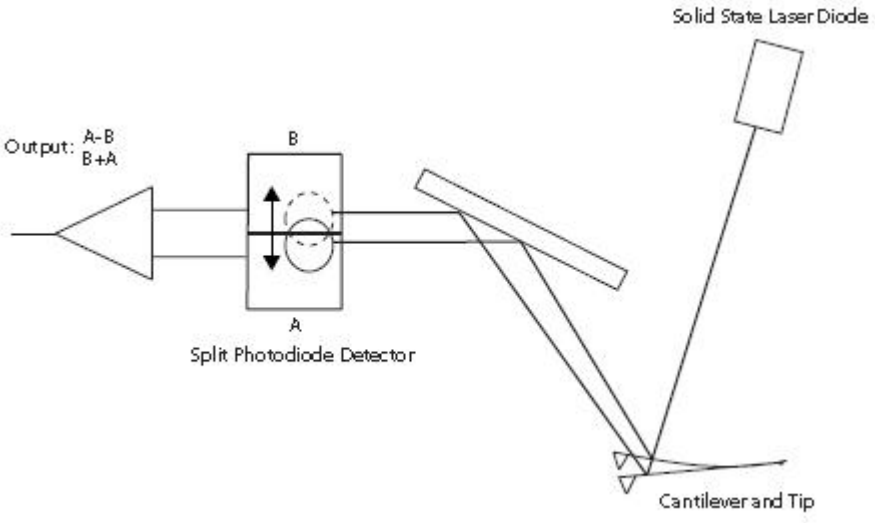
In ambient conditions, most samples develop a liquid meniscus layer. Because of this, keeping the probe tip close enough to the sample for short-range forces to become detectable while preventing the tip from sticking to the surface presents a major problem for non-contact dynamic mode in ambient conditions. Dynamic contact mode (also called intermittent contact or tapping mode) was developed to bypass this problem.

In *tapping mode*, the cantilever is driven to oscillate up and down at near its resonance frequency by a small piezoelectric element mounted in the AFM tip holder similar to non-contact mode. However, the amplitude of this oscillation is greater than 10 nm, typically 100 to 200 nm. Due to the interaction of forces acting on the cantilever when the tip comes close to the surface, Van der Waals force, dipole-dipole interaction, electrostatic forces, etc. cause the amplitude of this oscillation to decrease as the tip gets closer to the sample. An electronic servo uses the piezoelectric actuator to control the height of the cantilever above the sample. The servo adjusts the height to maintain a set cantilever oscillation amplitude as the cantilever is scanned over the sample. A *tapping AFM* image is therefore produced by imaging the force of the intermittent contacts of the tip with the sample surface.

This method of "tapping" lessens the damage done to the surface and the tip compared to the amount done in contact mode. Tapping mode is gentle enough even for the visualization of supported lipid bilayers or adsorbed single polymer molecules (for instance, 0.4 nm thick chains of synthetic polyelectrolytes) under liquid medium. With

proper scanning parameters, the conformation of single molecules can remain unchanged for hours.

AFM cantilever deflection measurement



AFM beam deflection detection

Laser light from a solid state diode is reflected off the back of the cantilever and collected by a position sensitive detector (PSD) consisting of two closely spaced photodiodes whose output signal is collected by a differential amplifier. Angular displacement of the cantilever results in one photodiode collecting more light than the other photodiode, producing an output signal (the difference between the photodiode signals normalized by their sum) which is proportional to the deflection of the cantilever. It detects cantilever deflections <10 nm (thermal noise limited). A long beam path (several centimeters) amplifies changes in beam angle.

Force spectroscopy

Another major application of AFM (besides imaging) is force spectroscopy, the direct measurement of tip-sample interaction forces as a function of the gap between the tip and sample (the result of this measurement is called a force-distance curve). For this method, the AFM tip is extended towards and retracted from the surface as the deflection of the cantilever is monitored as a function of piezoelectric displacement. These measurements have been used to measure nanoscale contacts, atomic bonding, Van der Waals forces, and Casimir forces, dissolution forces in liquids and single molecule stretching and rupture forces. Furthermore, AFM was used to measure, in an aqueous environment, the dispersion force due to polymer adsorbed on the substrate. Forces of the order of a few piconewtons can now be routinely measured with a vertical distance resolution of better than 0.1 nanometers. Force spectroscopy can be performed with either static or dynamic modes. In dynamic modes, information about the cantilever vibration is monitored in addition to the static deflection.

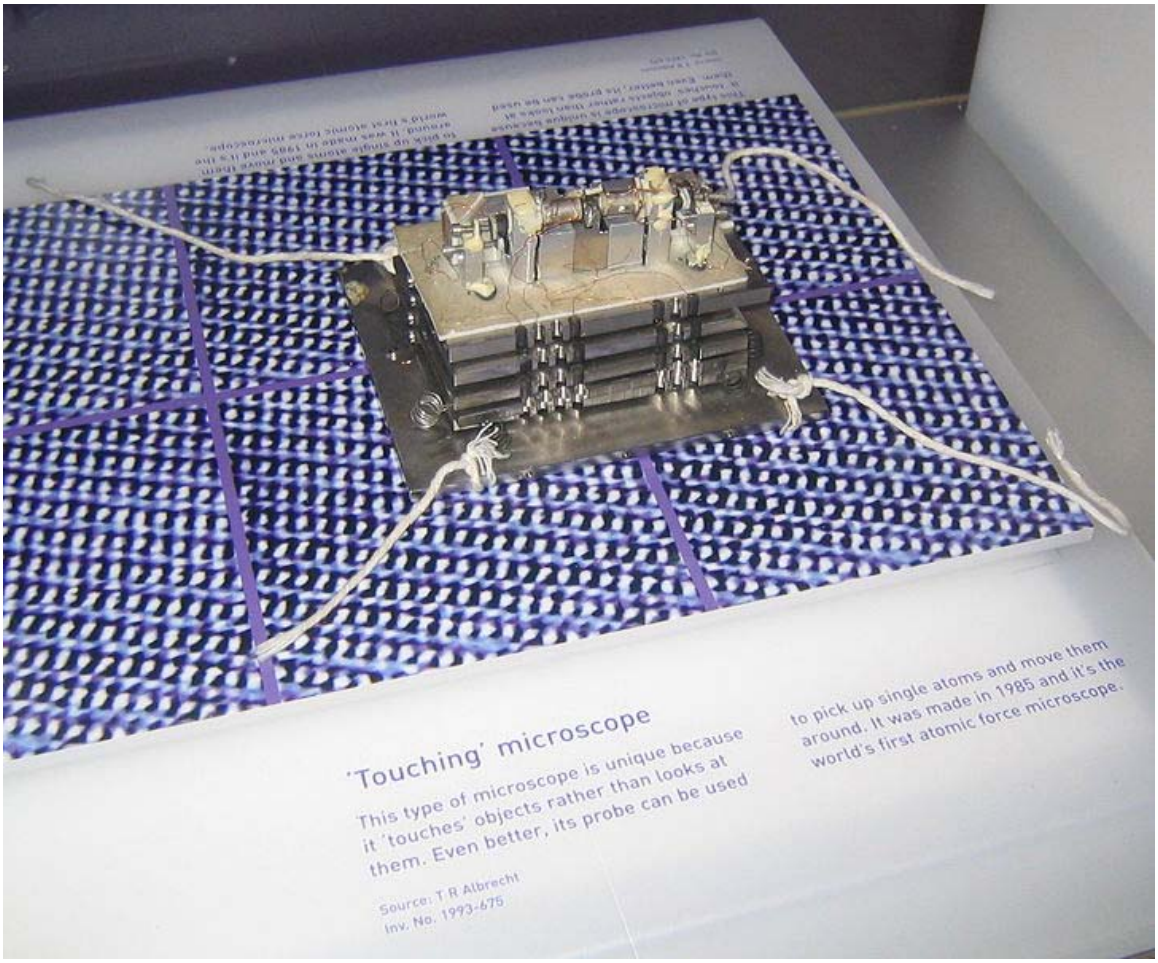
Problems with the technique include no direct measurement of the tip-sample separation and the common need for low stiffness cantilevers which tend to 'snap' to the surface. The snap-in can be reduced by measuring in liquids or by using stiffer cantilevers, but in the latter case a more sensitive deflection sensor is needed. By applying a small dither to the tip, the stiffness (force gradient) of the bond can be measured as well.

Identification of individual surface atoms

The AFM can be used to image and manipulate atoms and structures on a variety of surfaces. The atom at the apex of the tip "senses" individual atoms on the underlying surface when it forms incipient chemical bonds with each atom. Because these chemical interactions subtly alter the tip's vibration frequency, they can be detected and mapped. This principle was used to distinguish between atoms of silicon, tin and lead on an alloy surface, by comparing these 'atomic fingerprints' to values obtained from large-scale density functional theory (DFT) simulations.

The trick is to first measure these forces precisely for each type of atom expected in the sample, and then to compare with forces given by DFT simulations. The team found that the tip interacted most strongly with silicon atoms, and interacted 23% and 41% less strongly with tin and lead atoms, respectively. Thus, each different type of atom can be identified in the matrix as the tip is moved across the surface.

Advantages and disadvantages



The first atomic force microscope

Just like any other tool, an AFM's usefulness has limitations. When determining whether or not analyzing a sample with an AFM is appropriate, there are various advantages and disadvantages that must be considered.

Advantages

AFM has several advantages over the scanning electron microscope (SEM). Unlike the electron microscope which provides a two-dimensional projection or a two-dimensional image of a sample, the AFM provides a three-dimensional surface profile. Additionally, samples viewed by AFM do not require any special treatments (such as metal/carbon coatings) that would irreversibly change or damage the sample. While an electron microscope needs an expensive vacuum environment for proper operation, most AFM modes can work perfectly well in ambient air or even a liquid environment. This makes it possible to study biological macromolecules and even living organisms. In principle, AFM can provide higher resolution than SEM. It has been shown to give true atomic resolution in ultra-high vacuum (UHV) and, more recently, in liquid environments. High

resolution AFM is comparable in resolution to scanning tunneling microscopy and transmission electron microscopy.

Disadvantages

A disadvantage of AFM compared with the scanning electron microscope (SEM) is the single scan image size. In one pass, the SEM can image an area on the order of square millimeters with a depth of field on the order of millimeters. Whereas the AFM can only image a maximum height on the order of 10-20 micrometers and a maximum scanning area of about 150×150 micrometers. One method of improving the scanned area size for AFM is by using parallel probes in a fashion similar to that of millipede data storage.

The scanning speed of an AFM is also a limitation. Traditionally, an AFM cannot scan images as fast as a SEM, requiring several minutes for a typical scan, while a SEM is capable of scanning at near real-time, although at relatively low quality. The relatively slow rate of scanning during AFM imaging often leads to thermal drift in the image making the AFM microscope less suited for measuring accurate distances between topographical features on the image. However, several fast-acting designs were suggested to increase microscope scanning productivity including what is being termed videoAFM (reasonable quality images are being obtained with videoAFM at video rate: faster than the average SEM). To eliminate image distortions induced by thermal drift, several methods have been introduced.

AFM images can also be affected by hysteresis of the piezoelectric material and cross-talk between the *x*, *y*, *z* axes that may require software enhancement and filtering. Such filtering could "flatten" out real topographical features. However, newer AFMs utilize closed-loop scanners which practically eliminate these problems. Some AFMs also use separated orthogonal scanners (as opposed to a single tube) which also serve to eliminate part of the cross-talk problems.

As with any other imaging technique, there is the possibility of image artifacts, which could be induced by an unsuitable tip, a poor operating environment, or even by the sample itself. These image artifacts are unavoidable however, their occurrence and effect on results can be reduced through various methods.

Due to the nature of AFM probes, they cannot normally measure steep walls or overhangs. Specially made cantilevers and AFMs can be used to modulate the probe sideways as well as up and down (as with dynamic contact and non-contact modes) to measure sidewalls, at the cost of more expensive cantilevers, lower lateral resolution and additional artifacts.

Piezoelectric scanners

AFM scanners are made from piezoelectric material, which expands and contracts proportionally to an applied voltage. Whether they elongate or contract depends upon the polarity of the voltage applied. The scanner is constructed by combining independently

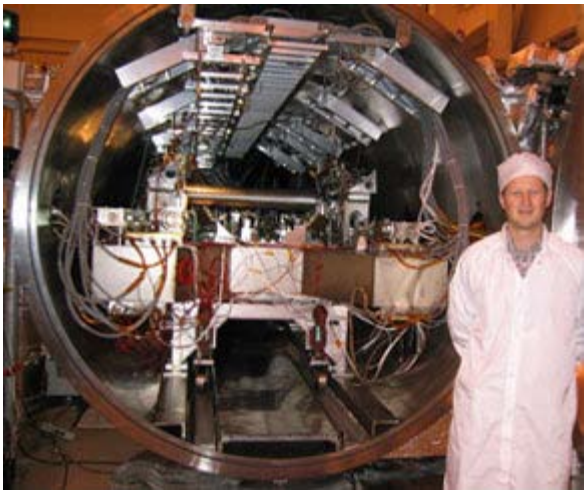
operated piezo electrodes for X, Y, and Z into a single tube, forming a scanner which can manipulate samples and probes with extreme precision in 3 dimensions.

Scanners are characterized by their sensitivity which is the ratio of piezo movement to piezo voltage, i.e., by how much the piezo material extends or contracts per applied volt. Because of differences in material or size, the sensitivity varies from scanner to scanner. Sensitivity varies non-linearly with respect to scan size. Piezo scanners exhibit more sensitivity at the end than at the beginning of a scan. This causes the forward and reverse scans to behave differently and display hysteresis between the two scan directions. This can be corrected by applying a non-linear voltage to the piezo electrodes to cause linear scanner movement and calibrating the scanner accordingly.

The sensitivity of piezoelectric materials decreases exponentially with time. This causes most of the change in sensitivity to occur in the initial stages of the scanner's life. Piezoelectric scanners are run for approximately 48 hours before they are shipped from the factory so that they are past the point where they may have large changes in sensitivity. As the scanner ages, the sensitivity will change less with time and the scanner would seldom require recalibration.

Chapter 5

Metrology



A scientist stands in front of a microarcsecond (1 millionth of 1 arcsecond or 1 millionth of 1/3600 degree) testbed.

Metrology is the science of measurement. Metrology includes all theoretical and practical aspects of measurement. The word comes from Greek μέτρον (*metron*), "measure" + "λόγος" (*logos*), amongst others meaning "speech, oration, discourse, quote, study, calculation, reason". In Ancient Greek the term μετρολογία (*metrologia*) meant "theory of ratios".

Introduction

Metrology is defined by the *International Bureau of Weights and Measures* (BIPM) as "the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology." The *ontology* and international vocabulary of metrology (VIM) is maintained by the International Organisation for Standardisation.

Metrology is a very broad field and may be divided into three subfields:

Subfield	Definition
Scientific or fundamental metrology	concerns the establishment of *quantity systems*, unit systems, *units of measurement*, the development of new measurement methods, realisation of measurement standards and the transfer of traceability from these standards to users in society.
Applied or industrial metrology	concerns the application of measurement science to manufacturing and other processes and their use in society, ensuring the suitability of measurement instruments, their calibration and quality control of measurements.
Legal metrology	concerns regulatory requirements of measurements and measuring instruments for the protection of health, public safety, the environment, enabling taxation, protection of consumers and fair trade.

A core concept in metrology is (metrological) traceability, defined as "the property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons, all having stated uncertainties." The level of traceability establishes the level of comparability of the measurement: whether the result of a measurement can be compared to the previous one, a measurement result a year ago, or to the result of a measurement performed anywhere else in the world.

Traceability is most often obtained by calibration, establishing the relation between the indication of a measuring instrument and the value of a measurement standard. These standards are usually coordinated by national metrological institutes: National Institute of Standards and Technology, National Physical Laboratory, UK, Physikalisch-Technische Bundesanstalt, etc.

Tracebility, accuracy, precision, systematic bias, evaluation of measurement uncertainty are critical parts of a quality management system.

Basics

Mistakes can make measurements and counts incorrect. Even if there are no mistakes, nearly all measurements are still inexact. The term 'error' is reserved for that inexactness, also called measurement uncertainty. Among the few exact measurements are:

- The absence of the quantity being measured, such as a voltmeter with its leads shorted together: the meter should read zero exactly.
- Measurement of an accepted constant under qualifying conditions, such as the triple point of pure water: the thermometer should read 273.16 kelvin (0.01 degrees Celsius, 32.018 degrees Fahrenheit) when qualified equipment is used correctly.

- Self-checking ratio metric measurements, such as a potentiometer: the ratio in between steps is independently adjusted and verified to be beyond influential inexactness.

All other measurements either have to be checked to be sufficiently correct or left to chance. Metrology is the science that establishes the correctness of specific measurement situations. This is done by anticipating and allowing for both mistakes and error. The precise distinction between measurement error and mistakes is not settled and varies by country. Repeatability and reproducibility studies help quantify the precision: one common method is an ANOVA Gauge R&R study.

Calibration is the process where metrology is applied to measurement equipment and processes to ensure conformity with a known standard of measurement, usually traceable to a national standards board.

Society

Sufficiently correct measurements are essential to commerce. About nine out of every ten people working in metrology specialize in commercial measurement, most at the technician level. Correct measurements are beneficial to manufacturing, but other methods are available and sometimes are more appropriate.

Metrology has thrived at the interface between science and manufacturing. Aerospace, commercial nuclear power, medicine, medical devices and semiconductors rely on metrology to translate theoretical science into mass produced reality.

The basic concepts of metrology appear simple on the surface, and metrology is rarely taught in a systematic manner above the technician level. Within most businesses, metrology core beliefs such as recording all setups and observations for possible future reference are opposed to the general business practice of minimizing recordkeeping to limit litigation effects.

Applied metrology

Metrology laboratories are places where both metrology and calibration work are performed. Calibration laboratories generally specialize in calibration work only.

Both metrology and calibration laboratories must isolate the work performed from influences that might affect the work. Temperature, humidity, vibration, electrical power supply, radiated energy and other influences are often controlled. Generally, it is the rate of change or instability that is more detrimental than whatever value prevails.

Calibration technicians execute calibration work. In large organizations, the work is further divided into three groups:

Group	Definition
Set-up people	arrange the equipment needed for calibration and verify that it works correctly.
Operators	execute the calibration procedures and collect data.
Tear-down people	dismantle set-ups, check the components for damage and then put the components into a stored state. This is the entry-level position for people who didn't start in the equipment warehouse or transportation functions

Alternately, the technicians can be divided by major discipline areas: physical, dimensional, electrical, RF, microwave and so on. But the principles are the same regardless of the equipment.

Metrology technicians perform investigation work in addition to calibrations. They also apply proven principles to known situations and evaluate unexpected or contradictory results.

Specific education in metrology was formerly limited to sub-professional work. Most of the branches of the US Military train 'enlisted-grade' technicians to meet their specific needs.

Large industrial organizations also develop people who demonstrate aptitude in testing functions. When this is combined with an engineering degree, it qualifies the person as a metrology engineer. Over the last 15 years, Universities such as the University of North Carolina at Charlotte created a specific curriculum in metrology engineering. In England, metrology was part of the fifth year of some undergraduate engineering programmes.

Metrologists are people who perform metrology work at and above the technician levels, generally without the benefit or acknowledgement of a college degree.

The metrology and calibration work described above is always accompanied by documentation. The documentation can be divided into two types; one related to the task and the other related the administrative program. Task documentation includes calibration procedures and the data collected. Administrative program documentation includes equipment identification data, 'calibration certificates', calibration time interval information and 'as-found' or 'out-of-tolerance' notifications.

Administrative programs provide standardization of the metrology and calibration work and make it possible to independently verify that the work was performed. Generally, the administrative program is specific to the organization performing the work and addresses customer requirements. General administrative program specifications created by industry groups, such as the ANS (ANSI) Z540 series may also be covered in the administrative program. Other specifications created by the US Food and Drug Administration, US Federal Aviation Administration or other agencies would supplement

or replace ANS Z540 for work performed in their domains. Often administrative programs can be as complicated and detailed as the measurement work itself.

An administrative program that has insufficient actual metrology or calibration capability is derisively referred to as a "lick and stick" program.

Standards

Standards are objects or ideas that are designated as being authoritative for some accepted reason. Whatever value they possess is useful for comparison to unknowns for the purpose of establishing or confirming an assigned value based on the standard. The design of this comparison process for measurements is metrology. The execution of measurement comparisons for the purpose of establishing the relationship between a standard and some other measuring device is calibration.

The ideal standard is independently reproducible without uncertainty. This is what the creators of the “meter” length standard were attempting to do in the 19th century when they defined a meter as one ten-millionth of the distance from the equator to one of the Earth’s poles. Later, it was learned that the Earth’s surface is an unreliable basis for a standard. The Earth is not spherical and it is constantly changing in shape. But the special alloy meter bars that were created and accepted in that time period standardized international length measurement until the 1950s. Careful calibrations allowed tolerances as small as 10 parts per million to be distributed and reproduced in metrology laboratories worldwide, regardless of whether the rest of the metric system was implemented and in spite of the shortfalls of the meter’s original basis.

Historical International Prototype Meter bars

Modern standards

Currently, only five independent units of measure are internationally recognized: temperature interval, linear distance, electrical current, frequency and mass. All measurements of all types are based on one or more of these independent units. Two supplemental independent units are also recognized internationally, both dealing with angle measurement.

For example, Ohm's law is a widely known concept in electrical study. Of the three units of measure involved, only current (ampere) is an independent unit. Voltage and resistance units are dependent on current units, as defined by Ohm's law.

In the United States, ASTM Standard Practice E 380,replaced by IEEE/ASTM SI10 , adapts independent unit of measure theory to practical measurement activity.

It is believed that each of independent units of measure will be defined in terms of the other four independent units eventually. Length (meter) and time (second) are already connected this way. If an accurate time base is available, then a length standard can be reproduced without a meter bar artifact, using the known constant speed of light. Lesser known is the relationship between the luminance (candela) and current (ampere). The candela is defined in terms of the watt, which in turn is derived from the ampere. This difficult to recreate standard is supplemented by an incandescent bulb design that is used as a secondary and transfer standard. These bulbs recreate the candela when a specific amount of current is applied.

The development of standards follows the needs of technology. As a result, some units of measure have much more resolution than others. The second is reproducible to 1 part in 10^{14}. As it became possible to measure time more precisely, solar time, believed to be a constant, proved to be very slightly irregular. This resulted in leap second adjustments to keep UTC synchronised with solar time.

Luminance (candela) can only be reproduced to 5% of reading despite having sensors that have accuracies of +/- 50 parts per million (0.005%) precision. This is due to the standard not being accurately reproducible.

Temperature (kelvin) is defined by agreed fixed points. These points are defined by the state changes of nearly pure materials, generally as they move from liquid to solid. Between these fixed points, Standard Platinum Resistance Thermometers (SPRTs), constructed a specified manner, are used to interpolate temperature values. This mosaic of approaches produces measurement uncertainty which is not uniform over the entire range of temperature measurement. Temperature measurement is coordinated by the International Practical Temperature Scale, maintained by the BIPM.

These non-commercial measurement details used to be academic curiosities. However, engineering, manufacturing and ordinary living now routinely challenge the limits of measurement.

Industry-specific standards

In addition to standards created by national and international standards organizations, many large and small industrial companies also define metrology standards and procedures to meet their particular needs for technically and economically competitive manufacturing. These standards and procedures, while drawing in part upon the national and international standards, also address the issues of what specific instrument

technology will be used to measure each quantity, how often each quantity will be measured, and which definition of each quantity will be used as the basis for accomplishing the process control that their manufacturing and product specifications require. Industrial metrology standards include dynamic control plans, also known as “dimensional control plans”, or “DCPs”, for their products.

In industrial metrology, several issues beyond accuracy constrain the usability of metrology methods. These include

- The speed with which measurements can be accomplished on parts or surfaces in the process of manufacturing, which must match the TAKT Time of the production line.
- The completeness with which the manufactured part can be measured such as described in high-definition metrology,
- The ability of the measurement mechanism to operate reliably in a manufacturing plant environment considering temperature, vibration, dust, and a host of other potential hostile factors,
- The ability of the measurement results, as they are presented, to be assimilated by the manufacturing operators or automation in time to effectively control the manufacturing process variables, and
- The total financial cost of measuring each part.

National standards

Every country maintains its own metrology system. In the United States, the National Institute of Standards and Technology (NIST) plays the dual role of maintaining and furthering both commercial and scientific metrology. NIST does not enforce measurement accuracy directly.

The accuracy and traceability of commercial measurements is enforced per the laws of the individual states. Commercial measurement generally involves any material sold by any unit of measure. Some intuitive or obvious measurement is generally exempted, such as selling cloth on a cutting table that has a yardstick fastened to it. All counting-based transactions are generally exempt also. But each state has its own rules, responding to the accumulated concerns of the state residents.

Commercial metrology is also known as "weights and measures" and is essential to commerce of any kind above the pure barter level. Every state maintains its own weights and measures functionality with traceability to the national standards maintained by NIST. Large states further divide this effort by county, where a "Sealer" or other appointee is responsible for the validity of most common commercial measurements such as mass balances (scales) in grocery stores and gasoline pump measurements of volume. The sealer's staff and agents make periodic inspections to catch merchant cheaters, maintaining the integrity of commercial measurements.

Typical State Seal application.

Depending on the specific state, other state government agencies can be involved. For example, electricity watt-hour meters and water delivery flow meters are commonly monitored by the state's "public utilities commission" who enforces the measurement tolerances and traceability to NIST through the utility providers. Highway State Police and the State Highway Department generally run the commercial truck weight measurement programs for safety purposes and to minimize the damage to road surfaces that overloaded trucks cause. Nearly all states license weighmasters, weighmistresses, scale calibrators and other specialists involved in commercial measuring equipment maintenance.

The term "commercial metrology" is also used to describe calibration laboratories that are not owned by the companies they serve.

Scientific metrology addresses measurement phenomena not quantified in ordinary commerce, such as the test bed pictured at the beginning. Calibration laboratories that serve scientific metrology are regulated as businesses only. They may choose to have their work accredited by voluntary certification organizations based on customer desires, but there is no requirement to do so. Irresolvable disputes involving scientific metrology are generally settled in the civil court systems. Some federal government entities like the Federal Communications Commission and the Environmental Protection Administration are considered to be the final authority in their domains rather than the NIST. Disputes

involving only metrology issues with those organizations probably would not be heard in any courts.

Historical development

Metrology has existed in some form or another since antiquity. The earliest forms of metrology were simply arbitrary standards set up by regional or local authorities, often based on practical measures such as the length of an arm. The earliest examples of these standardized measures are length, time, and weight. These standards were established in order to facilitate commerce and record human activity.

Little progress was made with regard to proto-metrology until various scientists, chemists, and physicists started making headway during the scientific revolution. With the advances in the sciences, the comparison of experiment to theory required a rational system of units, and something more closely resembling modern metrology began to come into being. The discovery of atoms, electricity, thermodynamics, and other fundamental scientific principles could be applied to standards of measurement, and many inventions made it easier to quantitatively or qualitatively assess physical properties, using the defined units of measurement established by science.

Metrology was thus one of the precursors to the Industrial Revolution, and was necessary for the implementation of mass production, equipment commonality, and assembly lines.

Modern metrology has its roots in the French Revolution, with the political motivation to harmonize units all over France and the concept of establishing units of measurement based on constants of nature, and thus making measurement units available "for all people, for all time". In this case deriving a unit of length from the dimensions of the Earth, and a unit of mass from a cube of water. The result was platinum standards for the meter and the kilogram established as the basis of the metric system on June 22, 1799. This further led to the creation of the Système International d'Unités, or the International System of Units. This system has gained unprecedented worldwide acceptance as definitions and standards of modern measurement units. Though not the official system of units of all nations, the definitions and specifications of SI are globally accepted and recognized. The SI is maintained under the auspices of the Metre Convention and its institutions, the General Conference on Weights and Measures, or CGPM, its executive branch the International Committee for Weights and Measures, or CIPM, and its technical institution the International Bureau of Weights and Measures, or BIPM.

As the authorities on SI, these organizations establish and promulgate the SI, with the ambition to be able to service all. This includes introducing new units, such as the relatively new unit, the mole, to encompass metrology in chemistry. These units are then established and maintained through various agencies in each country, and establish a hierarchy of measurement standards that can be traced back to the established standard unit, a concept known as metrological traceability. The U.S. agencies holding this responsibility are the National Institute of Standards and Technology (NIST) and the American National Standards Institute (ANSI).

The development of standards also does involve individual and small group achievements. In 1893, Edward Weston (chemist) and his company perfected his Saturated Standard Cell design, which allowed the volt to be reproduced to 1 part in ten to the fourth power directly. This advance made a huge practical difference at a critical moment in the development of modern electrical devices. Groupings of saturated cells, called banks, can still be found in some metrology and calibration laboratories today. Edward Weston did not pursue patents for his cell design. By doing this, his superior design quickly replaced similar but inferior patented devices worldwide without much discussion.

Mechanisms

At the base of metrology is the definition, realisation and dissemination of units of measurement. Physical or chemical properties are quantised by assigning a property value in some multiple of a measurement unit.

The basic 'lineage' of measurement standards are:

- The definition of a unit, based on some physical constant, such as absolute zero, the freezing point of water, etc.; or an agreed-upon arbitrary standard.
- The realisation of the unit by experimental methods and the scaling into multiples and submultiples, by establishment of primary standards. In some cases an approximation is used, when the realisation of the units is less precise than other methods of generating a scale of the quantity in question. This is presently the situation for the electrical units in the SI, where voltage and resistance are defined in terms of the ampere, but are used in practice from realisations based on the Josephson effect and the quantised Hall effect.
- the transfer of traceability from the primary standards to secondary and working standards. This is achieved by calibration.

Theoretically, metrology, as the science of measurement, attempts to validate the data obtained from test equipment. Though metrology is the science of measurement, in practical applications, it is the enforcement, verification and validation of predefined standards for:

Criterion	Definition
Accuracy	is the degree of exactness which the final product corresponds to the measurement standard.
Precision	refers to the ability of a measurement to be consistently reproduced.
Reliability	refers to the consistency of accurate results over consecutive measurements over time.
Traceability	refers to the ongoing validations that the measurement of the final product conforms to the original standard of measurement.

These standards can vary widely, but are often mandated by governments, agencies, and treaties such as the International Organization for Standardization, the Metre Convention, or the FDA. These agencies promulgate policies and regulations that standardize industries, countries, and streamline international trade, products, and measurements. Metrology is, at its core, an analysis of the uncertainty of individual measurements, and attempts to validate each measurement made with a given instrument, and the data obtained from it. The dissemination of traceability to consumers in society is often performed by a dedicated calibration laboratory with a recognized quality system in compliance with such standards. National laboratory accreditation schemes have been established to offer third-party assessment of such quality systems. A central requirement of these accreditations is documented traceability to national or international standards.

Some common standards include:

- ISO 17025:2005—General Requirements for Calibration Laboratories
- ISO 9000—Quality Systems Management
- ISO 14000—Environmental Management
- 21 CFR Part 210/211—FDA Regulations concerning GMP (Good Manufacturing Practices) Quality Systems
- 21 CFR Part 110—FDA Regulations concerning Food Industry GMP's.

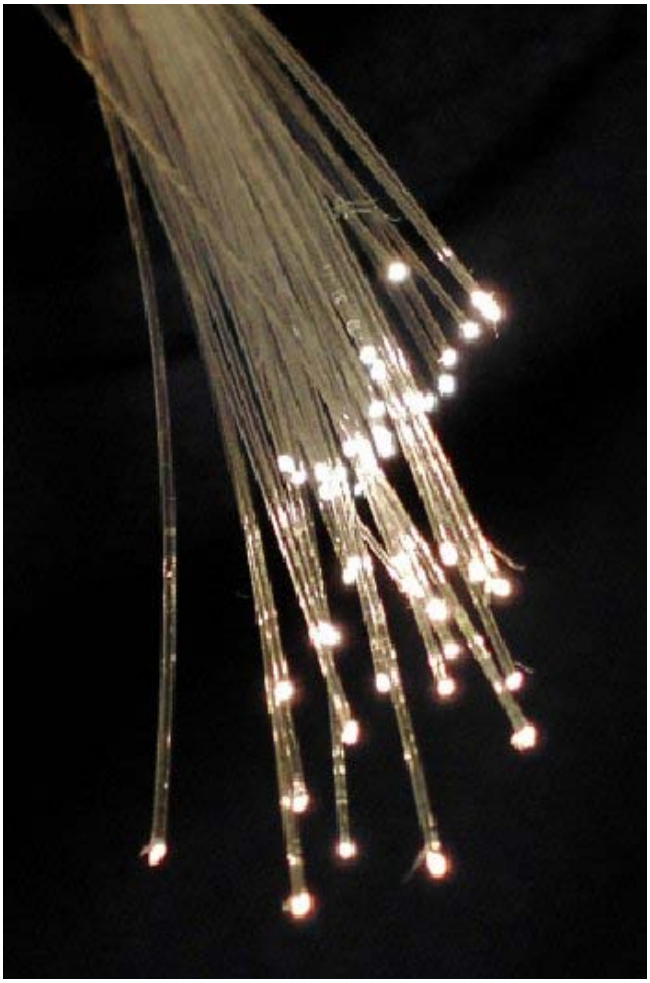
Time and frequency metrology

This area of metrology studies components and their characteristics, especially

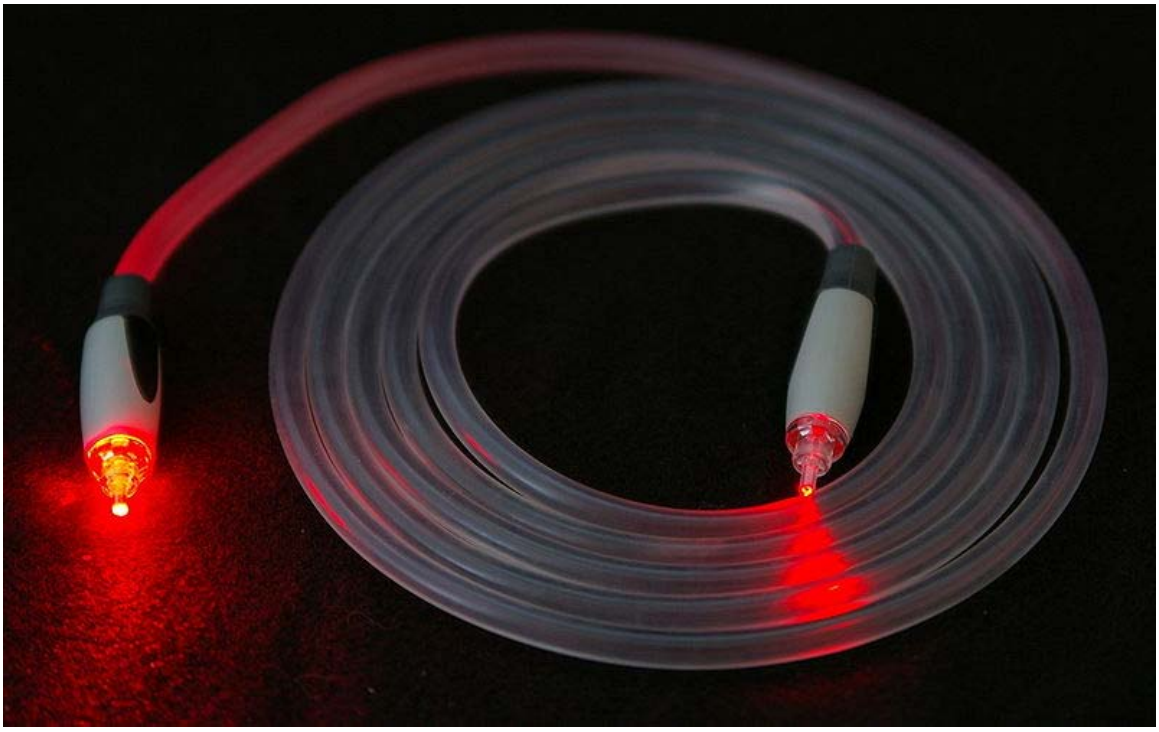
- frequency standards
- synthesizers
- oscillators
- digital clocks

Chapter 6

Optical Fiber



A bundle of optical fibers



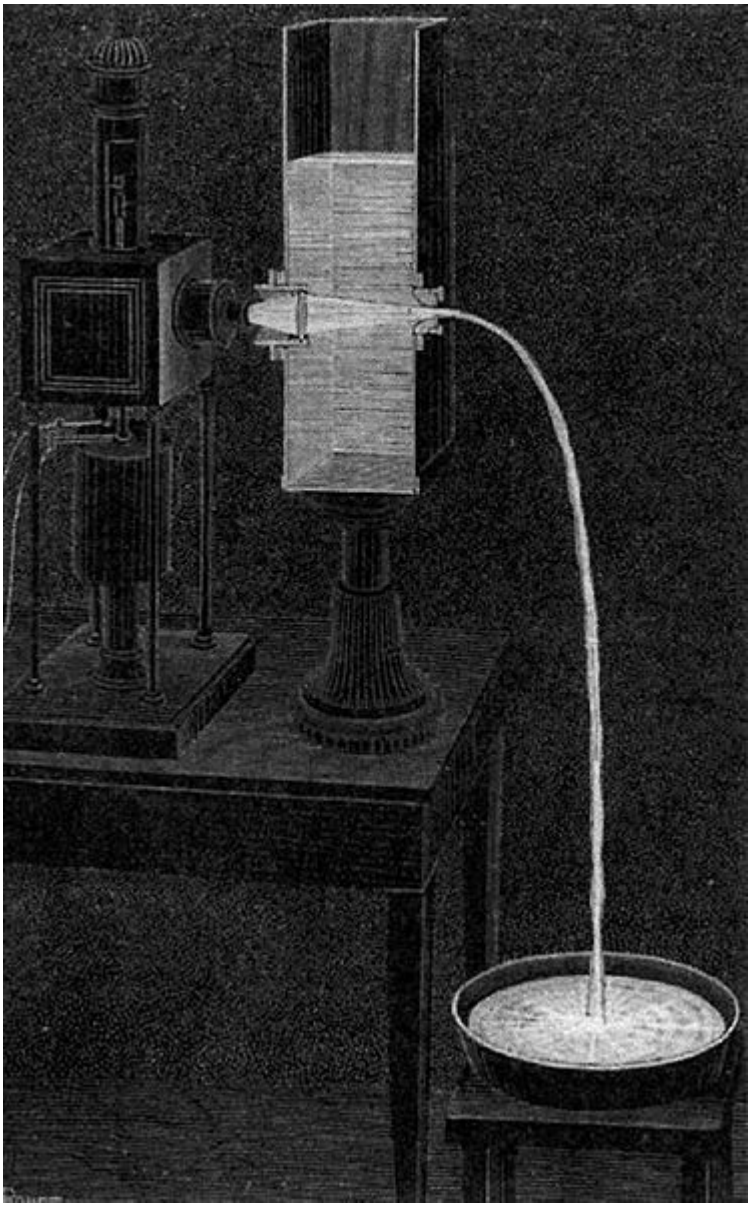
A TOSLINK fiber optic audio cable being illuminated at one end

An *optical fiber* is a thin, flexible, transparent fiber that acts as a waveguide, or "light pipe", to transmit light between the two ends of the fiber. The field of applied science and engineering concerned with the design and application of optical fibers is known as **fiber optics**. Optical fibers are widely used in fiber-optic communications, which permits transmission over longer distances and at higher bandwidths (data rates) than other forms of communication. Fibers are used instead of metal wires because signals travel along them with less loss and are also immune to electromagnetic interference. Fibers are also used for illumination, and are wrapped in bundles so they can be used to carry images, thus allowing viewing in tight spaces. Specially designed fibers are used for a variety of other applications, including sensors and fiber lasers.

Optical fiber typically consists of a transparent core surrounded by a transparent cladding material with a lower index of refraction. Light is kept in the core by total internal reflection. This causes the fiber to act as a waveguide. Fibers which support many propagation paths or transverse modes are called multi-mode fibers (MMF), while those which can only support a single mode are called single-mode fibers (SMF). Multi-mode fibers generally have a larger core diameter, and are used for short-distance communication links and for applications where high power must be transmitted. Single-mode fibers are used for most communication links longer than 1,050 meters (3,440 ft).

Joining lengths of optical fiber is more complex than joining electrical wire or cable. The ends of the fibers must be carefully cleaved, and then spliced together either mechanically or by fusing them together with heat. Special optical fiber connectors are used to make removable connections.

History



Daniel Colladon first described this "light fountain" or "light pipe" in an 1842 article titled *On the reflections of a ray of light inside a parabolic liquid stream*. This particular illustration comes from a later article by Colladon, in 1884.

Fiber optics, though used extensively in the modern world, is a fairly simple and old technology. Guiding of light by refraction, the principle that makes fiber optics possible, was first demonstrated by Daniel Colladon and Jacques Babinet in Paris in the early 1840s. John Tyndall included a demonstration of it in his public lectures in London a dozen years later. Tyndall also wrote about the property of total internal reflection in an introductory book about the nature of light in 1870: "When the light passes from air into water, the refracted ray is bent *towards* the perpendicular... When the ray passes from water to air it is bent *from* the perpendicular... If the angle which the ray in water encloses

with the perpendicular to the surface be greater than 48 degrees, the ray will not quit the water at all: it will be *totally reflected* at the surface.... The angle which marks the limit where total reflection begins is called the limiting angle of the medium. For water this angle is 48°27', for flint glass it is 38°41', while for diamond it is 23°42'." Unpigmented human hairs have also been to shown to act as an optical fibre.

Practical applications, such as close internal illumination during dentistry, appeared early in the twentieth century. Image transmission through tubes was demonstrated independently by the radio experimenter Clarence Hansell and the television pioneer John Logie Baird in the 1920s. The principle was first used for internal medical examinations by Heinrich Lamm in the following decade. In 1952, physicist Narinder Singh Kapany conducted experiments that led to the invention of optical fiber. Modern optical fibers, where the glass fiber is coated with a transparent cladding to offer a more suitable refractive index, appeared later in the decade. Development then focused on fiber bundles for image transmission. The first fiber optic semi-flexible gastroscope was patented by Basil Hirschowitz, C. Wilbur Peters, and Lawrence E. Curtiss, researchers at the University of Michigan, in 1956. In the process of developing the gastroscope, Curtiss produced the first glass-clad fibers; previous optical fibers had relied on air or impractical oils and waxes as the low-index cladding material. A variety of other image transmission applications soon followed.

In the late 19th and early 20th centuries, light was guided through bent glass rods to illuminate body cavities. Alexander Graham Bell invented a 'Photophone' to transmit voice signals over an optical beam.

Jun-ichi Nishizawa, a Japanese scientist at Tohoku University, also proposed the use of optical fibers for communications in 1963, as stated in his book published in 2004 in India. Nishizawa invented other technologies which contributed to the development of optical fiber communications, such as the graded-index optical fiber as a channel for transmitting light from semiconductor lasers. Charles K. Kao and George A. Hockham of the British company Standard Telephones and Cables (STC) were the first to promote the idea that the attenuation in optical fibers could be reduced below 20 decibels per kilometer (dB/km), allowing fibers to be a practical medium for communication. They proposed that the attenuation in fibers available at the time was caused by impurities, which could be removed, rather than fundamental physical effects such as scattering. They correctly and systematically theorized the light-loss properties for optical fiber, and pointed out the right material to manufacture such fibers — silica glass with high purity. This discovery led to Kao being awarded the Nobel Prize in Physics in 2009.

NASA used fiber optics in the television cameras sent to the moon. At the time such use in the cameras was 'classified confidential' and only those with the right security clearance or those accompanied by someone with the right security clearance were permitted to handle the cameras.

The crucial attenuation limit of 20 dB/km was first achieved in 1970, by researchers Robert D. Maurer, Donald Keck, Peter C. Schultz, and Frank Zimar working for

American glass maker Corning Glass Works, now Corning Incorporated. They demonstrated a fiber with 17 dB/km attenuation by doping silica glass with titanium. A few years later they produced a fiber with only 4 dB/km attenuation using germanium dioxide as the core dopant. Such low attenuation ushered in optical fiber telecommunication. In 1981, General Electric produced fused quartz ingots that could be drawn into fiber optic strands 25 miles (40 km) long.

Attenuation in modern optical cables is far less than in electrical copper cables, leading to long-haul fiber connections with repeater distances of 70–150 kilometers (43–93 mi). The erbium-doped fiber amplifier, which reduced the cost of long-distance fiber systems by reducing or eliminating optical-electrical-optical repeaters, was co-developed by teams led by David N. Payne of the University of Southampton and Emmanuel Desurvire at Bell Labs in 1986. Robust modern optical fiber uses glass for both core and sheath and is therefore less prone to aging processes. It was invented by Gerhard Bernsee of Schott Glass in Germany in 1973.

The emerging field of photonic crystals led to the development in 1991 of photonic-crystal fiber which guides light by diffraction from a periodic structure, rather than by total internal reflection. The first photonic crystal fibers became commercially available in 2000. Photonic crystal fibers can carry higher power than conventional fibers and their wavelength-dependent properties can be manipulated to improve performance.

Applications

Optical fiber communication

Optical fiber can be used as a medium for telecommunication and networking because it is flexible and can be bundled as cables. It is especially advantageous for long-distance communications, because light propagates through the fiber with little attenuation compared to electrical cables. This allows long distances to be spanned with few repeaters. Additionally, the per-channel light signals propagating in the fiber have been modulated at rates as high as 111 gigabits per second by NTT, although 10 or 40 Gbit/s is typical in deployed systems. Each fiber can carry many independent channels, each using a different wavelength of light (wavelength-division multiplexing (WDM)). The net data rate (data rate without overhead bytes) per fiber is the per-channel data rate reduced by the FEC overhead, multiplied by the number of channels (usually up to eighty in commercial dense WDM systems as of 2008). The current laboratory fiber optic data rate record, held by Bell Labs in Villarceaux, France, is multiplexing 155 channels, each carrying 100 Gbit/s over a 7000 km fiber. Nippon Telegraph and Telephone Corporation have also managed 69.1 Tbit/s over a single 240 km fiber (multiplexing 432 channels, equating to 171 Gbit/s per channel). Bell Labs also broke a 100 Petabit per second *kilometer* barrier (15.5 Tbit/s over a single 7000 km fiber).

For short distance applications, such as creating a network within an office building, fiber-optic cabling can be used to save space in cable ducts. This is because a single fiber can often carry much more data than many electrical cables, such as 4 pair Cat-5 Ethernet

cabling. Fiber is also immune to electrical interference; there is no cross-talk between signals in different cables and no pickup of environmental noise. Non-armored fiber cables do not conduct electricity, which makes fiber a good solution for protecting communications equipment located in high voltage environments such as power generation facilities, or metal communication structures prone to lightning strikes. They can also be used in environments where explosive fumes are present, without danger of ignition. Wiretapping is more difficult compared to electrical connections, and there are concentric dual core fibers that are said to be tap-proof.

Fiber optic sensors

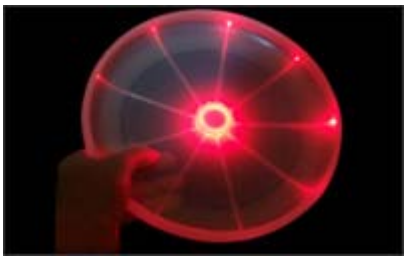
Fibers have many uses in remote sensing. In some applications, the sensor is itself an optical fiber. In other cases, fiber is used to connect a non-fiberoptic sensor to a measurement system. Depending on the application, fiber may be used because of its small size, or the fact that no electrical power is needed at the remote location, or because many sensors can be multiplexed along the length of a fiber by using different wavelengths of light for each sensor, or by sensing the time delay as light passes along the fiber through each sensor. Time delay can be determined using a device such as an optical time-domain reflectometer.

Optical fibers can be used as sensors to measure strain, temperature, pressure and other quantities by modifying a fiber so that the quantity to be measured modulates the intensity, phase, polarization, wavelength or transit time of light in the fiber. Sensors that vary the intensity of light are the simplest, since only a simple source and detector are required. A particularly useful feature of such fiber optic sensors is that they can, if required, provide distributed sensing over distances of up to one meter.

Extrinsic fiber optic sensors use an optical fiber cable, normally a multi-mode one, to transmit modulated light from either a non-fiber optical sensor, or an electronic sensor connected to an optical transmitter. A major benefit of extrinsic sensors is their ability to reach places which are otherwise inaccessible. An example is the measurement of temperature inside aircraft jet engines by using a fiber to transmit radiation into a radiation pyrometer located outside the engine. Extrinsic sensors can also be used in the same way to measure the internal temperature of electrical transformers, where the extreme electromagnetic fields present make other measurement techniques impossible. Extrinsic sensors are used to measure vibration, rotation, displacement, velocity, acceleration, torque, and twisting. A solid state version of the gyroscope using the interference of light has been developed. The fiber optic gyroscope (FOG) has no moving parts and exploits the Sagnac effect to detect mechanical rotation.

A common use for fiber optic sensors are in advanced intrusion detection security systems, where the light is transmitted along the fiber optic sensor cable, which is placed on a fence, pipeline or communication cabling, and the returned signal is monitored and analysed for disturbances. This return signal is digitally processed to identify if there is a disturbance, and if an intrusion has occurred an alarm is triggered by the fiber optic security system.

Other uses of optical fibers



A frisbee illuminated by fiber optics

Light reflected from optical fiber illuminates exhibited model

Fiber optic front sight on a hand gun

Fibers are widely used in illumination applications. They are used as light guides in medical and other applications where bright light needs to be shone on a target without a clear line-of-sight path. In some buildings, optical fibers are used to route sunlight from the roof to other parts of the building. Optical fiber illumination is also used for decorative applications, including signs, art, and artificial Christmas trees. Swarovski boutiques use optical fibers to illuminate their crystal showcases from many different angles while only employing one light source. Optical fiber is an intrinsic part of the light-transmitting concrete building product, LiTraCon.

Optical fiber is also used in imaging optics. A coherent bundle of fibers is used, sometimes along with lenses, for a long, thin imaging device called an endoscope, which is used to view objects through a small hole. Medical endoscopes are used for minimally invasive exploratory or surgical procedures (endoscopy). Industrial endoscopes are used for inspecting anything hard to reach, such as jet engine interiors.

In spectroscopy, optical fiber bundles are used to transmit light from a spectrometer to a substance which cannot be placed inside the spectrometer itself, in order to analyze its composition. A spectrometer analyzes substances by bouncing light off of and through them. By using fibers, a spectrometer can be used to study objects that are too large to fit inside, or gasses, or reactions which occur in pressure vessels.

An optical fiber doped with certain rare earth elements such as erbium can be used as the gain medium of a laser or optical amplifier. Rare-earth doped optical fibers can be used to provide signal amplification by splicing a short section of doped fiber into a regular (undoped) optical fiber line. The doped fiber is optically pumped with a second laser wavelength that is coupled into the line in addition to the signal wave. Both wavelengths of light are transmitted through the doped fiber, which transfers energy from the second pump wavelength to the signal wave. The process that causes the amplification is stimulated emission.

Optical fibers doped with a wavelength shifter are used to collect scintillation light in physics experiments.

Optical fiber can be used to supply a low level of power (around one watt) to electronics situated in a difficult electrical environment. Examples of this are electronics in high-powered antenna elements and measurement devices used in high voltage transmission equipment.

A growing trend in iron sights for arms, is the use of short pieces of optical fiber for contrast enhancement dots, made in such a way that ambient light falling on the length of the fiber is concentrated at the tip, making the dots slightly brighter than the surroundings. This method is most commonly used in front sights, but many makers offer sights that use fiber optics on front and rear sights. Fiber optic sights can now be found on handguns, rifles, and shotguns, both as aftermarket accessories and a growing number of factory guns.

Principle of operation

An optical fiber is a cylindrical dielectric waveguide (nonconducting waveguide) that transmits light along its axis, by the process of total internal reflection. The fiber consists of a *core* surrounded by a cladding layer, both of which are made of dielectric materials. To confine the optical signal in the core, the refractive index of the core must be greater than that of the cladding. The boundary between the core and cladding may either be abrupt, in *step-index fiber*, or gradual, in *graded-index fiber*.

Index of refraction

The index of refraction is a way of measuring the speed of light in a material. Light travels fastest in a vacuum, such as outer space. The speed of light in a vacuum is about 300,000 kilometres (186 thousand miles) per second. Index of refraction is calculated by dividing the speed of light in a vacuum by the speed of light in some other medium. The index of refraction of a vacuum is therefore 1, by definition. The typical value for the cladding of an optical fiber is 1.46. The core value is typically 1.48. The larger the index of refraction, the slower light travels in that medium. From this information, a good rule of thumb is that signal using optical fiber for communication will travel at around 200 million meters per second. Or to put it another way, to travel 1000 kilometers in fiber, the signal will take 5 milliseconds to propagate. Thus a phone call carried by fiber between

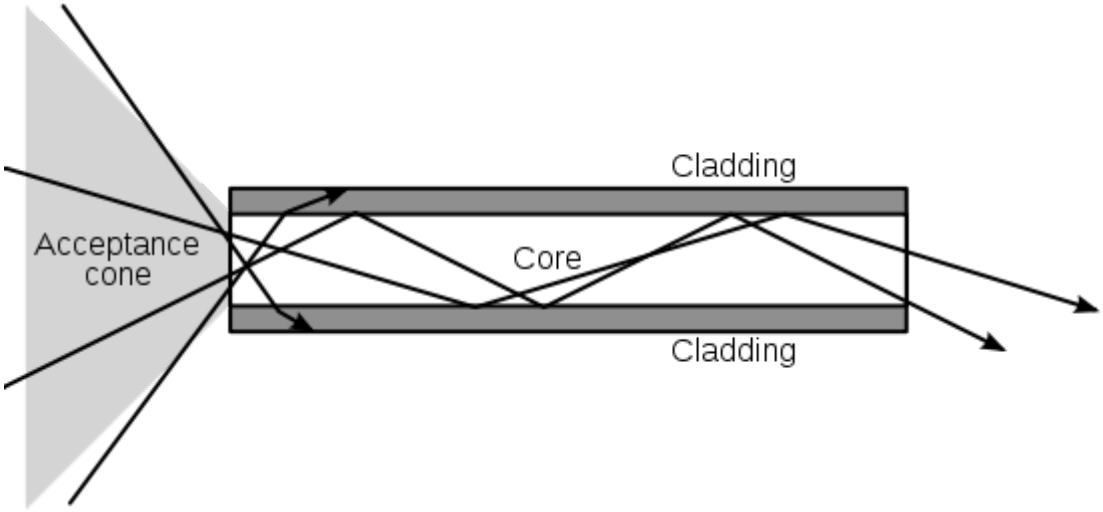
Sydney and New York, a 12000 kilometer distance, means that there is an absolute minimum delay of 60 milliseconds (or around 1/16 of a second) between when one caller speaks to when the other hears. (Of course the fiber in this case will probably travel a longer route, and there will be additional delays due to communication equipment switching and the process of encoding and decoding the voice onto the fiber).

Total internal reflection

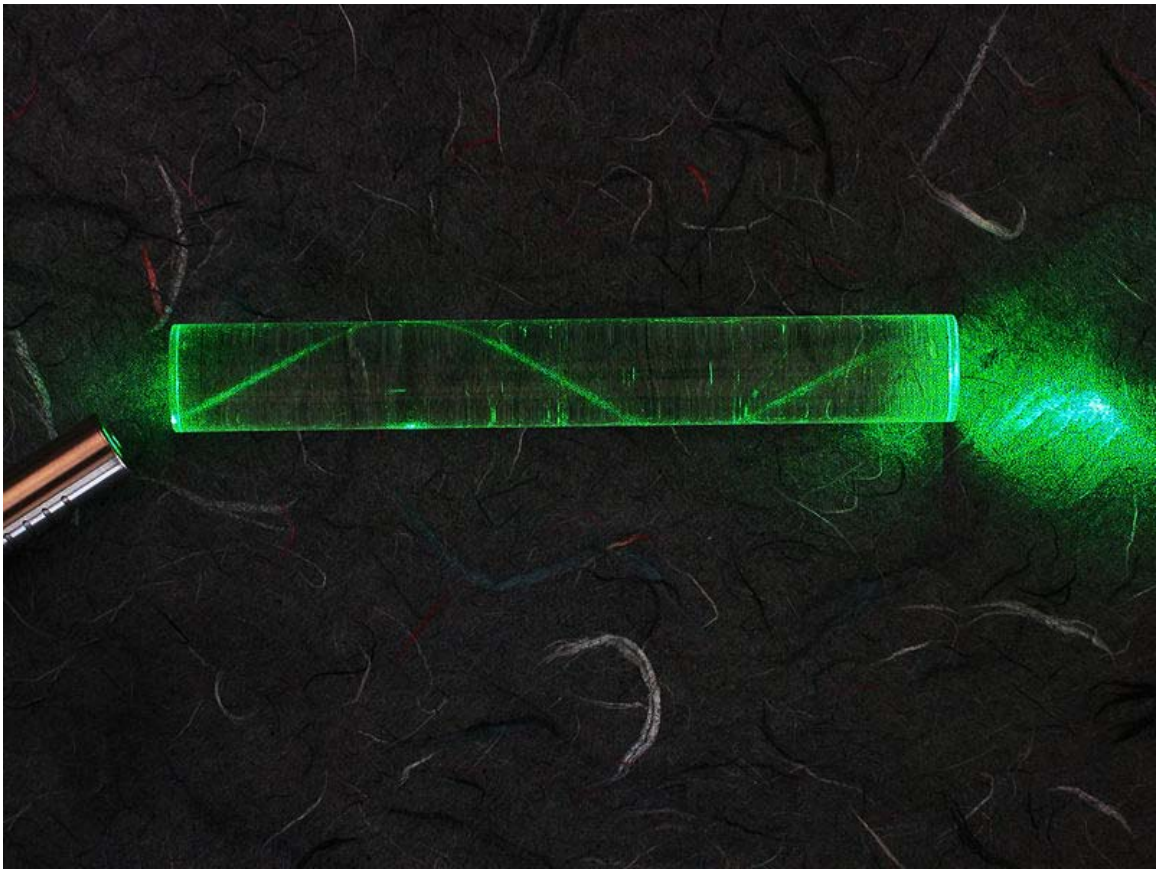
When light traveling in a dense medium hits a boundary at a steep angle (larger than the "critical angle" for the boundary), the light will be completely reflected. This effect is used in optical fibers to confine light in the core. Light travels along the fiber bouncing back and forth off of the boundary. Because the light must strike the boundary with an angle greater than the critical angle, only light that enters the fiber within a certain range of angles can travel down the fiber without leaking out. This range of angles is called the acceptance cone of the fiber. The size of this acceptance cone is a function of the refractive index difference between the fiber's core and cladding.

In simpler terms, there is a maximum angle from the fiber axis at which light may enter the fiber so that it will propagate, or travel, in the core of the fiber. The sine of this maximum angle is the numerical aperture (NA) of the fiber. Fiber with a larger NA requires less precision to splice and work with than fiber with a smaller NA. Single-mode fiber has a small NA.

Multi-mode fiber

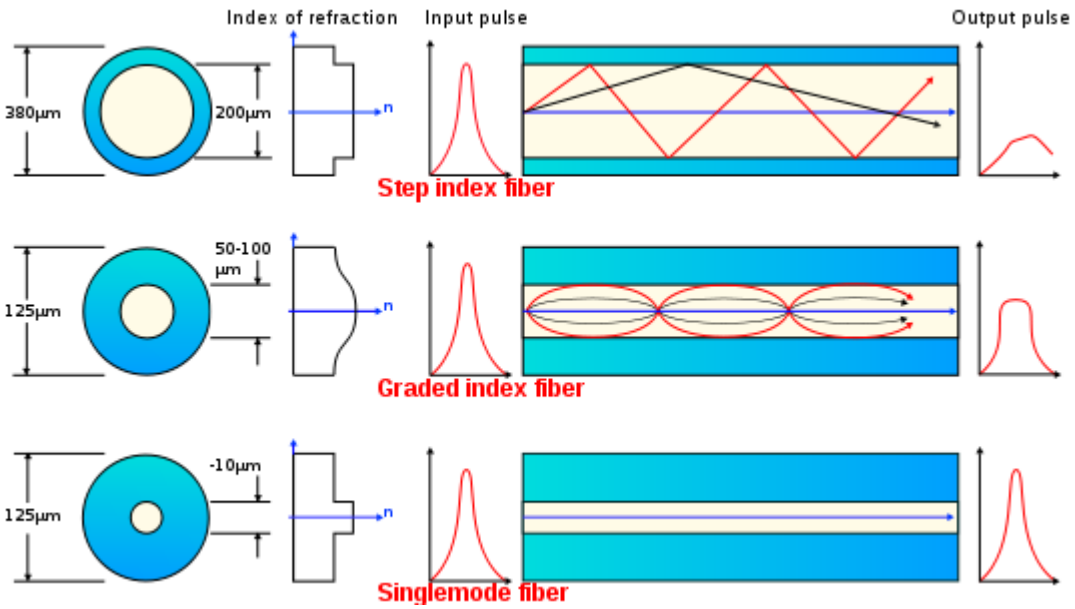


The propagation of light through a multi-mode optical fiber.



A laser bouncing down an acrylic rod, illustrating the total internal reflection of light in a multi-mode optical fiber.

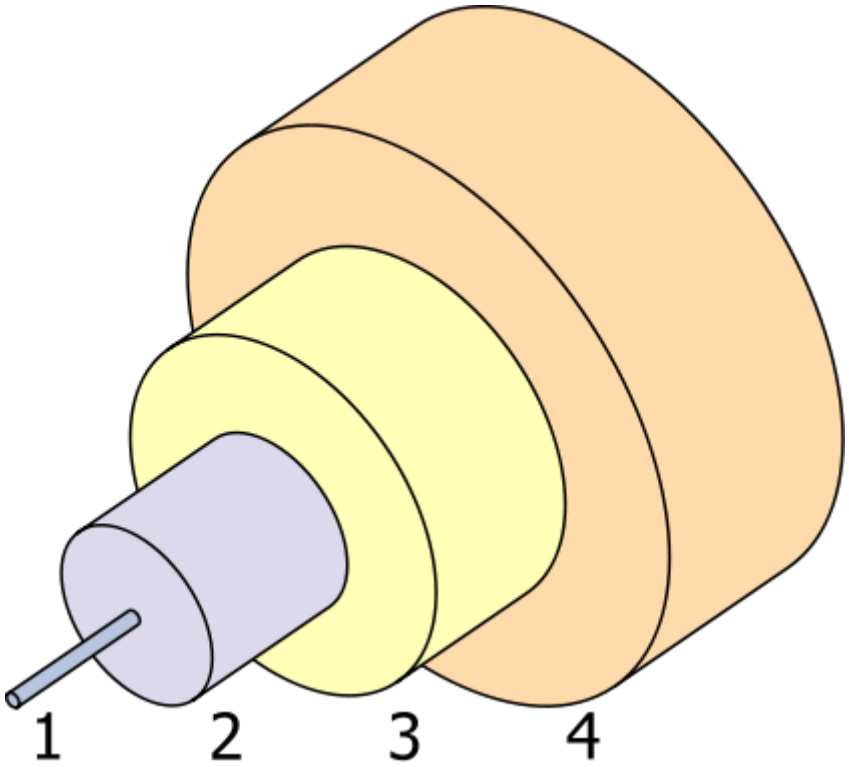
Fiber with large core diameter (greater than 10 micrometers) may be analyzed by geometrical optics. Such fiber is called *multi-mode fiber*, from the electromagnetic analysis (see below). In a step-index multi-mode fiber, rays of light are guided along the fiber core by total internal reflection. Rays that meet the core-cladding boundary at a high angle (measured relative to a line normal to the boundary), greater than the critical angle for this boundary, are completely reflected. The critical angle (minimum angle for total internal reflection) is determined by the difference in index of refraction between the core and cladding materials. Rays that meet the boundary at a low angle are refracted from the core into the cladding, and do not convey light and hence information along the fiber. The critical angle determines the acceptance angle of the fiber, often reported as a numerical aperture. A high numerical aperture allows light to propagate down the fiber in rays both close to the axis and at various angles, allowing efficient coupling of light into the fiber. However, this high numerical aperture increases the amount of dispersion as rays at different angles have different path lengths and therefore take different times to traverse the fiber.



Optical fiber types.

In graded-index fiber, the index of refraction in the core decreases continuously between the axis and the cladding. This causes light rays to bend smoothly as they approach the cladding, rather than reflecting abruptly from the core-cladding boundary. The resulting curved paths reduce multi-path dispersion because high angle rays pass more through the lower-index periphery of the core, rather than the high-index center. The index profile is chosen to minimize the difference in axial propagation speeds of the various rays in the fiber. This ideal index profile is very close to a parabolic relationship between the index and the distance from the axis.

Single-mode fiber



The structure of a typical single-mode fiber.
1. Core: 8 µm diameter
2. Cladding: 125 µm dia.
3. Buffer: 250 µm dia.
4. Jacket: 400 µm dia.

Fiber with a core diameter less than about ten times the wavelength of the propagating light cannot be modeled using geometric optics. Instead, it must be analyzed as an electromagnetic structure, by solution of Maxwell's equations as reduced to the electromagnetic wave equation. The electromagnetic analysis may also be required to understand behaviors such as speckle that occur when coherent light propagates in multi-mode fiber. As an optical waveguide, the fiber supports one or more confined transverse modes by which light can propagate along the fiber. Fiber supporting only one mode is called *single-mode* or *mono-mode fiber*. The behavior of larger-core multi-mode fiber can also be modeled using the wave equation, which shows that such fiber supports more than one mode of propagation (hence the name). The results of such modeling of multi-mode fiber approximately agree with the predictions of geometric optics, if the fiber core is large enough to support more than a few modes.

The waveguide analysis shows that the light energy in the fiber is not completely confined in the core. Instead, especially in single-mode fibers, a significant fraction of the energy in the bound mode travels in the cladding as an evanescent wave.

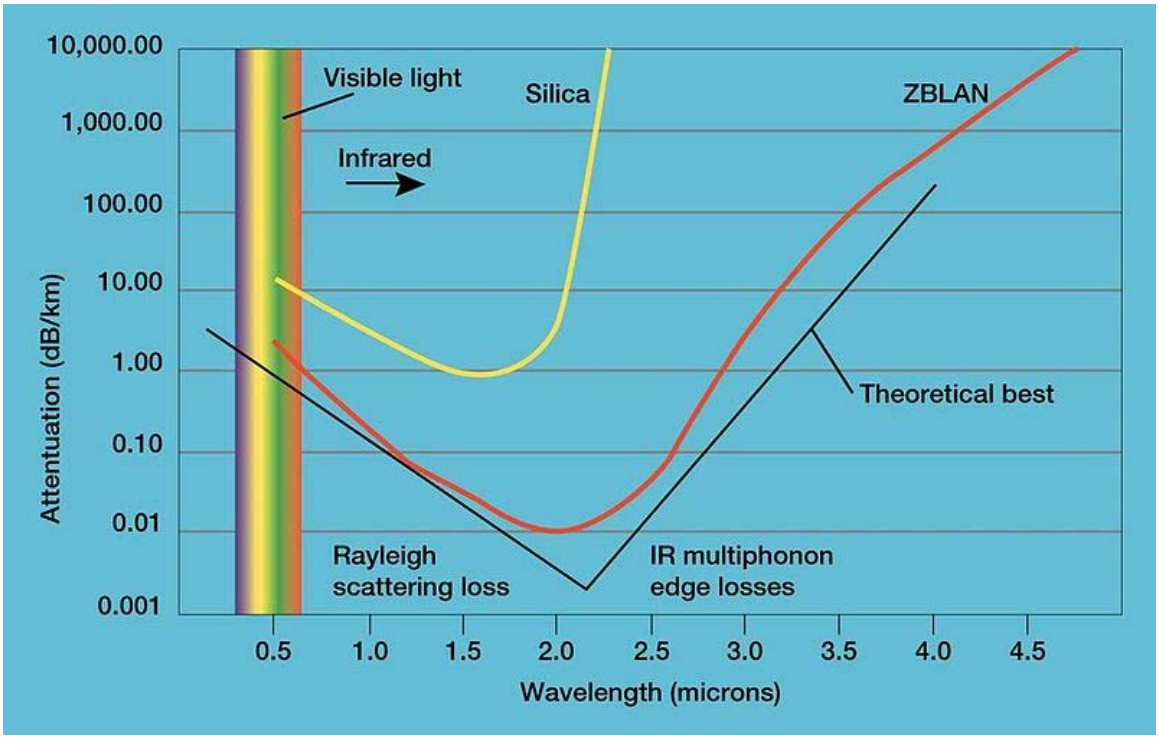
The most common type of single-mode fiber has a core diameter of 8–10 micrometers and is designed for use in the near infrared. The mode structure depends on the wavelength of the light used, so that this fiber actually supports a small number of additional modes at visible wavelengths. Multi-mode fiber, by comparison, is manufactured with core diameters as small as 50 micrometers and as large as hundreds of micrometers. The normalized frequency V for this fiber should be less than the first zero of the Bessel function J_0 (approximately 2.405).

Special-purpose fiber

Some special-purpose optical fiber is constructed with a non-cylindrical core and/or cladding layer, usually with an elliptical or rectangular cross-section. These include polarization-maintaining fiber and fiber designed to suppress whispering gallery mode propagation.

Photonic-crystal fiber is made with a regular pattern of index variation (often in the form of cylindrical holes that run along the length of the fiber). Such fiber uses diffraction effects instead of or in addition to total internal reflection, to confine light to the fiber's core. The properties of the fiber can be tailored to a wide variety of applications.

Mechanisms of attenuation

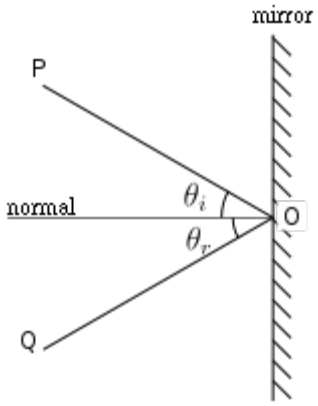


Light attenuation by ZBLAN and silica fibers

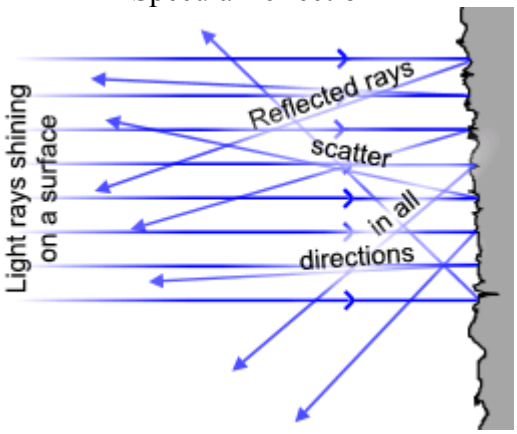
Attenuation in fiber optics, also known as transmission loss, is the reduction in intensity of the light beam (or signal) with respect to distance traveled through a transmission

medium. Attenuation coefficients in fiber optics usually use units of dB/km through the medium due to the relatively high quality of transparency of modern optical transmission media. The medium is usually a fiber of silica glass that confines the incident light beam to the inside. Attenuation is an important factor limiting the transmission of a digital signal across large distances. Thus, much research has gone into both limiting the attenuation and maximizing the amplification of the optical signal. Empirical research has shown that attenuation in optical fiber is caused primarily by both scattering and absorption.

Light scattering



Specular reflection



Diffuse reflection

The propagation of light through the core of an optical fiber is based on total internal reflection of the lightwave. Rough and irregular surfaces, even at the molecular level, can cause light rays to be reflected in random directions. This is called diffuse reflection or scattering, and it is typically characterized by wide variety of reflection angles.

Light scattering depends on the wavelength of the light being scattered. Thus, limits to spatial scales of visibility arise, depending on the frequency of the incident light-wave and the physical dimension (or spatial scale) of the scattering center, which is typically in the form of some specific micro-structural feature. Since visible light has a wavelength of

the order of one micrometre (one millionth of a meter) scattering centers will have dimensions on a similar spatial scale.

Thus, attenuation results from the incoherent scattering of light at internal surfaces and interfaces. In (poly)crystalline materials such as metals and ceramics, in addition to pores, most of the internal surfaces or interfaces are in the form of grain boundaries that separate tiny regions of crystalline order. It has recently been shown that when the size of the scattering center (or grain boundary) is reduced below the size of the wavelength of the light being scattered, the scattering no longer occurs to any significant extent. This phenomenon has given rise to the production of transparent ceramic materials.

Similarly, the scattering of light in optical quality glass fiber is caused by molecular level irregularities (compositional fluctuations) in the glass structure. Indeed, one emerging school of thought is that a glass is simply the limiting case of a polycrystalline solid. Within this framework, "domains" exhibiting various degrees of short-range order become the building blocks of both metals and alloys, as well as glasses and ceramics. Distributed both between and within these domains are micro-structural defects which will provide the most ideal locations for the occurrence of light scattering. This same phenomenon is seen as one of the limiting factors in the transparency of IR missile domes.

At high optical powers, scattering can also be caused by nonlinear optical processes in the fiber.

UV-Vis-IR absorption

In addition to light scattering, attenuation or signal loss can also occur due to selective absorption of specific wavelengths, in a manner similar to that responsible for the appearance of color. Primary material considerations include both electrons and molecules as follows:

1) At the electronic level, it depends on whether the electron orbitals are spaced (or "quantized") such that they can absorb a quantum of light (or photon) of a specific wavelength or frequency in the ultraviolet (UV) or visible ranges. This is what gives rise to color.

2) At the atomic or molecular level, it depends on the frequencies of atomic or molecular vibrations or chemical bonds, how close-packed its atoms or molecules are, and whether or not the atoms or molecules exhibit long-range order. These factors will determine the capacity of the material transmitting longer wavelengths in the infrared (IR), far IR, radio and microwave ranges.

The design of any optically transparent device requires the selection of materials based upon knowledge of its properties and limitations. The lattice absorption characteristics observed at the lower frequency regions (mid IR to far-infrared wavelength range) define the long-wavelength transparency limit of the material. They are the result of the

interactive coupling between the motions of thermally induced vibrations of the constituent atoms and molecules of the solid lattice and the incident light wave radiation. Hence, all materials are bounded by limiting regions of absorption caused by atomic and molecular vibrations (bond-stretching)in the far-infrared (>10 μm).

Thus, multi-phonon absorption occurs when two or more phonons simultaneously interact to produce electric dipole moments with which the incident radiation may couple. These dipoles can absorb energy from the incident radiation, reaching a maximum coupling with the radiation when the frequency is equal to the fundamental vibrational mode of the molecular dipole (e.g. Si-O bond) in the far-infrared, or one of its harmonics.

The selective absorption of infrared (IR) light by a particular material occurs because the selected frequency of the light wave matches the frequency (or an integer multiple of the frequency) at which the particles of that material vibrate. Since different atoms and molecules have different natural frequencies of vibration, they will selectively absorb different frequencies (or portions of the spectrum) of infrared (IR) light.

Reflection and transmission of light waves occur because the frequencies of the light waves do not match the natural resonant frequencies of vibration of the objects. When IR light of these frequencies strikes an object, the energy is either reflected or transmitted.

Manufacturing

Materials

Glass optical fibers are almost always made from silica, but some other materials, such as fluorozirconate, fluoroaluminate, and chalcogenide glasses as well as crystalline materials like sapphire, are used for longer-wavelength infrared or other specialized applications. Silica and fluoride glasses usually have refractive indices of about 1.5, but some materials such as the chalcogenides can have indices as high as 3. Typically the index difference between core and cladding is less than one percent.

Plastic optical fibers (POF) are commonly step-index multi-mode fibers with a core diameter of 0.5 millimeters or larger. POF typically have higher attenuation coefficients than glass fibers, 1 dB/m or higher, and this high attenuation limits the range of POF-based systems.

Silica

Silica exhibits fairly good optical transmission over a wide range of wavelengths. In the near-infrared (near IR) portion of the spectrum, particularly around 1.5 μm, silica can have extremely low absorption and scattering losses of the order of 0.2 dB/km. A high transparency in the 1.4-μm region is achieved by maintaining a low concentration of hydroxyl groups (OH). Alternatively, a high OH concentration is better for transmission in the ultraviolet (UV) region.

Silica can be drawn into fibers at reasonably high temperatures, and has a fairly broad glass transformation range. One other advantage is that fusion splicing and cleaving of silica fibers is relatively effective. Silica fiber also has high mechanical strength against both pulling and even bending, provided that the fiber is not too thick and that the surfaces have been well prepared during processing. Even simple cleaving (breaking) of the ends of the fiber can provide nicely flat surfaces with acceptable optical quality. Silica is also relatively chemically inert. In particular, it is not hygroscopic (does not absorb water).

Silica glass can be doped with various materials. One purpose of doping is to raise the refractive index (e.g. with Germanium dioxide (GeO_2) or Aluminium oxide (Al_2O_3)) or to lower it (e.g. with fluorine or Boron trioxide (B_2O_3)). Doping is also possible with laser-active ions (for example, rare earth-doped fibers) in order to obtain active fibers to be used, for example, in fiber amplifiers or laser applications. Both the fiber core and cladding are typically doped, so that the entire assembly (core and cladding) is effectively the same compound (e.g. an aluminosilicate, germanosilicate, phosphosilicate or borosilicate glass).

Particularly for active fibers, pure silica is usually not a very suitable host glass, because it exhibits a low solubility for rare earth ions. This can lead to quenching effects due to clustering of dopant ions. Aluminosilicates are much more effective in this respect.

Silica fiber also exhibits a high threshold for optical damage. This property ensures a low tendency for laser-induced breakdown. This is important for fiber amplifiers when utilized for the amplification of short pulses.

Because of these properties silica fibers are the material of choice in many optical applications, such as communications (except for very short distances with plastic optical fiber), fiber lasers, fiber amplifiers, and fiber-optic sensors. The large efforts which have been put forth in the development of various types of silica fibers have further increased the performance of such fibers over other materials.

Fluorides

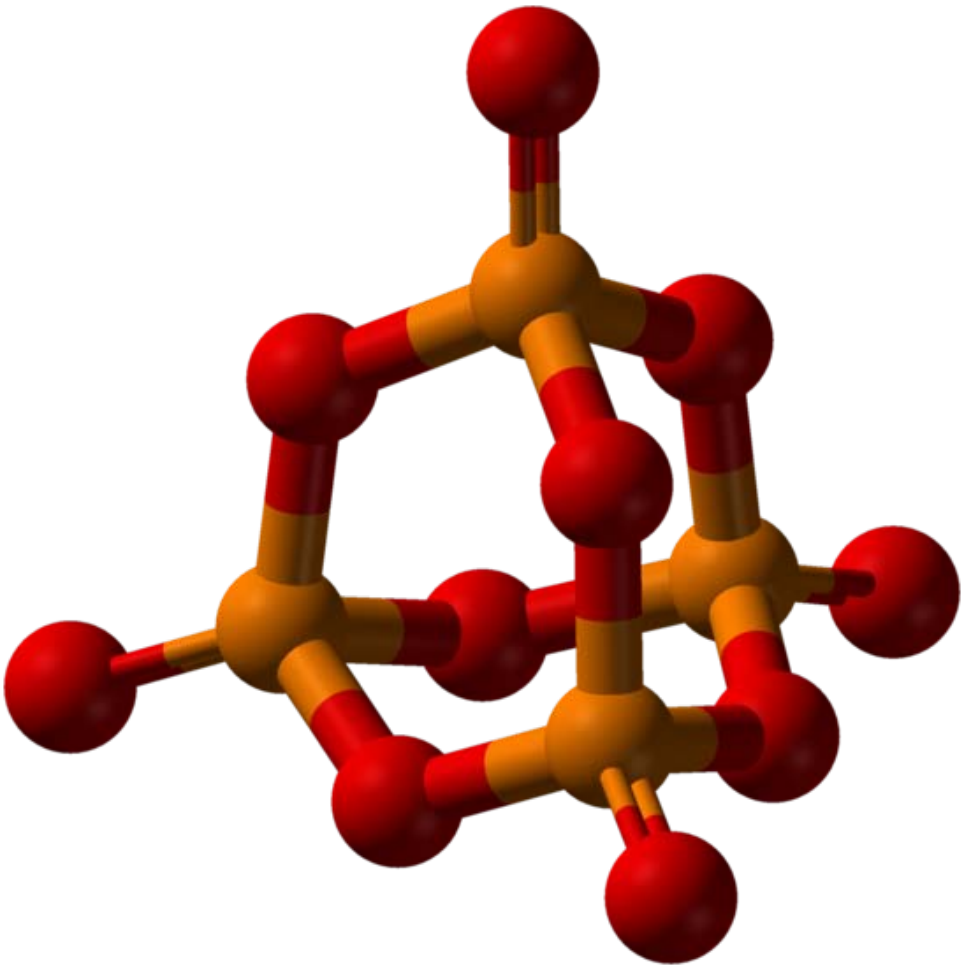
Fluoride glass is a class of non-oxide optical quality glasses composed of fluorides of various metals. Because of their low viscosity, it is very difficult to completely avoid crystallization while processing it through the glass transition (or drawing the fiber from the melt). Thus, although heavy metal fluoride glasses (HMFG) exhibit very low optical attenuation, they are not only difficult to manufacture, but are quite fragile, and have poor resistance to moisture and other environmental attacks. Their best attribute is that they lack the absorption band associated with the hydroxyl (OH) group (3200–3600 cm^{-1}), which is present in nearly all oxide-based glasses.

An example of a heavy metal fluoride glass is the ZBLAN glass group, composed of zirconium, barium, lanthanum, aluminium, and sodium fluorides. Their main

technological application is as optical waveguides in both planar and fiber form. They are advantageous especially in the mid-infrared (2000–5000 nm) range.

HMFGs were initially slated for optical fiber applications, because the intrinsic losses of a mid-IR fiber could in principle be lower than those of silica fibers, which are transparent only up to about 2 μm. However, such low losses were never realized in practice, and the fragility and high cost of fluoride fibers made them less than ideal as primary candidates. Later, the utility of fluoride fibers for various other applications was discovered. These include mid-IR spectroscopy, fiber optic sensors, thermometry, and imaging. Also, fluoride fibers can be used for guided lightwave transmission in media such as YAG (yttria-alumina garnet) lasers at 2.9 μm, as required for medical applications (e.g. ophthalmology and dentistry).

Phosphates



The P_4O_{10} cagelike structure—the basic building block for phosphate glass.

Phosphate glass constitutes a class of optical glasses composed of metaphosphates of various metals. Instead of the SiO_4 tetrahedra observed in silicate glasses, the building block for this glass former is Phosphorus pentoxide (P_2O_5), which crystallizes in at least four different forms. The most familiar polymorph (see figure) comprises molecules of P_4O_{10}.

Phosphate glasses can be advantageous over silica glasses for optical fibers with a high concentration of doping rare earth ions. A mix of fluoride glass and phosphate glass is fluorophosphate glass.

Chalcogenides

The chalcogens—the elements in group 16 of the periodic table—particularly sulfur (S), selenium (Se) and tellurium (Te)—react with more electropositive elements, such as silver, to form chalcogenides. These are extremely versatile compounds, in that they can be crystalline or amorphous, metallic or semiconducting, and conductors of ions or electrons.

Process

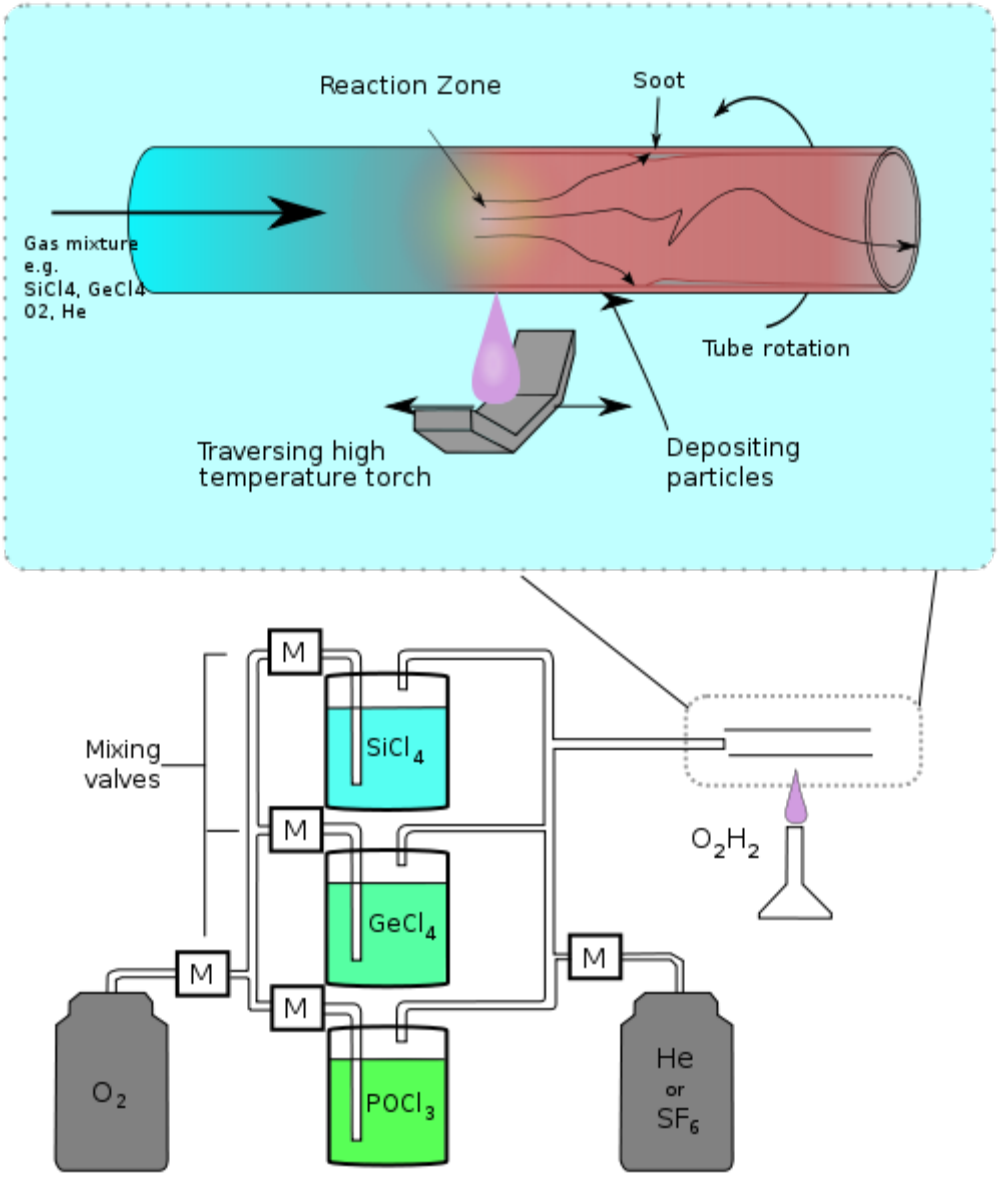


Illustration of the modified chemical vapor deposition (inside) process

Standard optical fibers are made by first constructing a large-diameter "preform", with a carefully controlled refractive index profile, and then "pulling" the preform to form the long, thin optical fiber. The preform is commonly made by three chemical vapor deposition methods: *inside vapor deposition*, *outside vapor deposition*, and *vapor axial deposition*.

With *inside vapor deposition*, the preform starts as a hollow glass tube approximately 40 centimeters (16 in) long, which is placed horizontally and rotated slowly on a lathe. Gases such as silicon tetrachloride ($SiCl_4$) or germanium tetrachloride ($GeCl_4$) are injected with oxygen in the end of the tube. The gases are then heated by means of an

external hydrogen burner, bringing the temperature of the gas up to 1900 K (1600 °C, 3000 °F), where the tetrachlorides react with oxygen to produce silica or germania (germanium dioxide) particles. When the reaction conditions are chosen to allow this reaction to occur in the gas phase throughout the tube volume, in contrast to earlier techniques where the reaction occurred only on the glass surface, this technique is called *modified chemical vapor deposition (MCVD)*.

The oxide particles then agglomerate to form large particle chains, which subsequently deposit on the walls of the tube as soot. The deposition is due to the large difference in temperature between the gas core and the wall causing the gas to push the particles outwards (this is known as thermophoresis). The torch is then traversed up and down the length of the tube to deposit the material evenly. After the torch has reached the end of the tube, it is then brought back to the beginning of the tube and the deposited particles are then melted to form a solid layer. This process is repeated until a sufficient amount of material has been deposited. For each layer the composition can be modified by varying the gas composition, resulting in precise control of the finished fiber's optical properties.

In outside vapor deposition or vapor axial deposition, the glass is formed by *flame hydrolysis*, a reaction in which silicon tetrachloride and germanium tetrachloride are oxidized by reaction with water (H_2O) in an oxyhydrogen flame. In outside vapor deposition the glass is deposited onto a solid rod, which is removed before further processing. In vapor axial deposition, a short *seed rod* is used, and a porous preform, whose length is not limited by the size of the source rod, is built up on its end. The porous preform is consolidated into a transparent, solid preform by heating to about 1800 K (1500 °C, 2800 °F).

The preform, however constructed, is then placed in a device known as a drawing tower, where the preform tip is heated and the optic fiber is pulled out as a string. By measuring the resultant fiber width, the tension on the fiber can be controlled to maintain the fiber thickness.

Coatings

The light is "guided" down the core of the fiber by an optical "cladding" with a lower refractive index that traps light in the core through "total internal reflection."

The cladding is coated by a "buffer" that protects it from moisture and physical damage. The buffer is what gets stripped off the fiber for termination or splicing. These coatings are UV-cured urethane acrylate composite materials applied to the outside of the fiber during the drawing process. The coatings protect the very delicate strands of glass fiber—about the size of a human hair—and allow it to survive the rigors of manufacturing, proof testing, cabling and installation.

Today’s glass optical fiber draw processes employ a dual-layer coating approach. An inner primary coating is designed to act as a shock absorber to minimize attenuation caused by microbending. An outer secondary coating protects the primary coating against

mechanical damage and acts as a barrier to lateral forces. Sometimes a metallic armour layer is added to provide extra protection.

These fiber optic coating layers are applied during the fiber draw, at speeds approaching 100 kilometers per hour (60 mph). Fiber optic coatings are applied using one of two methods: wet-on-dry, in which the fiber passes through a primary coating application, which is then UV cured, then through the secondary coating application which is subsequently cured; and wet-on-wet, in which the fiber passes through both the primary and secondary coating applications and then goes to UV curing.

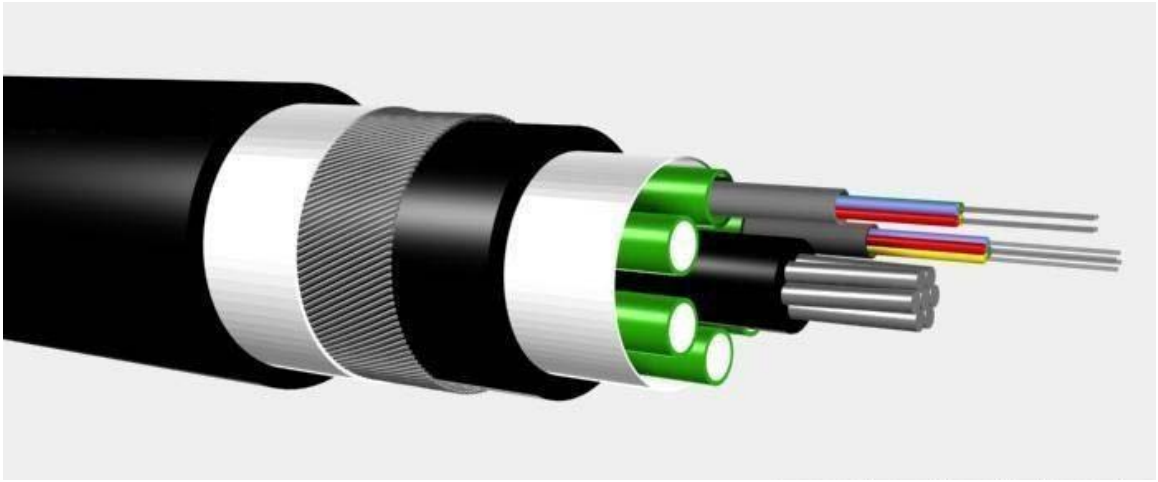
Fiber optic coatings are applied in concentric layers to prevent damage to the fiber during the drawing application and to maximize fiber strength and microbend resistance. Unevenly coated fiber will experience non-uniform forces when the coating expands or contracts, and is susceptible to greater signal attenuation. Under proper drawing and coating processes, the coatings are concentric around the fiber, continuous over the length of the application and have constant thickness.

Fiber optic coatings protect the glass fibers from scratches that could lead to strength degradation. The combination of moisture and scratches accelerates the aging and deterioration of fiber strength. When fiber is subjected to low stresses over a long period, fiber fatigue can occur. Over time or in extreme conditions, these factors combine to cause microscopic flaws in the glass fiber to propagate, which can ultimately result in fiber failure.

Three key characteristics of fiber optic waveguides can be affected by environmental conditions: strength, attenuation and resistance to losses caused by microbending. External fiber optic coatings protect glass optical fiber from environmental conditions that can affect the fiber’s performance and long-term durability. On the inside, coatings ensure the reliability of the signal being carried and help minimize attenuation due to microbending.

Practical issues

Optical fiber cables



An optical fiber cable

In practical fibers, the cladding is usually coated with a tough resin *buffer* layer, which may be further surrounded by a *jacket* layer, usually glass. These layers add strength to the fiber but do not contribute to its optical wave guide properties. Rigid fiber assemblies sometimes put light-absorbing ("dark") glass between the fibers, to prevent light that leaks out of one fiber from entering another. This reduces cross-talk between the fibers, or reduces flare in fiber bundle imaging applications.

Modern cables come in a wide variety of sheathings and armor, designed for applications such as direct burial in trenches, high voltage isolation, dual use as power lines, installation in conduit, lashing to aerial telephone poles, submarine installation, and insertion in paved streets. The cost of small fiber-count pole-mounted cables has greatly decreased due to the high demand for fiber to the home (FTTH) installations in Japan and South Korea.

Fiber cable can be very flexible, but traditional fiber's loss increases greatly if the fiber is bent with a radius smaller than around 30 mm. This creates a problem when the cable is bent around corners or wound around a spool, making FTTX installations more complicated. "Bendable fibers", targeted towards easier installation in home environments, have been standardized as ITU-T G.657. This type of fiber can be bent with a radius as low as 7.5 mm without adverse impact. Even more bendable fibers have been developed. Bendable fiber may also be resistant to fiber hacking, in which the signal in a fiber is surreptitiously monitored by bending the fiber and detecting the leakage.

Another important feature of cable is cable withstanding against the horizontally applied force. It is technically called max tensile strength defining how much force can applied to the cable during the installation period.

Telecom Anatolia fiber optic cable versions are reinforced with aramid yarns or glass yarns as intermediary strength member. In commercial terms, usage of the glass yarns are more cost effective while no loss in mechanical durability of the cable. Glass yarns also protect the cable core against rodents and termites.

Termination and splicing

ST connectors on multi-mode fiber.

Optical fibers are connected to terminal equipment by optical fiber connectors. These connectors are usually of a standard type such as *FC*, *SC*, *ST*, *LC*, or *MTRJ*.

Optical fibers may be connected to each other by connectors or by *splicing*, that is, joining two fibers together to form a continuous optical waveguide. The generally

accepted splicing method is arc fusion splicing, which melts the fiber ends together with an electric arc. For quicker fastening jobs, a "mechanical splice" is used.

Fusion splicing is done with a specialized instrument that typically operates as follows: The two cable ends are fastened inside a splice enclosure that will protect the splices, and the fiber ends are stripped of their protective polymer coating (as well as the more sturdy outer jacket, if present). The ends are *cleaved* (cut) with a precision cleaver to make them perpendicular, and are placed into special holders in the splicer. The splice is usually inspected via a magnified viewing screen to check the cleaves before and after the splice. The splicer uses small motors to align the end faces together, and emits a small spark between electrodes at the gap to burn off dust and moisture. Then the splicer generates a larger spark that raises the temperature above the melting point of the glass, fusing the ends together permanently. The location and energy of the spark is carefully controlled so that the molten core and cladding do not mix, and this minimizes optical loss. A splice loss estimate is measured by the splicer, by directing light through the cladding on one side and measuring the light leaking from the cladding on the other side. A splice loss under 0.1 dB is typical. The complexity of this process makes fiber splicing much more difficult than splicing copper wire.

Mechanical fiber splices are designed to be quicker and easier to install, but there is still the need for stripping, careful cleaning and precision cleaving. The fiber ends are aligned and held together by a precision-made sleeve, often using a clear index-matching gel that enhances the transmission of light across the joint. Such joints typically have higher optical loss and are less robust than fusion splices, especially if the gel is used. All splicing techniques involve the use of an enclosure into which the splice is placed for protection afterward.

Fibers are terminated in connectors so that the fiber end is held at the end face precisely and securely. A fiber-optic connector is basically a rigid cylindrical barrel surrounded by a sleeve that holds the barrel in its mating socket. The mating mechanism can be "push and click", "turn and latch" ("bayonet"), or screw-in (threaded). A typical connector is installed by preparing the fiber end and inserting it into the rear of the connector body. Quick-set adhesive is usually used so the fiber is held securely, and a strain relief is secured to the rear. Once the adhesive has set, the fiber's end is polished to a mirror finish. Various polish profiles are used, depending on the type of fiber and the application. For single-mode fiber, the fiber ends are typically polished with a slight curvature, such that when the connectors are mated the fibers touch only at their cores. This is known as a "physical contact" (PC) polish. The curved surface may be polished at an angle, to make an "angled physical contact" (APC) connection. Such connections have higher loss than PC connections, but greatly reduced back reflection, because light that reflects from the angled surface leaks out of the fiber core; the resulting loss in signal strength is known as gap loss. APC fiber ends have low back reflection even when disconnected.

In the 1990s, terminating fiber optic cables was very labor intensive. The number of parts per connector, polishing of the fibers, and the need to oven-bake the epoxy in each

connector made terminating fiber optic cables very difficult. Today, many different connectors are on the market and offer an easier, less labor intensive way of terminating the cables. Some of the most popular connectors have already been polished from the factory and include a gel inside the connector and those two steps help save money on labor especially on large projects. A cleave is made at a required length in order to get as close to the polished piece already inside the connector, with the gel surrounding the point where the two piece meet inside the connector very little light loss is exposed.

Free-space coupling

It is often necessary to align an optical fiber with another optical fiber, or with an optoelectronic device such as a light-emitting diode, a laser diode, or a modulator. This can involve either carefully aligning the fiber and placing it in contact with the device, or can use a lens to allow coupling over an air gap. In some cases the end of the fiber is polished into a curved form that is designed to allow it to act as a lens.

In a laboratory environment, a bare fiber end is coupled using a fiber launch system, which uses a microscope objective lens to focus the light down to a fine point. A precision translation stage (micro-positioning table) is used to move the lens, fiber, or device to allow the coupling efficiency to be optimized. Fibers with a connector on the end make this process much simpler: the connector is simply plugged into a pre-aligned fiberoptic collimator, which contains a lens that is either accurately positioned with respect to the fiber, or is adjustable. To achieve the best injection efficiency into single-mode fiber, the direction, position, size and divergence of the beam must all be optimized. With good beams, 70 to 90% coupling efficiency can be achieved.

With properly polished single-mode fibers, the emitted beam has an almost perfect Gaussian shape—even in the far field—if a good lens is used. The lens needs to be large enough to support the full numerical aperture of the fiber, and must not introduce aberrations in the beam. Aspheric lenses are typically used.

Fiber fuse

At high optical intensities, above 2 megawatts per square centimeter, when a fiber is subjected to a shock or is otherwise suddenly damaged, a *fiber fuse* can occur. The reflection from the damage vaporizes the fiber immediately before the break, and this new defect remains reflective so that the damage propagates back toward the transmitter at 1–3 meters per second (4−11 km/h, 2–8 mph). The open fiber control system, which ensures laser eye safety in the event of a broken fiber, can also effectively halt propagation of the fiber fuse. In situations, such as undersea cables, where high power levels might be used without the need for open fiber control, a "fiber fuse" protection device at the transmitter can break the circuit to prevent any damage.

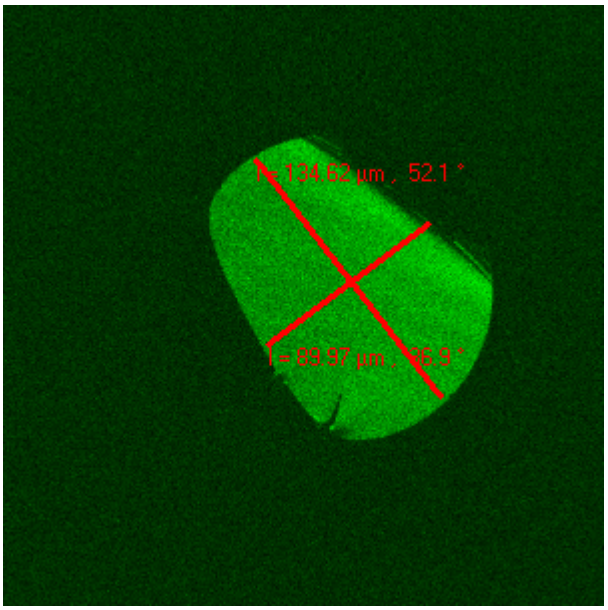
Example

Fiber connections can be used for various types of connections. For example, most high definition televisions offer a digital audio optical connection. This allows the streaming of audio over light, using the TOSLink protocol.

Electric power transmission

Optical fiber can be used to transmit electricity. While the efficiency is not nearly that of traditional copper wire, it is especially useful in situations where it is desirable to not have a metallic conductor as in the case of use near MRI machines which produce strong magnetic currents.

Preform



Cross-section of a fiber drawn from a D-shaped **preform**

A preform is a piece of glass used to draw an optical fiber. The preform may consist of several pieces of a glass with different refractive index, to provide the core and cladding of the fiber. The shape of the preform may be circular, although for some applications such as double-clad fibers another form is preferred. In fiber lasers based on double-clad fiber, an asymmetric shape improves the filling factor for laser pumping.

Due to the surface tension, the shape is smoothed during the drawing process, and the shape of the resulting fiber does not reproduce the sharp edges of the preform. Nevertheless, the careful polishing of the **preform** is important, any defects of the **preform** surface affect the optical and mechanical properties of the resulting fiber. In particular, the preform for the test-fiber shown in the figure was not polished well, and the cracks are seen with confocal optical microscope.

Chapter 7

Semiconductor

A **semiconductor** is a material with electrical conductivity due to electron flow (as opposed to ionic conductivity) intermediate in magnitude between that of a conductor and an insulator. This means a conductivity roughly in the range of 10^3 to 10^{-8} siemens per centimeter. Semiconductor materials are the foundation of modern electronics, including radio, computers, telephones, and many other devices. Such devices include transistors, solar cells, many kinds of diodes including the light-emitting diode, the silicon controlled rectifier, and digital and analog integrated circuits. Similarly, semiconductor solar photovoltaic panels directly convert light energy into electrical energy. In a metallic conductor, current is carried by the flow of electrons. In semiconductors, current is often schematized as being carried either by the flow of electrons or by the flow of positively charged "holes" in the electron structure of the material. Actually, however, in both cases only electron movements are involved.

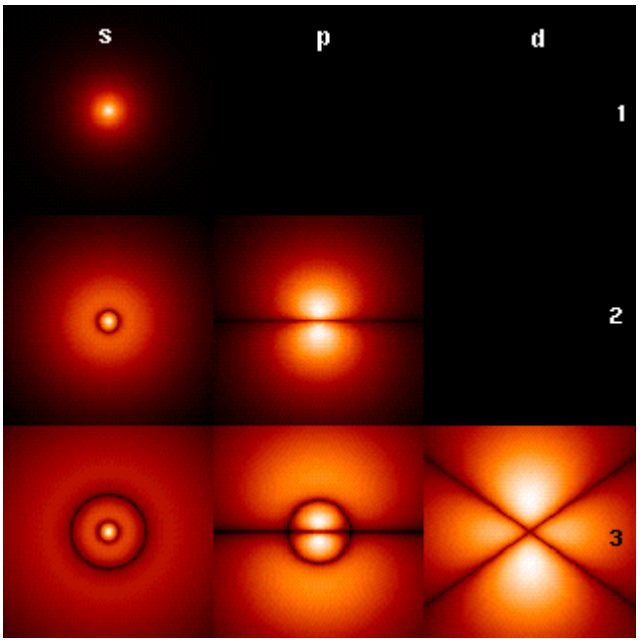
Common semiconducting materials are crystalline solids, but amorphous and liquid semiconductors are known. These include hydrogenated amorphous silicon and mixtures of arsenic, selenium and tellurium in a variety of proportions. Such compounds share with better known semiconductors intermediate conductivity and a rapid variation of conductivity with temperature, as well as occasional negative resistance. Such disordered materials lack the rigid crystalline structure of conventional semiconductors such as silicon and are generally used in thin film structures, which are less demanding for as concerns the electronic quality of the material and thus are relatively insensitive to impurities and radiation damage. Organic semiconductors, that is, organic materials with properties resembling conventional semiconductors, are also known.

Silicon is used to create most semiconductors commercially. Dozens of other materials are used, including germanium, gallium arsenide, and silicon carbide. A pure semiconductor is often called an “intrinsic” semiconductor. The electronic properties and the conductivity of a semiconductor can be changed in a controlled manner by adding very small quantities of other elements, called “dopants”, to the intrinsic material. In crystalline silicon typically this is achieved by adding impurities of boron or phosphorus to the melt and then allowing the melt to solidify into the crystal. This process is called "doping".

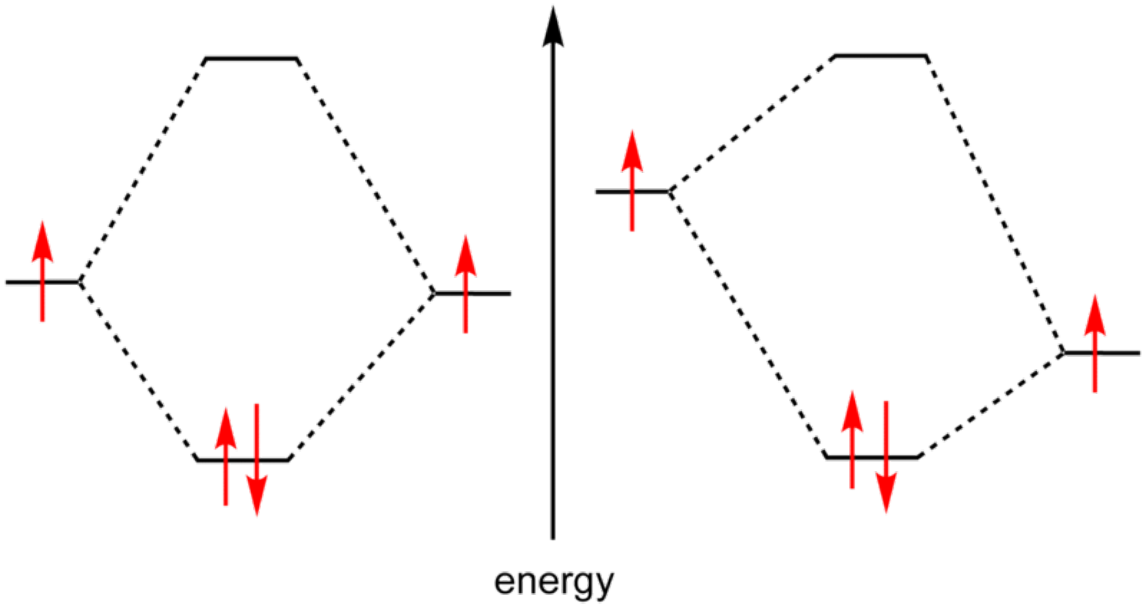
Explaining semiconductor energy bands

There are three popular ways to classify the electronic structure of a crystal.

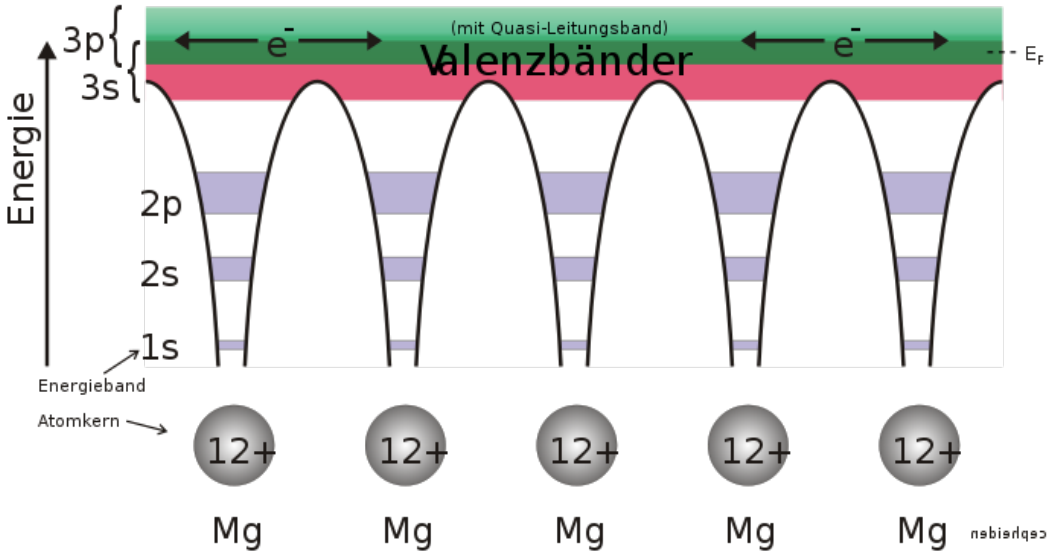
- Band structure
- atoms – crystal – vacuum



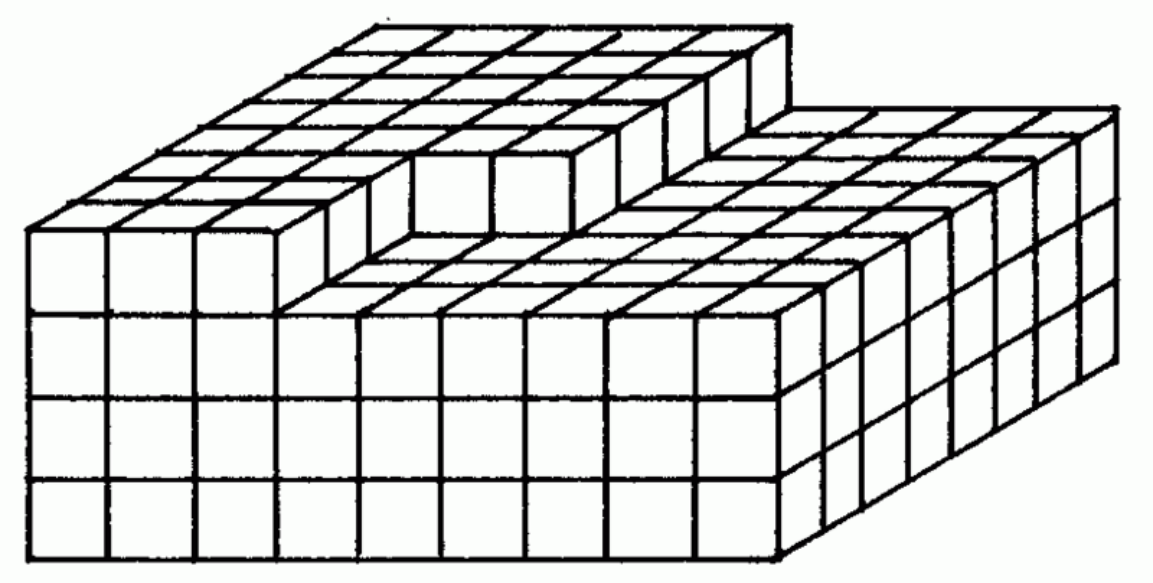
In a single H-atom an electron resides in well known orbitals. Note that the orbitals are called s,p,d in order of increasing circular current.



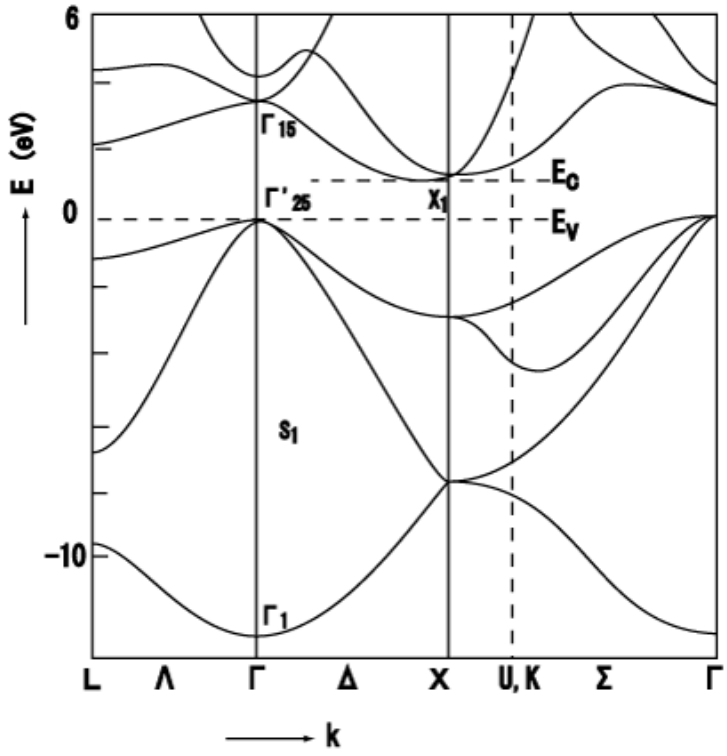
Putting two atoms together leads to delocalized orbitals across two atoms, yielding a covalent bond. Due to the Pauli exclusion principle, every state can contain only one electron.



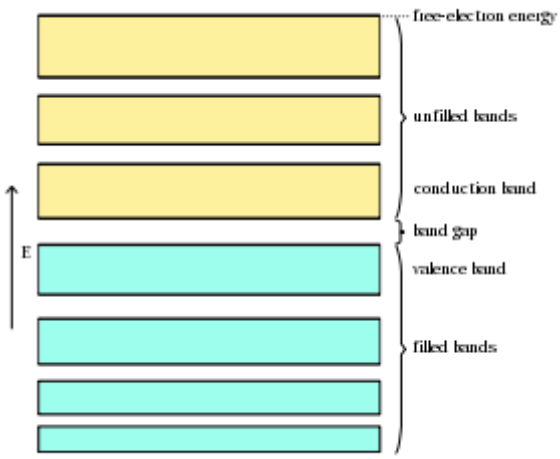
This can be continued with more atoms. Note: This picture shows a metal, not an actual semiconductor.



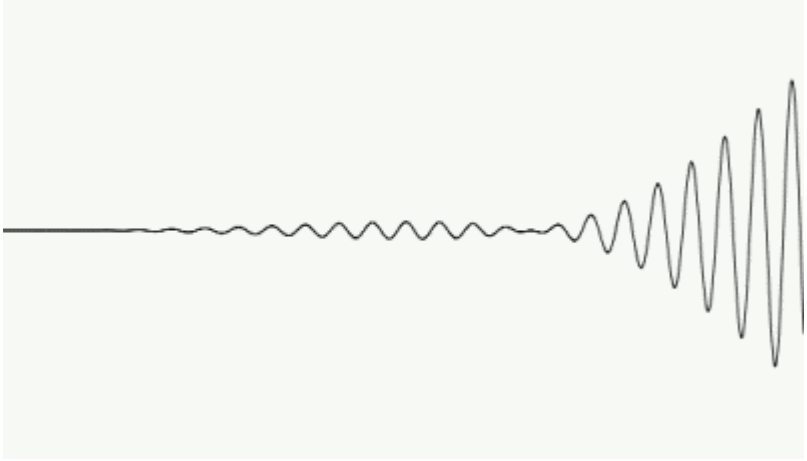
Continuing to add creates a crystal, which may then be cut into a tape and fused together at the ends to allow circular currents.



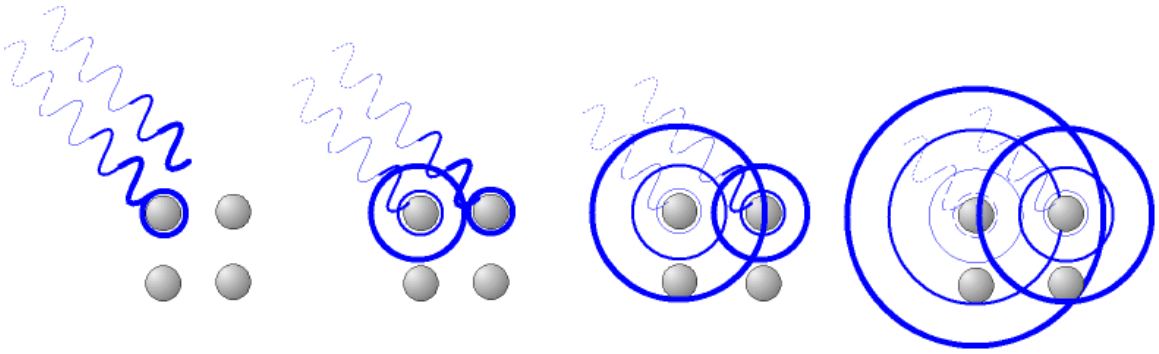
For this regular solid the band structure can be calculated or measured.



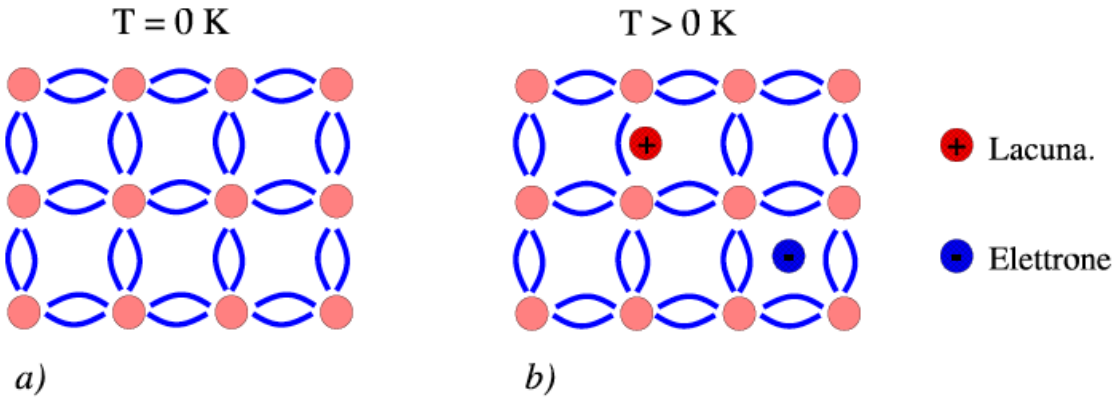
Integrating over the k axis gives the bands of a semiconductor showing a full valence band and an empty conduction band. Generally stopping at the vacuum level is undesirable, because some people want to calculate: photoemission, inverse photoemission



After the band structure is determined states can be combined to generate wave packets. As this is analogous to wave packages in free space, the results are similar.



An alternative description, which does not really appreciate the strong Coulomb interaction, shoots free electrons into the crystal and looks at the scattering.



A third alternative description uses strongly localized unpaired electrons in chemical bonds, which looks almost like a Mott insulator.

Energy bands and electrical conduction

In classic crystalline semiconductors, the electrons can have energies only within certain bands (i.e. ranges of levels of energy). Energetically, these bands are located between the energy of the ground state, corresponding to electrons tightly bound to the atomic nuclei of the material, and the free electron energy. The latter is the energy required for an electron to escape entirely from the material. The energy bands each correspond to a large number of discrete quantum states of the electrons, and most of the states with low energy (closer to the nucleus) are full, up to a particular band called the *valence band*. Semiconductors and insulators are distinguished from metals because the valence band in them is nearly filled with electrons under usual operating conditions, while very few (semiconductor) or virtually none (insulator) of them are available in the *conduction band*, the band immediately above the valence band.

The ease with which electrons in a semiconductor can be excited from the valence band to the conduction band depends on the band gap between the bands. The size of this energy bandgap serves as an arbitrary dividing line (roughly 4 eV) between semiconductors and insulators.

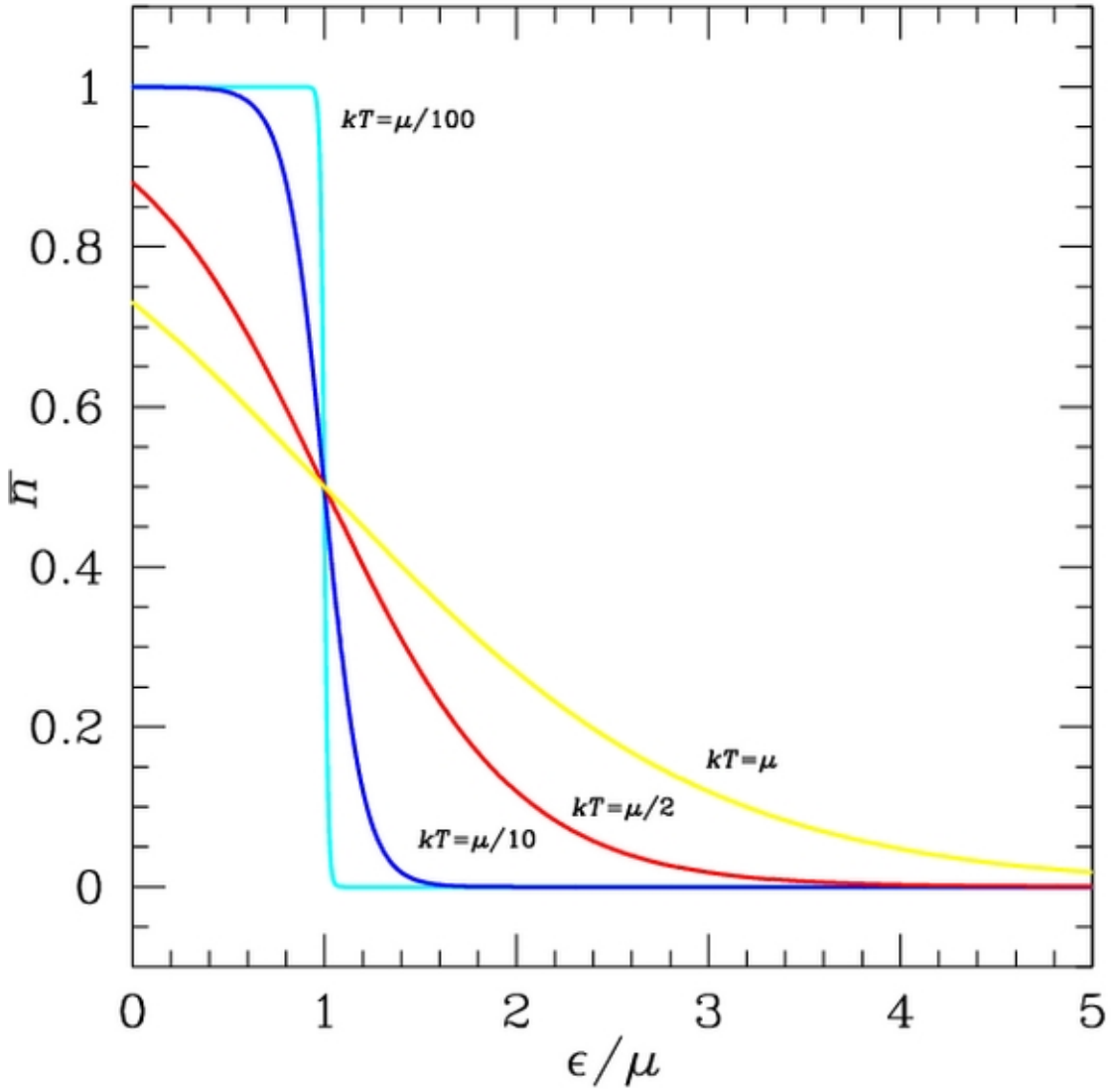
With covalent bonds, an electron moves by hopping to a neighboring bond. The Pauli exclusion principle requires the electron to be lifted into the higher anti-bonding state of that bond. For delocalized states, for example in one dimension – that is in a nanowire, for every energy there is a state with electrons flowing in one direction and another state with the electrons flowing in the other. For a net current to flow, more states for one direction than for the other direction must be occupied. For this to occur, energy is required, as in the semiconductor the next higher states lie above the band gap. Often this is stated as: full bands do not contribute to the electrical conductivity. However, as the temperature of a semiconductor rises above absolute zero, there is more energy in the semiconductor to spend on lattice vibration and — more importantly for us — on lifting some electrons into an energy states of the conduction band. The current-carrying electrons in the conduction band are known as "free electrons", although they are often simply called "electrons" if context allows this usage to be clear.

Electrons excited to the conduction band also leave behind electron holes, or unoccupied states in the valence band. Both the conduction band electrons and the valence band holes contribute to electrical conductivity. The holes themselves don't actually move, but a neighboring electron can move to fill the hole, leaving a hole at the place it has just come from, and in this way the holes appear to move, and the holes behave as if they were actual positively charged particles.

One covalent bond between neighboring atoms in the solid is ten times stronger than the binding of the single electron to the atom, so freeing the electron does not imply destruction of the crystal structure.

Holes: electron absence as a charge carrier

The concept of holes can also be applied to metals, where the Fermi level lies *within* the conduction band. With most metals the Hall effect indicates electrons are the charge carriers. However, some metals have a mostly filled conduction band. In these, the Hall effect reveals positive charge carriers, which are not the ion-cores, but holes. In the case of a metal, only a small amount of energy is needed for the electrons to find other unoccupied states to move into, and hence for current to flow. Sometimes even in this case it may be said that a hole was left behind, to explain why the electron does not fall back to lower energies: It cannot find a hole. In the end in both materials electron-phonon scattering and defects are the dominant causes for resistance.



Fermi-Dirac distribution. States with energy ε below the Fermi energy, here μ, have higher probability *n* to be occupied, and those above are less likely to be occupied. Smearing of the distribution increases with temperature.

The energy distribution of the electrons determines which of the states are filled and which are empty. This distribution is described by Fermi-Dirac statistics. The distribution is characterized by the temperature of the electrons, and the *Fermi energy* or *Fermi level*. Under absolute zero conditions the Fermi energy can be thought of as the energy up to which available electron states are occupied. At higher temperatures, the Fermi energy is the energy at which the probability of a state being occupied has fallen to 0.5.

The dependence of the electron energy distribution on temperature also explains why the conductivity of a semiconductor has a strong temperature dependency, as a

semiconductor operating at lower temperatures will have fewer available free electrons and holes able to do the work.

Energy–momentum dispersion

In the preceding description an important fact is ignored for the sake of simplicity: the *dispersion* of the energy. The reason that the energies of the states are broadened into a band is that the energy depends on the value of the wave vector, or *k-vector*, of the electron. The k-vector, in quantum mechanics, is the representation of the momentum of a particle.

The dispersion relationship determines the effective mass, m^*, of electrons or holes in the semiconductor, according to the formula:

$$m^* = \hbar^2 \cdot \left[\frac{d^2 E(k)}{dk^2}\right]^{-1}.$$

The effective mass is important as it affects many of the electrical properties of the semiconductor, such as the electron or hole mobility, which in turn influences the *diffusivity* of the charge carriers and the electrical conductivity of the semiconductor.

Typically the effective mass of electrons and holes are different. This affects the relative performance of *p-channel* and *n-channel* IGFETs.

The top of the valence band and the bottom of the conduction band might not occur at that same value of *k*. Materials with this situation, such as silicon and germanium, are known as *indirect bandgap* materials. Materials in which the band extrema are aligned in *k*, for example gallium arsenide, are called *direct bandgap* semiconductors. Direct gap semiconductors are particularly important in optoelectronics because they are much more efficient as light emitters than indirect gap materials.

Carrier generation and recombination

When ionizing radiation strikes a semiconductor, it may excite an electron out of its energy level and consequently leave a hole. This process is known as *electron–hole pair generation*. Electron-hole pairs are constantly generated from thermal energy as well, in the absence of any external energy source.

Electron-hole pairs are also apt to recombine. Conservation of energy demands that these recombination events, in which an electron loses an amount of energy larger than the band gap, be accompanied by the emission of thermal energy (in the form of phonons) or radiation (in the form of photons).

In some states, the generation and recombination of electron–hole pairs are in equipoise. The number of electron-hole pairs in the steady state at a given temperature is determined

by quantum statistical mechanics. The precise quantum mechanical mechanisms of generation and recombination are governed by conservation of energy and conservation of momentum.

As the probability that electrons and holes meet together is proportional to the product of their amounts, the product is in steady state nearly constant at a given temperature, providing that there is no significant electric field (which might "flush" carriers of both types, or move them from neighbour regions containing more of them to meet together) or externally driven pair generation. The product is a function of the temperature, as the probability of getting enough thermal energy to produce a pair increases with temperature, being approximately $\exp(-E_G/kT)$, where k is Boltzmann's constant, T is absolute temperature and E_G is band gap.

The probability of meeting is increased by carrier traps—impurities or dislocations which can trap an electron or hole and hold it until a pair is completed. Such carrier traps are sometimes purposely added to reduce the time needed to reach the steady state.

Semi-insulators

Some materials are classified as **semi-insulators**. These have electrical conductivity nearer to that of electrical insulators. Semi-insulators find niche applications in micro-electronics, such as substrates for HEMT. An example of a common semi-insulator is gallium arsenide.

Doping

The property of semiconductors that makes them most useful for constructing electronic devices is that their conductivity may easily be modified by introducing impurities into their crystal lattice. The process of adding controlled impurities to a semiconductor is known as *doping*. The amount of impurity, or dopant, added to an *intrinsic* (pure) semiconductor varies its level of conductivity. Doped semiconductors are often referred to as *extrinsic*. By adding impurity to pure semiconductors, the electrical conductivity may be varied not only by the number of impurity atoms but also, by the type of impurity atom and the changes may be thousand folds and million folds. For example, 1 cm^3 of a metal or semiconductor specimen has a number of atoms on the order of 10^{22}. Since every atom in metal donates at least one free electron for conduction in metal, 1 cm^3 of metal contains free electrons on the order of 10^{22}. At the temperature close to 20 °C , 1 cm^3 of pure germanium contains about 4.2×10^{22} atoms and 2.5×10^{13} free electrons and 2.5×10^{13} holes (empty spaces in crystal lattice having positive charge) The addition of 0.001% of arsenic (an impurity) donates an extra 10^{17} free electrons in the same volume and the electrical conductivity increases about 10,000 times."

Dopants

The materials chosen as suitable dopants depend on the atomic properties of both the dopant and the material to be doped. In general, dopants that produce the desired

controlled changes are classified as either electron acceptors or donors. A donor atom that activates (that is, becomes incorporated into the crystal lattice) donates weakly bound valence electrons to the material, creating excess negative charge carriers. These weakly bound electrons can move about in the crystal lattice relatively freely and can facilitate conduction in the presence of an electric field. (The donor atoms introduce some states under, but very close to the conduction band edge. Electrons at these states can be easily excited to the conduction band, becoming free electrons, at room temperature.) Conversely, an activated acceptor produces a hole. Semiconductors doped with *donor* impurities are called *n-type*, while those doped with *acceptor* impurities are known as *p-type*. The n and p type designations indicate which charge carrier acts as the material's majority carrier. The opposite carrier is called the minority carrier, which exists due to thermal excitation at a much lower concentration compared to the majority carrier.

For example, the pure semiconductor silicon has four valence electrons. In silicon, the most common dopants are IUPAC group 13 (commonly known as *group III*) and group 15 (commonly known as *group V*) elements. Group 13 elements all contain three valence electrons, causing them to function as acceptors when used to dope silicon. Group 15 elements have five valence electrons, which allows them to act as a donor. Therefore, a silicon crystal doped with boron creates a p-type semiconductor whereas one doped with phosphorus results in an n-type material.

Carrier concentration

The concentration of dopant introduced to an intrinsic semiconductor determines its concentration and indirectly affects many of its electrical properties. The most important factor that doping directly affects is the material's carrier concentration. In an intrinsic semiconductor under thermal equilibrium, the concentration of electrons and holes is equivalent. That is,

$$n = p = n_i.$$

If we have a non-intrinsic semiconductor in thermal equilibrium the relation becomes:

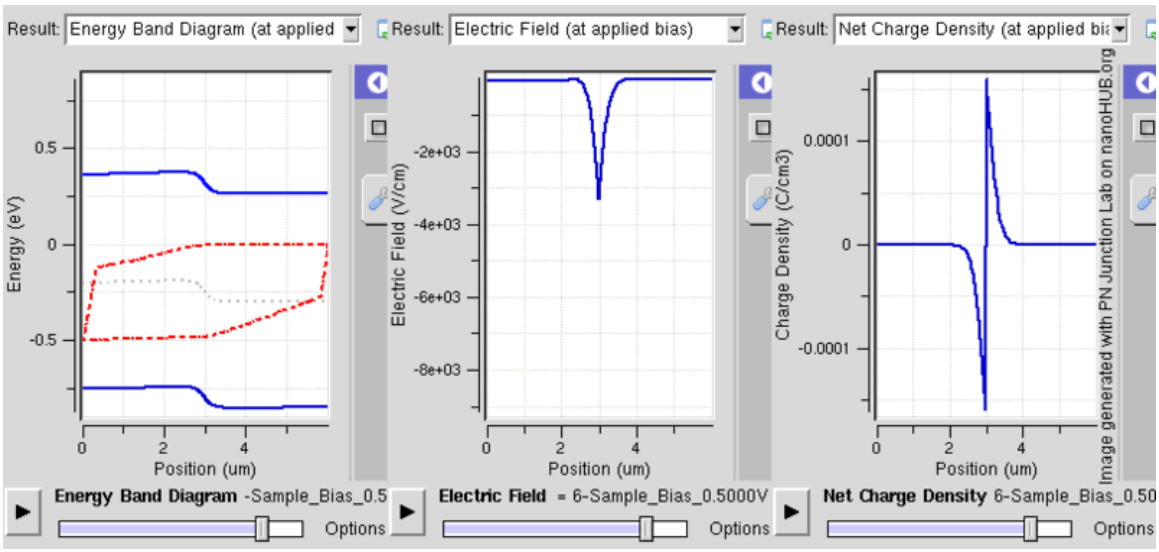
$$n_0 \cdot p_0 = n_i^2$$

where n_0 is the concentration of conducting electrons, p_0 is the electron hole concentration, and n_i is the material's intrinsic carrier concentration. Intrinsic carrier concentration varies between materials and is dependent on temperature. Silicon's n_i, for example, is roughly 1.08×10^{10} cm^{-3} at 300 kelvins (room temperature).

In general, an increase in doping concentration affords an increase in conductivity due to the higher concentration of carriers available for conduction. Degenerately (very highly) doped semiconductors have conductivity levels comparable to metals and are often used in modern integrated circuits as a replacement for metal. Often superscript plus and minus symbols are used to denote relative doping concentration in semiconductors. For example, n^+ denotes an n-type semiconductor with a high, often degenerate, doping

concentration. Similarly, p^- would indicate a very lightly doped p-type material. It is useful to note that even degenerate levels of doping imply low concentrations of impurities with respect to the base semiconductor. In crystalline intrinsic silicon, there are approximately 5×10^{22} atoms/cm³. Doping concentration for silicon semiconductors may range anywhere from 10^{13} cm^{-3} to 10^{18} cm^{-3}. Doping concentration above about 10^{18} cm^{-3} is considered degenerate at room temperature. Degenerately doped silicon contains a proportion of impurity to silicon on the order of parts per thousand. This proportion may be reduced to parts per billion in very lightly doped silicon. Typical concentration values fall somewhere in this range and are tailored to produce the desired properties in the device that the semiconductor is intended for.

Effect on band structure



Band diagram of PN junction operation in forward bias mode showing reducing depletion width. Both p and n junctions are doped at a 1e15/cm3 doping level, leading to built-in potential of ~0.59V. Reducing depletion width can be inferred from the shrinking charge profile, as fewer dopants are exposed with increasing forward bias.

Doping a semiconductor crystal introduces allowed energy states within the band gap but very close to the energy band that corresponds to the dopant type. In other words, donor impurities create states near the conduction band while acceptors create states near the valence band. The gap between these energy states and the nearest energy band is usually referred to as dopant-site bonding energy or E_B and is relatively small. For example, the E_B for boron in silicon bulk is 0.045 eV, compared with silicon's band gap of about 1.12 eV. Because E_B is so small, it takes little energy to ionize the dopant atoms and create free carriers in the conduction or valence bands. Usually the thermal energy available at room temperature is sufficient to ionize most of the dopant.

Dopants also have the important effect of shifting the material's Fermi level towards the energy band that corresponds with the dopant with the greatest concentration. Since the Fermi level must remain constant in a system in thermodynamic equilibrium, stacking

layers of materials with different properties leads to many useful electrical properties. For example, the p-n junction's properties are due to the energy band bending that happens as a result of lining up the Fermi levels in contacting regions of p-type and n-type material.

This effect is shown in a *band diagram*. The band diagram typically indicates the variation in the valence band and conduction band edges versus some spatial dimension, often denoted *x*. The Fermi energy is also usually indicated in the diagram. Sometimes the *intrinsic Fermi energy*, E_i, which is the Fermi level in the absence of doping, is shown. These diagrams are useful in explaining the operation of many kinds of semiconductor devices.

Preparation of semiconductor materials

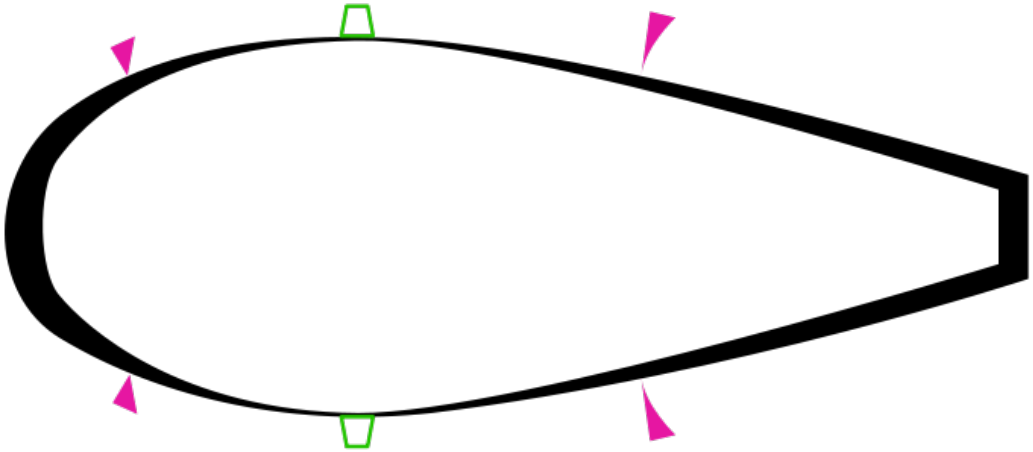
Semiconductors with predictable, reliable electronic properties are necessary for mass production. The level of chemical purity needed is extremely high because the presence of impurities even in very small proportions can have large effects on the properties of the material. A high degree of crystalline perfection is also required, since faults in crystal structure (such as dislocations, twins, and stacking faults) interfere with the semiconducting properties of the material. Crystalline faults are a major cause of defective semiconductor devices. The larger the crystal, the more difficult it is to achieve the necessary perfection. Current mass production processes use crystal ingots between 100 mm and 300 mm (4–12 inches) in diameter which are grown as cylinders and sliced into wafers.

Because of the required level of chemical purity and the perfection of the crystal structure which are needed to make semiconductor devices, special methods have been developed to produce the initial semiconductor material. A technique for achieving high purity includes growing the crystal using the Czochralski process. An additional step that can be used to further increase purity is known as zone refining. In zone refining, part of a solid crystal is melted. The impurities tend to concentrate in the melted region, while the desired material recrystalizes leaving the solid material more pure and with fewer crystalline faults.

In manufacturing semiconductor devices involving heterojunctions between different semiconductor materials, the lattice constant, which is the length of the repeating element of the crystal structure, is important for determining the compatibility of materials.

Chapter 8

Fluid Dynamics



Typical aerodynamic teardrop shape, assuming a viscous medium passing from left to right, the diagram shows the pressure distribution as the thickness of the black line and shows the velocity in the boundary layer as the violet triangles. The green vortex generators prompt the transition to turbulent flow and prevent back-flow also called flow separation from the high pressure region in the back. The surface in front is as smooth as possible or even employs shark like skin, as any turbulence here will reduce the energy of the airflow. The truncation on the right, known as a Kammback, also prevents back flow from the high pressure region in the back across the spoilers to the convergent part.

In physics, **fluid dynamics** is a sub-discipline of fluid mechanics that deals with **fluid flow**—the natural science of fluids (liquids and gases) in motion. It has several subdisciplines itself, including aerodynamics (the study of air and other gases in motion) and **hydrodynamics** (the study of liquids in motion). Fluid dynamics has a wide range of applications, including calculating forces and moments on aircraft, determining the mass flow rate of petroleum through pipelines, predicting weather patterns, understanding nebulae in interstellar space and reportedly modeling fission weapon detonation. Some of its principles are even used in traffic engineering, where traffic is treated as a continuous fluid.

Fluid dynamics offers a systematic structure that underlies these practical disciplines, that embraces empirical and semi-empirical laws derived from flow measurement and used to solve practical problems. The solution to a fluid dynamics problem typically involves calculating various properties of the fluid, such as velocity, pressure, density, and temperature, as functions of space and time.

Historically, *hydrodynamics* meant something different than it does today. Before the twentieth century, hydrodynamics was synonymous with fluid dynamics. This is still reflected in names of some fluid dynamics topics, like magnetohydrodynamics and hydrodynamic stability—both also applicable in, as well as being applied to, gases.

Equations of fluid dynamics

The foundational axioms of fluid dynamics are the conservation laws, specifically, conservation of mass, conservation of linear momentum (also known as Newton's Second Law of Motion), and conservation of energy (also known as First Law of Thermodynamics). These are based on classical mechanics and are modified in quantum mechanics and general relativity. They are expressed using the Reynolds Transport Theorem.

In addition to the above, fluids are assumed to obey the *continuum assumption*. Fluids are composed of molecules that collide with one another and solid objects. However, the continuum assumption considers fluids to be continuous, rather than discrete. Consequently, properties such as density, pressure, temperature, and velocity are taken to be well-defined at infinitesimally small points, and are assumed to vary continuously from one point to another. The fact that the fluid is made up of discrete molecules is ignored.

For fluids which are sufficiently dense to be a continuum, do not contain ionized species, and have velocities small in relation to the speed of light, the momentum equations for Newtonian fluids are the Navier-Stokes equations, which is a non-linear set of differential equations that describes the flow of a fluid whose stress depends linearly on velocity gradients and pressure. The unsimplified equations do not have a general closed-form solution, so they are primarily of use in Computational Fluid Dynamics. The equations can be simplified in a number of ways, all of which make them easier to solve. Some of them allow appropriate fluid dynamics problems to be solved in closed form.

In addition to the mass, momentum, and energy conservation equations, a thermodynamical equation of state giving the pressure as a function of other thermodynamic variables for the fluid is required to completely specify the problem. An example of this would be the perfect gas equation of state:

$$p = \frac{\rho R_u T}{M}$$

where p is pressure, ρ is density, R_u is the gas constant, M is the molar mass and T is temperature.

Compressible vs incompressible flow

All fluids are compressible to some extent, that is changes in pressure or temperature will result in changes in density. However, in many situations the changes in pressure and temperature are sufficiently small that the changes in density are negligible. In this case the flow can be modeled as an incompressible flow. Otherwise the more general compressible flow equations must be used.

Mathematically, incompressibility is expressed by saying that the density ρ of a fluid parcel does not change as it moves in the flow field, i.e.,

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = 0\,,$$

where D / Dt is the substantial derivative, which is the sum of local and convective derivatives. This additional constraint simplifies the governing equations, especially in the case when the fluid has a uniform density.

For flow of gases, to determine whether to use compressible or incompressible fluid dynamics, the Mach number of the flow is to be evaluated. As a rough guide, compressible effects can be ignored at Mach numbers below approximately 0.3. For liquids, whether the incompressible assumption is valid depends on the fluid properties (specifically the critical pressure and temperature of the fluid) and the flow conditions (how close to the critical pressure the actual flow pressure becomes). Acoustic problems always require allowing compressibility, since sound waves are compression waves involving changes in pressure and density of the medium through which they propagate.

Viscous vs inviscid flow

Viscous problems are those in which fluid friction has significant effects on the fluid motion.

The Reynolds number, which is a ratio between inertial and viscous forces, can be used to evaluate whether viscous or inviscid equations are appropriate to the problem.

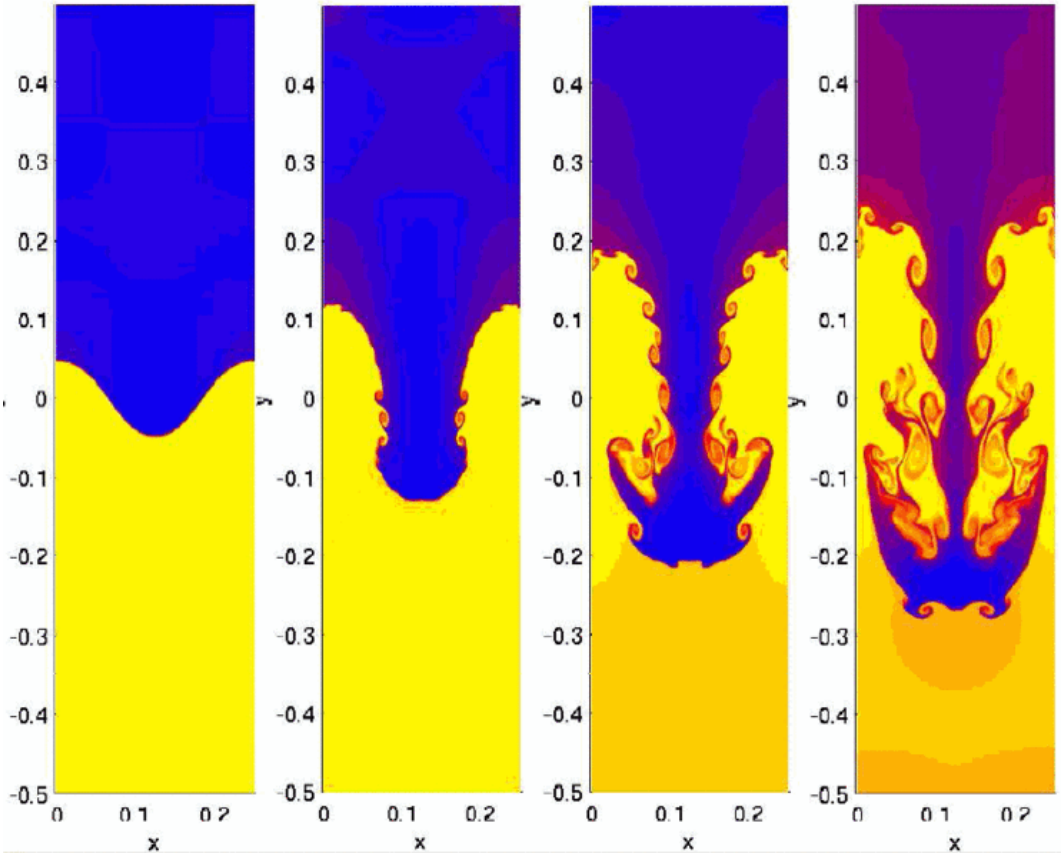
Stokes flow is flow at very low Reynolds numbers, $Re \ll 1$, such that inertial forces can be neglected compared to viscous forces.

On the contrary, high Reynolds numbers indicate that the inertial forces are more significant than the viscous (friction) forces. Therefore, we may assume the flow to be an inviscid flow, an approximation in which we neglect viscosity completely, compared to inertial terms.

This idea can work fairly well when the Reynolds number is high. However, certain problems such as those involving solid boundaries, may require that the viscosity be included. Viscosity often cannot be neglected near solid boundaries because the no-slip condition can generate a thin region of large strain rate (known as Boundary layer) which enhances the effect of even a small amount of viscosity, and thus generating vorticity. Therefore, to calculate net forces on bodies (such as wings) we should use viscous flow equations. As illustrated by d'Alembert's paradox, a body in an inviscid fluid will experience no drag force. The standard equations of inviscid flow are the Euler equations. Another often used model, especially in computational fluid dynamics, is to use the Euler equations away from the body and the boundary layer equations, which incorporates viscosity, in a region close to the body.

The Euler equations can be integrated along a streamline to get Bernoulli's equation. When the flow is everywhere irrotational and inviscid, Bernoulli's equation can be used throughout the flow field. Such flows are called potential flows.

Steady vs unsteady flow



Hydrodynamics simulation of the Rayleigh–Taylor instability

When all the time derivatives of a flow field vanish, the flow is considered to be a **steady flow**. Steady-state flow refers to the condition where the fluid properties at a point in the system do not change over time. Otherwise, flow is called unsteady. Whether a particular

flow is steady or unsteady, can depend on the chosen frame of reference. For instance, laminar flow over a sphere is steady in the frame of reference that is stationary with respect to the sphere. In a frame of reference that is stationary with respect to a background flow, the flow is unsteady.

Turbulent flows are unsteady by definition. A turbulent flow can, however, be statistically stationary. According to Pope:

The random field $U(x,t)$ is statistically stationary if all statistics are invariant under a shift in time.

This roughly means that all statistical properties are constant in time. Often, the mean field is the object of interest, and this is constant too in a statistically stationary flow.

Steady flows are often more tractable than otherwise similar unsteady flows. The governing equations of a steady problem have one dimension fewer (time) than the governing equations of the same problem without taking advantage of the steadiness of the flow field.

Laminar vs turbulent flow

Turbulence is flow characterized by recirculation, eddies, and apparent randomness. Flow in which turbulence is not exhibited is called laminar. It should be noted, however, that the presence of eddies or recirculation alone does not necessarily indicate turbulent flow—these phenomena may be present in laminar flow as well. Mathematically, turbulent flow is often represented via a Reynolds decomposition, in which the flow is broken down into the sum of an average component and a perturbation component.

It is believed that turbulent flows can be described well through the use of the Navier–Stokes equations. Direct numerical simulation (DNS), based on the Navier–Stokes equations, makes it possible to simulate turbulent flows at moderate Reynolds numbers. Restrictions depend on the power of the computer used and the efficiency of the solution algorithm. The results of DNS have been found to agree well with experimental data for some flows.

Most flows of interest have Reynolds numbers much too high for DNS to be a viable option, given the state of computational power for the next few decades. Any flight vehicle large enough to carry a human ($L > 3$ m), moving faster than 72 km/h (20 m/s) is well beyond the limit of DNS simulation ($Re = 4$ million). Transport aircraft wings (such as on an Airbus A300 or Boeing 747) have Reynolds numbers of 40 million (based on the wing chord). In order to solve these real-life flow problems, turbulence models will be a necessity for the foreseeable future. Reynolds-averaged Navier–Stokes equations (RANS) combined with turbulence modeling provides a model of the effects of the turbulent flow. Such a modeling mainly provides the additional momentum transfer by

the Reynolds stresses, although the turbulence also enhances the heat and mass transfer. Another promising methodology is large eddy simulation (LES), especially in the guise of detached eddy simulation (DES)—which is a combination of RANS turbulence modeling and large eddy simulation.

Newtonian vs non-Newtonian fluids

Sir Isaac Newton showed how stress and the rate of strain are very close to linearly related for many familiar fluids, such as water and air. These Newtonian fluids are modeled by a coefficient called viscosity, which depends on the specific fluid.

However, some of the other materials, such as emulsions and slurries and some visco-elastic materials (e.g. blood, some polymers), have more complicated *non-Newtonian* stress-strain behaviours. These materials include *sticky liquids* such as latex, honey, and lubricants which are studied in the sub-discipline of rheology.

Subsonic vs transonic, supersonic and hypersonic flows

While many terrestrial flows (e.g. flow of water through a pipe) occur at low mach numbers, many flows of practical interest (e.g. in aerodynamics) occur at high fractions of the Mach Number M=1 or in excess of it (supersonic flows). New phenomena occur at these Mach number regimes (e.g. shock waves for supersonic flow, transonic instability in a regime of flows with M nearly equal to 1, non-equilibrium chemical behavior due to ionization in hypersonic flows) and it is necessary to treat each of these flow regimes separately.

Magnetohydrodynamics

Magnetohydrodynamics is the multi-disciplinary study of the flow of electrically conducting fluids in electromagnetic fields. Examples of such fluids include plasmas, liquid metals, and salt water. The fluid flow equations are solved simultaneously with Maxwell's equations of electromagnetism.

Other approximations

There are a large number of other possible approximations to fluid dynamic problems. Some of the more commonly used are listed below.

- The **Boussinesq approximation** neglects variations in density except to calculate buoyancy forces. It is often used in free convection problems where density changes are small.
- **Lubrication theory** and **Hele-Shaw flow** exploits the large aspect ratio of the domain to show that certain terms in the equations are small and so can be neglected.
- **Slender-body theory** is a methodology used in Stokes flow problems to estimate the force on, or flow field around, a long slender object in a viscous fluid.

- The **shallow-water equations** can be used to describe a layer of relatively inviscid fluid with a free surface, in which surface gradients are small.
- The **Boussinesq equations** are applicable to surface waves on thicker layers of fluid and with steeper surface slopes.
- **Darcy's law** is used for flow in porous media, and works with variables averaged over several pore-widths.
- In rotating systems, the **quasi-geostrophic approximation** assumes an almost perfect balance between pressure gradients and the Coriolis force. It is useful in the study of atmospheric dynamics.

Terminology in fluid dynamics

The concept of pressure is central to the study of both fluid statics and fluid dynamics. A pressure can be identified for every point in a body of fluid, regardless of whether the fluid is in motion or not. Pressure can be measured using an aneroid, Bourdon tube, mercury column, or various other methods.

Some of the terminology that is necessary in the study of fluid dynamics is not found in other similar areas of study. In particular, some of the terminology used in fluid dynamics is not used in fluid statics.

Terminology in incompressible fluid dynamics

The concepts of total pressure and dynamic pressure arise from Bernoulli's equation and are significant in the study of all fluid flows. (These two pressures are not pressures in the usual sense—they cannot be measured using an aneroid, Bourdon tube or mercury column.) To avoid potential ambiguity when referring to pressure in fluid dynamics, many authors use the term static pressure to distinguish it from total pressure and dynamic pressure. Static pressure is identical to pressure and can be identified for every point in a fluid flow field.

In *Aerodynamics*, L.J. Clancy writes: *To distinguish it from the total and dynamic pressures, the actual pressure of the fluid, which is associated not with its motion but with its state, is often referred to as the static pressure, but where the term pressure alone is used it refers to this static pressure.*

A point in a fluid flow where the flow has come to rest (i.e. speed is equal to zero adjacent to some solid body immersed in the fluid flow) is of special significance. It is of such importance that it is given a special name—a stagnation point. The static pressure at the stagnation point is of special significance and is given its own name—stagnation pressure. In incompressible flows, the stagnation pressure at a stagnation point is equal to the total pressure throughout the flow field.

Terminology in compressible fluid dynamics

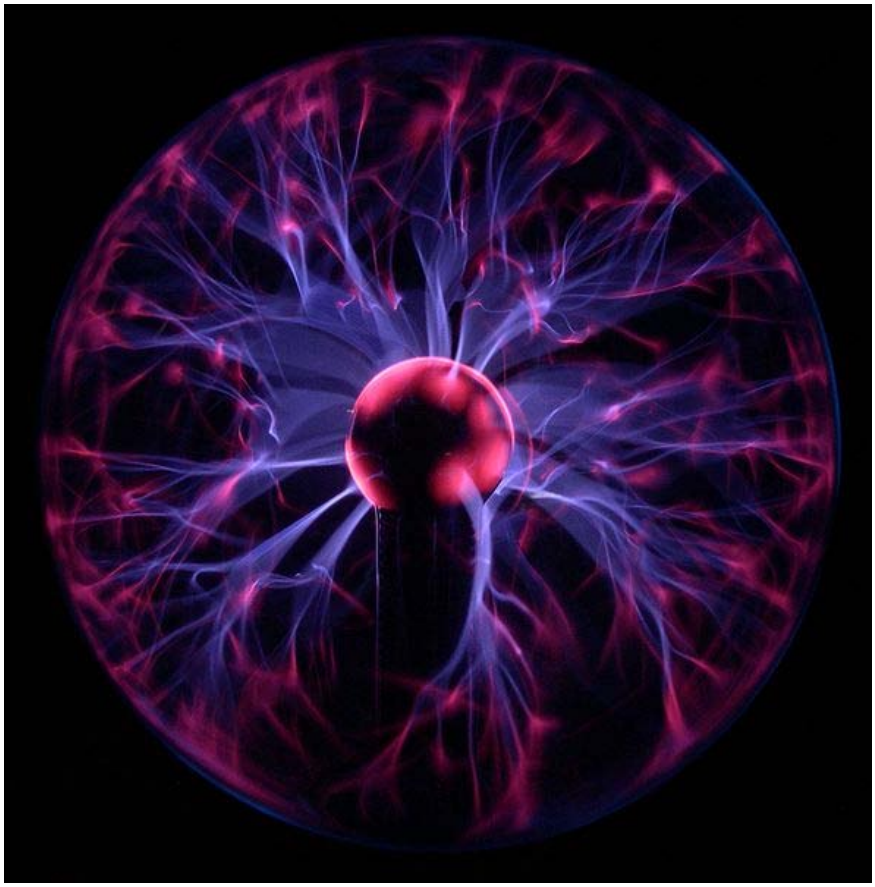
In a compressible fluid, such as air, the temperature and density are essential when determining the state of the fluid. In addition to the concept of total pressure (also known as stagnation pressure), the concepts of total (or stagnation) temperature and total (or stagnation) density are also essential in any study of compressible fluid flows. To avoid potential ambiguity when referring to temperature and density, many authors use the terms static temperature and static density. Static temperature is identical to temperature; and static density is identical to density; and both can be identified for every point in a fluid flow field.

The temperature and density at a stagnation point are called stagnation temperature and stagnation density.

A similar approach is also taken with the thermodynamic properties of compressible fluids. Many authors use the terms total (or stagnation) enthalpy and total (or stagnation) entropy. The terms static enthalpy and static entropy appear to be less common, but where they are used they mean nothing more than enthalpy and entropy respectively, and the prefix "static" is being used to avoid ambiguity with their 'total' or 'stagnation' counterparts. Because the 'total' flow conditions are defined by isentropically bringing the fluid to rest, the total (or stagnation) entropy is by definition always equal to the "static" entropy.

Chapter 9

Plasma (Physics)



Plasma lamp, illustrating some of the more complex phenomena of a plasma, including *filamentation*. The colors are a result of relaxation of electrons in excited states to lower energy states after they have recombined with ions. These processes emit light in a spectrum characteristic of the gas being excited.

In physics and chemistry, **plasma** is a state of matter similar to gas in which a certain portion of the particles are ionized. The basic premise is that heating a gas dissociates its molecular bonds, rendering it into its constituent atoms. Further heating leads to

ionization (a loss of electrons), turning it into a plasma: containing charged particles, positive ions and negative electrons.

The presence of a non-negligible number of charge carriers makes the plasma electrically conductive so that it responds strongly to electromagnetic fields. Plasma, therefore, has properties quite unlike those of solids, liquids, or gases and is considered a distinct state of matter. Like gas, plasma does not have a definite shape or a definite volume unless enclosed in a container; unlike gas, under the influence of a magnetic field, it may form structures such as filaments, beams and double layers. Some common plasmas are stars and neon signs.

Plasma was first identified in a Crookes tube, and so described by Sir William Crookes in 1879 (he called it "radiant matter"). The nature of the Crookes tube "cathode ray" matter was subsequently identified by British physicist Sir J.J. Thomson in 1897, and dubbed "plasma" by Irving Langmuir in 1928, perhaps because it reminded him of a blood plasma. Langmuir wrote:

Except near the electrodes, where there are *sheaths* containing very few electrons, the ionized gas contains ions and electrons in about equal numbers so that the resultant space charge is very small. We shall use the name *plasma* to describe this region containing balanced charges of ions and electrons.

Common plasmas

Plasmas are by far the most common phase of matter in the universe, both by mass and by volume. All the stars are made of plasma, and even the space between the stars is filled with a plasma, albeit a very sparse one. In our solar system, the planet Jupiter accounts for most of the *non*-plasma, only about 0.1% of the mass and 10^{-15}% of the volume within the orbit of Pluto. Very small grains within a gaseous plasma will also pick up a net negative charge, so that they in turn may act like a very heavy negative ion component of the plasma.

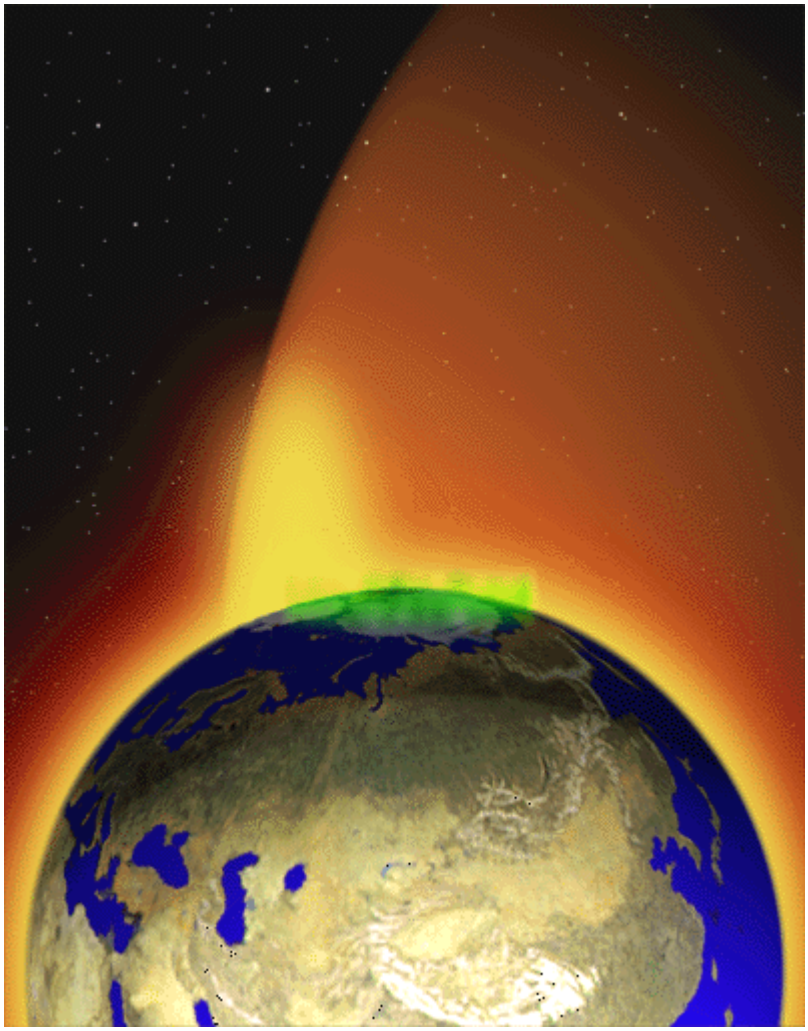
Common forms of plasma

Artificially produced	**Terrestrial plasmas**	**Space and Astrophysical plasmas**
• Those found in plasma displays, including TVs • Inside fluorescent lamps (low energy lighting), neon signs • Rocket exhaust and ion thrusters • The area in front of a spacecraft's heat shield during reentry	• Lightning • Ball lightning • St. Elmo's fire • Upper-atmospheric lightning • The ionosphere • The polar aurorae • Most flames	• The Sun and other stars (plasmas heated by nuclear fusion) • The solar wind • The interplanetary medium (space between planets) • The interstellar medium (space between star systems) • The Intergalactic medium (space between galaxies) • The Io-Jupiter flux tube • Accretion discs

into the atmosphere
- Inside a corona discharge ozone generator
- Fusion energy research
- The electric arc in an arc lamp, an arc welder or plasma torch
- Plasma ball (sometimes called a plasma sphere or plasma globe)
- Arcs produced by Tesla coils (resonant air core transformer or disruptor coil that produces arcs similar to lightning but with alternating current rather than static electricity)
- Plasmas used in semiconductor device fabrication including reactive-ion etching, sputtering, surface cleaning and plasma-enhanced chemical vapor deposition
- Laser-produced plasmas (LPP), found when high power lasers interact with materials.
- Inductively coupled plasmas (ICP), formed typically in argon gas for optical emission spectroscopy or mass
- Interstellar nebulae

spectrometry
- Magnetically induced plasmas (MIP), typically produced using microwaves as a resonant coupling method

Plasma properties and parameters



Artist's rendition of the Earth's "plasma fountain", showing oxygen, helium, and hydrogen ions that gush into space from regions near the Earth's poles. The faint yellow area shown above the north pole represents gas lost from Earth into space; the green area is the aurora borealis, where plasma energy pours back into the atmosphere.

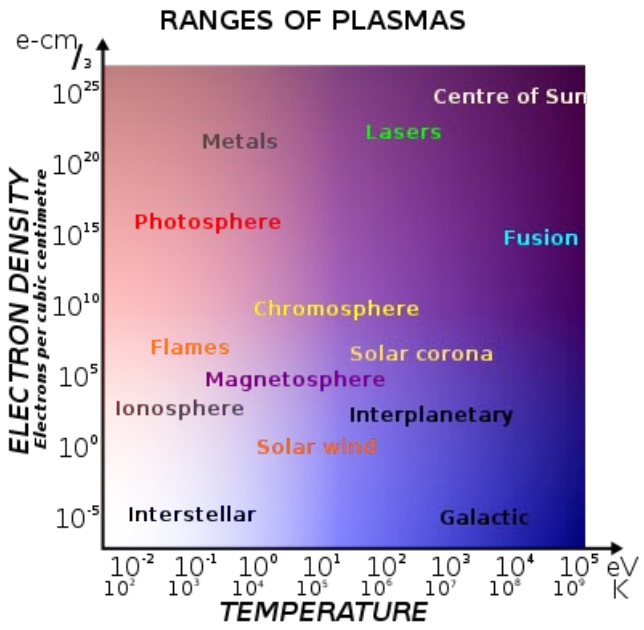
Definition of a plasma

Plasma is loosely described as an electrically neutral medium of positive and negative particles (i.e. the overall charge of a plasma is roughly zero). It is important to note that although they are unbound, these particles are not 'free'. When the charges move they generate electrical currents with magnetic fields, and as a result, they are affected by each other's fields. This governs their collective behavior with many degrees of freedom. A definition can have three criteria:

1. **The plasma approximation**: Charged particles must be close enough together that each particle influences many nearby charged particles, rather than just interacting with the closest particle (these collective effects are a distinguishing feature of a plasma). The plasma approximation is valid when the number of charge carriers within the sphere of influence (called the *Debye sphere* whose radius is the Debye screening length) of a particular particle is higher than unity to provide collective behavior of the charged particles. The average number of particles in the Debye sphere is given by the plasma parameter, "Λ" (the Greek letter Lambda).
2. **Bulk interactions**: The Debye screening length (defined above) is short compared to the physical size of the plasma. This criterion means that interactions in the bulk of the plasma are more important than those at its edges, where boundary effects may take place. When this criterion is satisfied, the plasma is quasineutral.
3. **Plasma frequency**: The electron plasma frequency (measuring plasma oscillations of the electrons) is large compared to the electron-neutral collision frequency (measuring frequency of collisions between electrons and neutral particles). When this condition is valid, electrostatic interactions dominate over the processes of ordinary gas kinetics.

Ranges of plasma parameters

Plasma parameters can take on values varying by many orders of magnitude, but the properties of plasmas with apparently disparate parameters may be very similar. The following chart considers only conventional atomic plasmas and not exotic phenomena like quark gluon plasmas:



Range of plasmas. Density increases upwards, temperature increases towards the right. The free electrons in a metal may be considered an electron plasma.

Typical ranges of plasma parameters: orders of magnitude

Characteristic	Terrestrial plasmas	Cosmic plasmas
Size in meters	10^{-6} m (lab plasmas) to 10^{2} m (lightning) (~8 OOM)	10^{-6} m (spacecraft sheath) to 10^{25} m (intergalactic nebula) (~31 OOM)
Lifetime in seconds	10^{-12} s (laser-produced plasma) to 10^{7} s (fluorescent lights) (~19 OOM)	10^{1} s (solar flares) to 10^{17} s (intergalactic plasma) (~16 OOM)
Density in particles per cubic meter	10^{7} m^{-3} to 10^{32} m^{-3} (inertial confinement plasma)	1 m^{-3} (intergalactic medium) to 10^{30} m^{-3} (stellar core)
Temperature in kelvins	~0 K (crystalline non-neutral plasma) to 10^{8} K (magnetic fusion plasma)	10^{2} K (aurora) to 10^{7} K (solar core)
Magnetic fields in teslas	10^{-4} T (lab plasma) to 10^{3} T (pulsed-power plasma)	10^{-12} T (intergalactic medium) to 10^{11} T (near neutron stars)

Degree of ionization

For plasma to exist, ionization is necessary. The term "plasma density" by itself usually refers to the "electron density", that is, the number of free electrons per unit volume. The

degree of ionization of a plasma is the proportion of atoms that have lost (or gained) electrons, and is controlled mostly by the temperature. Even a partially ionized gas in which as little as 1% of the particles are ionized can have the characteristics of a plasma (i.e., response to magnetic fields and high electrical conductivity). The degree of ionization, α is defined as $\alpha = n_i/(n_i + n_a)$ where n_i is the number density of ions and n_a is the number density of neutral atoms. The *electron density* is related to this by the average charge state <Z> of the ions through $n_e = \langle Z \rangle n_i$ where n_e is the number density of electrons.

Temperatures

Plasma temperature is commonly measured in kelvins or electronvolts and is, informally, a measure of the thermal kinetic energy per particle. Very high temperatures are usually needed to sustain ionization, which is a defining feature of a plasma. The degree of plasma ionization is determined by the "electron temperature" relative to the ionization energy, (and more weakly by the density), in a relationship called the Saha equation. At low temperatures, ions and electrons tend to recombine into bound states—atoms, and the plasma will eventually become a gas.

In most cases the electrons are close enough to thermal equilibrium that their temperature is relatively well-defined, even when there is a significant deviation from a Maxwellian energy distribution function, for example, due to UV radiation, energetic particles, or strong electric fields. Because of the large difference in mass, the electrons come to thermodynamic equilibrium amongst themselves much faster than they come into equilibrium with the ions or neutral atoms. For this reason, the "ion temperature" may be very different from (usually lower than) the "electron temperature". This is especially common in weakly ionized technological plasmas, where the ions are often near the ambient temperature.

Based on the relative temperatures of the electrons, ions and neutrals, plasmas are classified as "thermal" or "non-thermal". Thermal plasmas have electrons and the heavy particles at the same temperature, i.e., they are in thermal equilibrium with each other. Non-thermal plasmas on the other hand have the ions and neutrals at a much lower temperature, (normally room temperature), whereas electrons are much "hotter".

A plasma is sometimes referred to as being "hot" if it is nearly fully ionized, or "cold" if only a small fraction (for example 1%), of the gas molecules are ionized, but other definitions of the terms "hot plasma" and "cold plasma" are common. Even in a "cold" plasma, the electron temperature is still typically several thousand degrees Celsius. Plasmas utilized in "plasma technology" ("technological plasmas") are usually cold in this sense.

Potentials

Lightning is an example of plasma present at Earth's surface. Typically, lightning discharges 30,000 amperes at up to 100 million volts, and emits light, radio waves, X-rays and even gamma rays. Plasma temperatures in lightning can approach ~28,000 kelvin and electron densities may exceed 10^{24} m^{-3}.

Since plasmas are very good conductors, electric potentials play an important role. The potential as it exists on average in the space between charged particles, independent of the question of how it can be measured, is called the "plasma potential", or the "space potential". If an electrode is inserted into a plasma, its potential will generally lie considerably below the plasma potential due to what is termed a Debye sheath. The good electrical conductivity of plasmas makes their electric fields very small. This results in the important concept of "quasineutrality", which says the density of negative charges is

approximately equal to the density of positive charges over large volumes of the plasma ($n_e = \langle Z \rangle n_i$), but on the scale of the Debye length there can be charge imbalance. In the special case that *double layers* are formed, the charge separation can extend some tens of Debye lengths.

The magnitude of the potentials and electric fields must be determined by means other than simply finding the net charge density. A common example is to assume that the electrons satisfy the "Boltzmann relation":

$$n_e \propto e^{e\Phi/k_B T_e}.$$

Differentiating this relation provides a means to calculate the electric field from the density:

$$\vec{E} = (k_B T_e/e)(\nabla n_e/n_e).$$

It is possible to produce a plasma that is not quasineutral. An electron beam, for example, has only negative charges. The density of a non-neutral plasma must generally be very low, or it must be very small, otherwise it will be dissipated by the repulsive electrostatic force.

In astrophysical plasmas, Debye screening prevents electric fields from directly affecting the plasma over large distances, i.e., greater than the Debye length. But the existence of charged particles causes the plasma to generate and can be affected by magnetic fields. This can and does cause extremely complex behavior, such as the generation of plasma double layers, an object that separates charge over a few tens of Debye lengths. The dynamics of plasmas interacting with external and self-generated magnetic fields are studied in the academic discipline of magnetohydrodynamics.

Magnetization

Plasma with a magnetic field strong enough to influence the motion of the charged particles is said to be magnetized. A common quantitative criterion is that a particle on average completes at least one gyration around the magnetic field before making a collision, i.e., $\omega_{ce}/\nu_{coll} > 1$, where ω_{ce} is the "electron gyrofrequency" and ν_{coll} is the "electron collision rate". It is often the case that the electrons are magnetized while the ions are not. Magnetized plasmas are *anisotropic*, meaning that their properties in the direction parallel to the magnetic field are different from those perpendicular to it. While electric fields in plasmas are usually small due to the high conductivity, the electric field associated with a plasma moving in a magnetic field is given by $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$ (where $\mathbf{E}$ is the electric field, $\mathbf{v}$ is the velocity, and $\mathbf{B}$ is the magnetic field), and is not affected by Debye shielding.

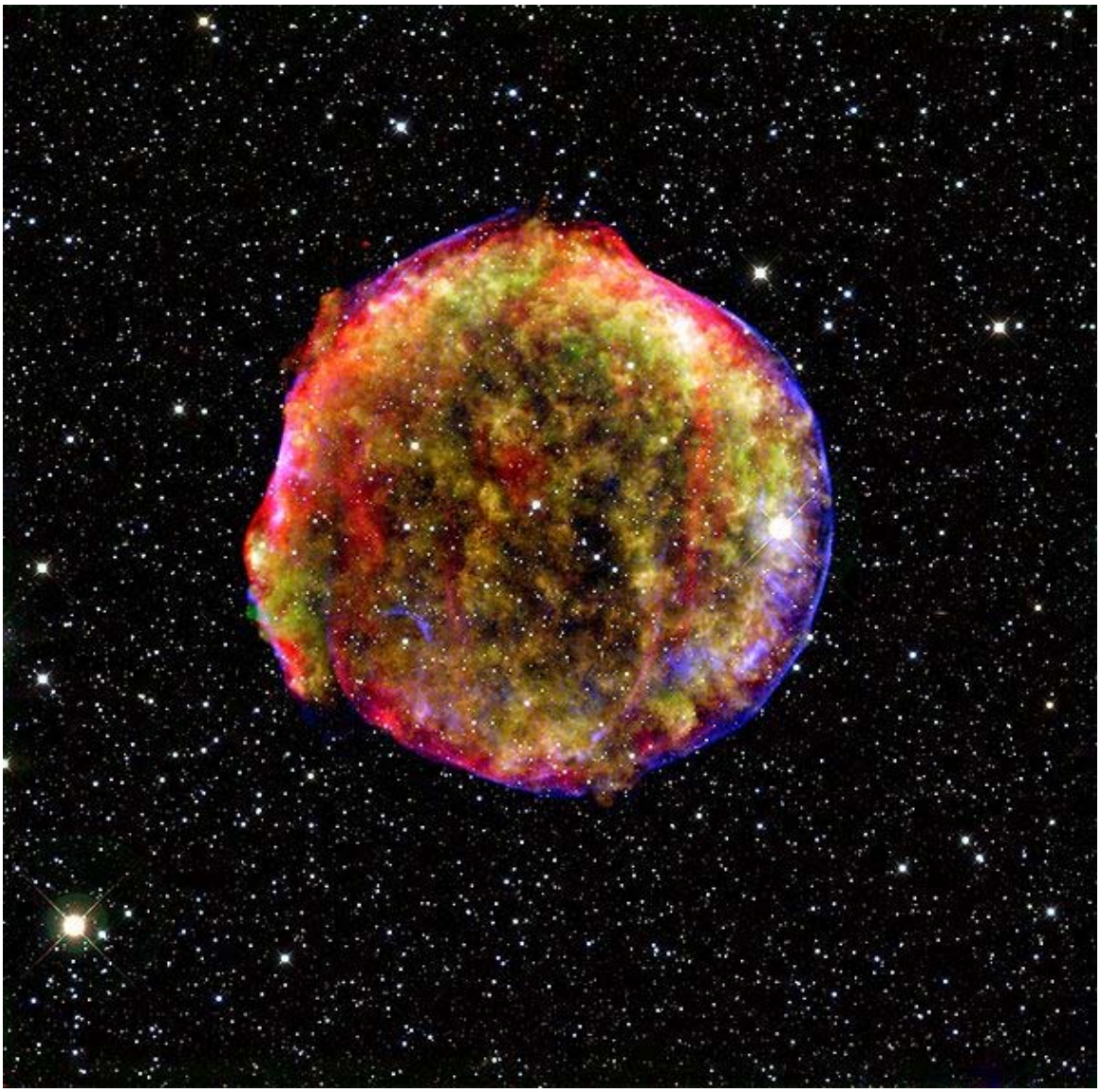
Comparison of plasma and gas phases

Plasma is often called the *fourth state of matter*. It is distinct from other lower-energy states of matter; most commonly solid, liquid, and gas. Although it is closely related to the gas phase in that it also has no definite form or volume, it differs in a number of ways, including the following:

Property	Gas	Plasma
Electrical Conductivity	**Very low** Air is an excellent insulator until it breaks down into plasma at electric field strengths above 30 kilovolts per centimeter.	**Usually very high** For many purposes, the conductivity of a plasma may be treated as infinite.
Independently acting species	**One** All gas particles behave in a similar way, influenced by gravity and by collisions with one another.	**Two or three** Electrons, ions, protons and neutrons can be distinguished by the sign and value of their charge so that they behave independently in many circumstances, with different bulk velocities and temperatures, allowing phenomena such as new types of waves and instabilities.
Velocity distribution	**Maxwellian** Collisions usually lead to a Maxwellian velocity distribution of all gas particles, with very few relatively fast particles.	**Often non-Maxwellian** Collisional interactions are often weak in hot plasmas and external forcing can drive the plasma far from local equilibrium and lead to a significant population of unusually fast particles.
Interactions	**Binary** Two-particle collisions are the rule, three-body collisions extremely rare.	**Collective** Waves, or organized motion of plasma, are very important because the particles can interact at long ranges through the electric and

magnetic forces.

Complex plasma phenomena



The remnant of "Tycho's Supernova", a huge ball of expanding plasma. The outer shell shown in blue is X-ray emission by high-speed electrons.

Although the underlying equations governing plasmas are relatively simple, plasma behavior is extraordinarily varied and subtle: the emergence of unexpected behavior from a simple model is a typical feature of a complex system. Such systems lie in some sense

on the boundary between ordered and disordered behavior and cannot typically be described either by simple, smooth, mathematical functions, or by pure randomness. The spontaneous formation of interesting spatial features on a wide range of length scales is one manifestation of plasma complexity. The features are interesting, for example, because they are very sharp, spatially intermittent (the distance between features is much larger than the features themselves), or have a fractal form. Many of these features were first studied in the laboratory, and have subsequently been recognized throughout the universe. Examples of complexity and complex structures in plasmas include:

Filamentation

Striations or string-like structures are seen in many plasmas, like the plasma ball, the aurora, lightning, electric arcs, solar flares, and supernova remnants. They are sometimes associated with larger current densities, and the interaction with the magnetic field can form a magnetic rope structure. High power microwave breakdown at atmospheric pressure also leads to the formation of filamentary structures.

Filamentation also refers to the self-focusing of a high power laser pulse. At high powers, the nonlinear part of the index of refraction becomes important and causes a higher index of refraction in the center of the laser beam, where the laser is brighter than at the edges, causing a feedback that focuses the laser even more. The tighter focused laser has a higher peak brightness (irradiance) that forms a plasma. The plasma has an index of refraction lower than one, and causes a defocusing of the laser beam. The interplay of the focusing index of refraction, and the defocusing plasma makes the formation of a long filament of plasma that can be micrometers to kilometers in length.

Shocks or double layers

Plasma properties change rapidly (within a few Debye lengths) across a two-dimensional sheet in the presence of a (moving) shock or (stationary) double layer. Double layers involve localized charge separation, which causes a large potential difference across the layer, but does not generate an electric field outside the layer. Double layers separate adjacent plasma regions with different physical characteristics, and are often found in current carrying plasmas. They accelerate both ions and electrons.

Electric fields and circuits

Quasineutrality of a plasma requires that plasma currents close on themselves in electric circuits. Such circuits follow Kirchhoff's circuit laws and possess a resistance and inductance. These circuits must generally be treated as a strongly coupled system, with the behavior in each plasma region dependent on the entire circuit. It is this strong coupling between system elements, together with nonlinearity, which may lead to complex behavior. Electrical circuits in plasmas store inductive (magnetic) energy, and should the circuit be disrupted, for example, by a plasma instability, the inductive energy will be released as plasma heating and acceleration. This is a common explanation for the heating that takes place in the solar corona. Electric currents, and in particular, magnetic-

field-aligned electric currents (which are sometimes generically referred to as "Birkeland currents"), are also observed in the Earth's aurora, and in plasma filaments.

Cellular structure

Narrow sheets with sharp gradients may separate regions with different properties such as magnetization, density and temperature, resulting in cell-like regions. Examples include the magnetosphere, heliosphere, and heliospheric current sheet. Hannes Alfvén wrote: "From the cosmological point of view, the most important new space research discovery is probably the cellular structure of space. As has been seen in every region of space accessible to in situ measurements, there are a number of 'cell walls', sheets of electric currents, which divide space into compartments with different magnetization, temperature, density, etc."

Critical ionization velocity

The critical ionization velocity is the relative velocity between an ionized plasma and a neutral gas, above which a runaway ionization process takes place. The critical ionization process is a quite general mechanism for the conversion of the kinetic energy of a rapidly streaming gas into ionization and plasma thermal energy. Critical phenomena in general are typical of complex systems, and may lead to sharp spatial or temporal features.

Ultracold plasma

Ultracold plasmas are created in a magneto-optical trap (MOT) by trapping and cooling neutral atoms, to temperatures of 1 mK or lower, and then using another laser to ionize the atoms by giving each of the outermost electrons just enough energy to escape the electrical attraction of its parent ion.

One advantage of ultracold plasmas is their well characterized and tunable initial conditions, including size and electron temperature. By adjusting the wavelength of the ionizing laser, the kinetic energy of the liberated electrons can be tuned as low as 0.1 K, a limit set by the frequency bandwidth of the laser pulse. The ions, inherit the millikelvin temperatures of the neutral atoms, but are quickly heated through a process known as disorder induced heating (DIH). This type of non-equilibrium ultracold plasma evolves rapidly, and displays many other interesting phenomena.

One of the metastable states of a strongly nonideal plasma is Rydberg matter, which forms upon condensation of excited atoms.

Non-neutral plasma

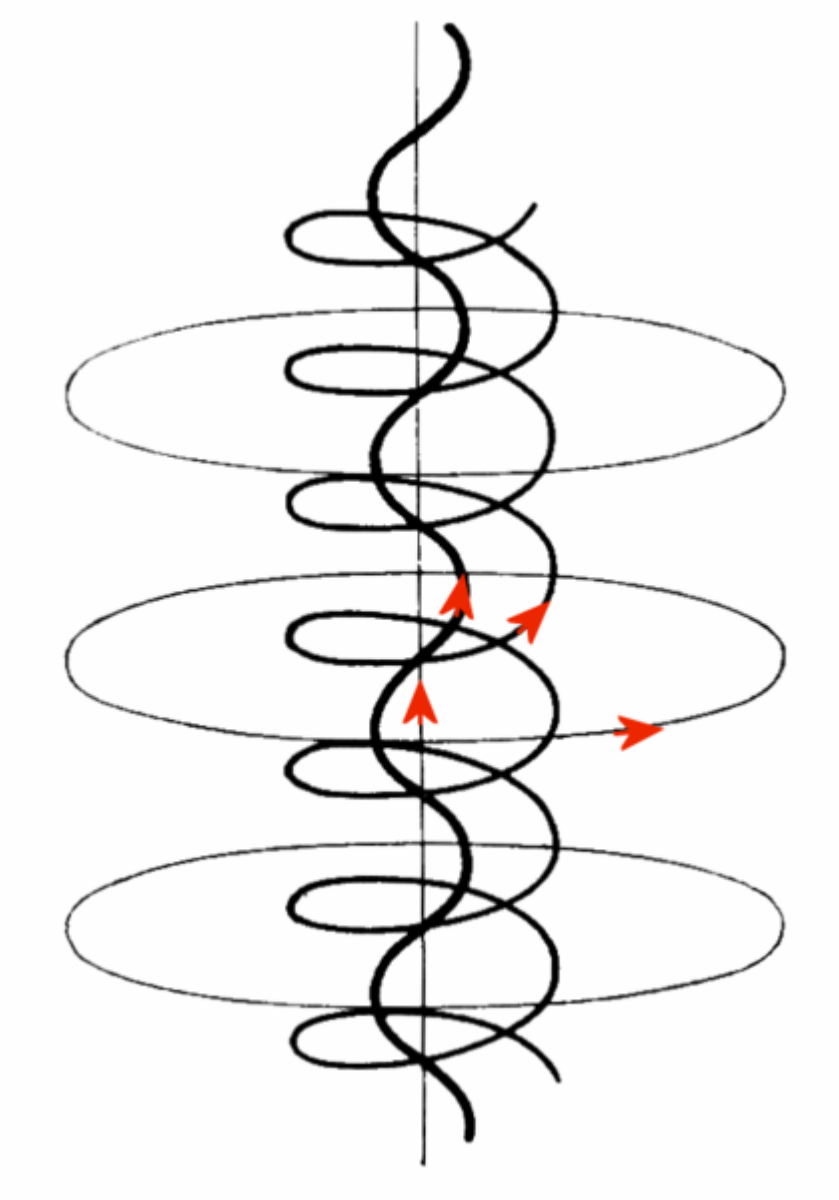
The strength and range of the electric force and the good conductivity of plasmas usually ensure that the densities of positive and negative charges in any sizeable region are equal ("quasineutrality"). A plasma with a significant excess of charge density, or, in the extreme case, is composed of only a single species, is called a non-neutral plasma. In

such a plasma, electric fields play a dominant role. Examples are charged particle beams, an electron cloud in a Penning trap and positron plasmas.

Dusty plasma and grain plasma

A dusty plasma contains tiny charged particles of dust (typically found in space), which also behaves like a plasma. A plasma that contains larger particles is called grain plasma.

Mathematical descriptions



The complex self-constricting magnetic field lines and current paths in a field-aligned Birkeland current that can develop in a plasma.

To completely describe the state of a plasma, we would need to write down all the particle locations and velocities and describe the electromagnetic field in the plasma region. However, it is generally not practical or necessary to keep track of all the particles in a plasma. Therefore, plasma physicists commonly use less detailed descriptions, of which there are two main types:

Fluid model

Fluid models describe plasmas in terms of smoothed quantities, like density and averaged velocity around each position. One simple fluid model, magnetohydrodynamics, treats the plasma as a single fluid governed by a combination of Maxwell's equations and the Navier–Stokes equations. A more general description is the two-fluid plasma picture, where the ions and electrons are described separately. Fluid models are often accurate when collisionality is sufficiently high to keep the plasma velocity distribution close to a Maxwell–Boltzmann distribution. Because fluid models usually describe the plasma in terms of a single flow at a certain temperature at each spatial location, they can neither capture velocity space structures like beams or double layers, nor resolve wave-particle effects.

Kinetic model

Kinetic models describe the particle velocity distribution function at each point in the plasma and therefore do not need to assume a Maxwell–Boltzmann distribution. A kinetic description is often necessary for collisionless plasmas. There are two common approaches to kinetic description of a plasma. One is based on representing the smoothed distribution function on a grid in velocity and position. The other, known as the particle-in-cell (PIC) technique, includes kinetic information by following the trajectories of a large number of individual particles. Kinetic models are generally more computationally intensive than fluid models. The Vlasov equation may be used to describe the dynamics of a system of charged particles interacting with an electromagnetic field. In magnetized plasmas, a gyrokinetic approach can substantially reduce the computational expense of a fully kinetic simulation.

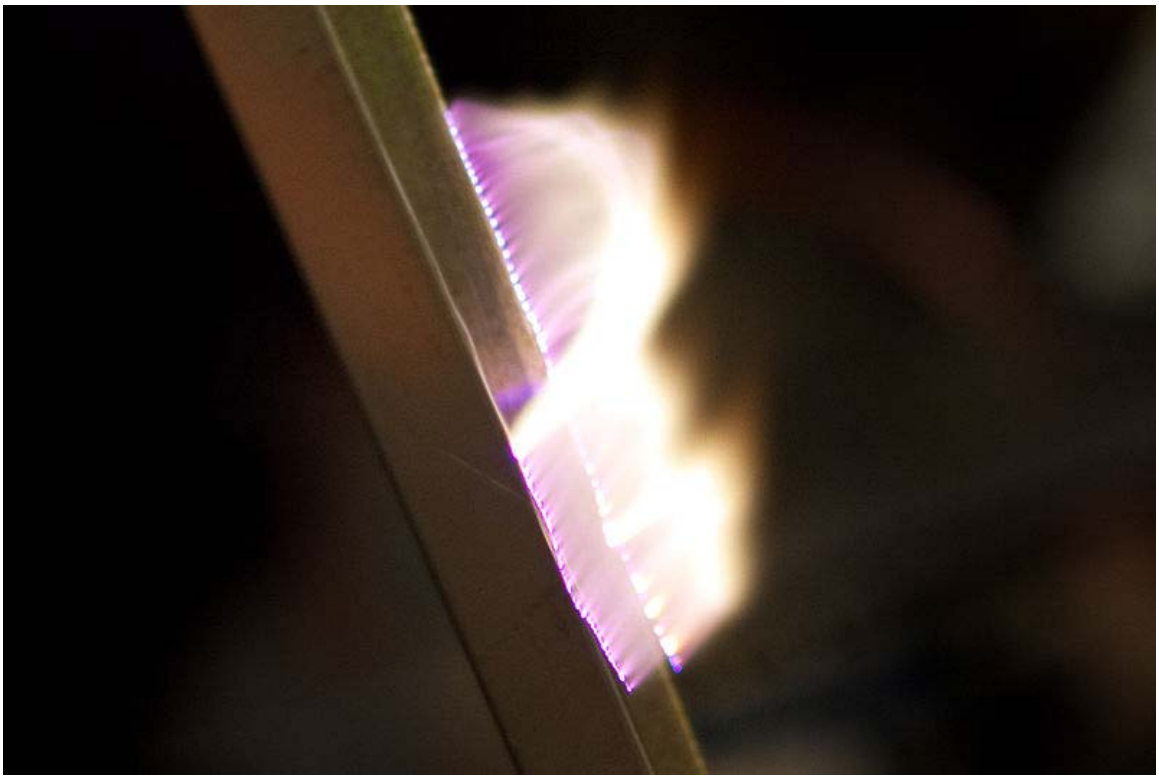
Artificial plasmas

Most artificial plasmas are generated by the application of electric and/or magnetic fields. Plasma generated in a laboratory setting and for industrial use can be generally categorized by:

- The type of power source used to generate the plasma—DC, RF and microwave
- The pressure they operate at—vacuum pressure (< 10 mTorr or 1 Pa), moderate pressure (~ 1 Torr or 100 Pa), atmospheric pressure (760 Torr or 100 kPa)
- The degree of ionization within the plasma—fully, partially, or weakly ionized
- The temperature relationships within the plasma—thermal plasma ($T_e = T_{\text{ion}} = T_{\text{gas}}$), non-thermal or "cold" plasma ($T_e >> T_{\text{ion}} = T_{\text{gas}}$)
- The electrode configuration used to generate the plasma

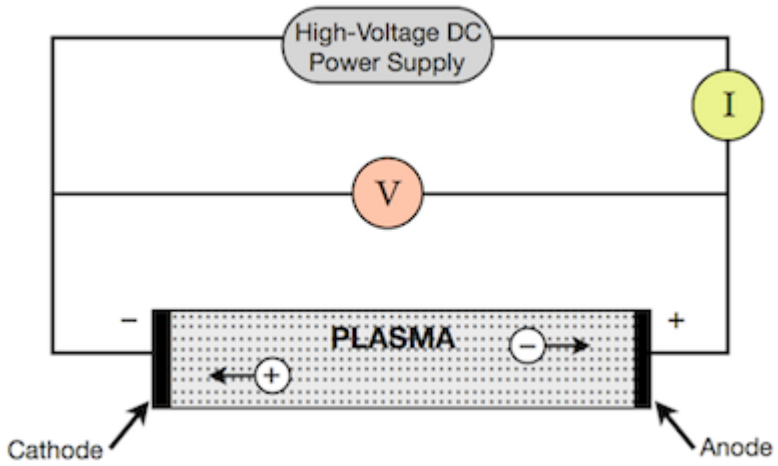
- The magnetization of the particles within the plasma—magnetized (both ion and electrons are trapped in Larmor orbits by the magnetic field), partially magnetized (the electrons but not the ions are trapped by the magnetic field), non-magnetized (the magnetic field is too weak to trap the particles in orbits but may generate Lorentz forces)
- The application

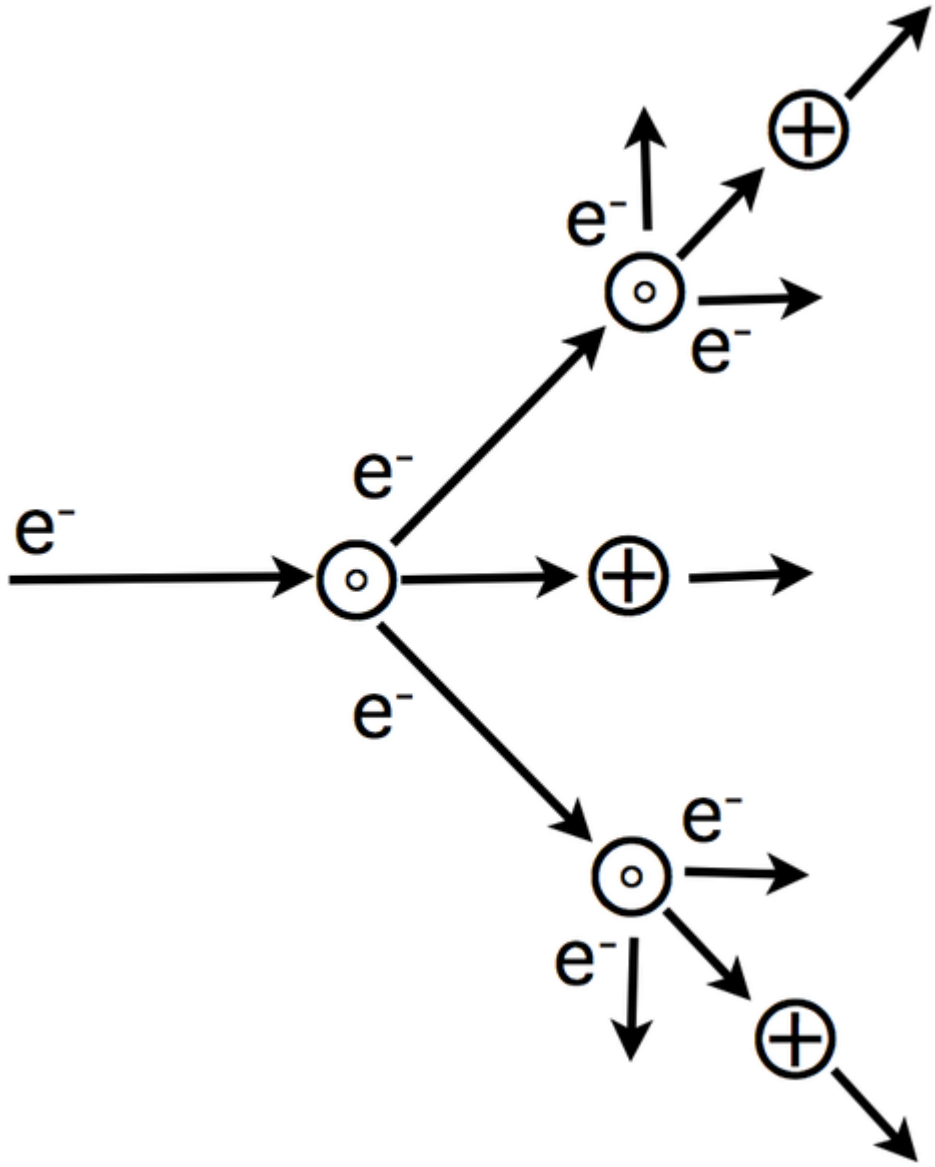
Generation of artificial plasma



Artificial plasma produced in air by a Jacob's Ladder

Just like the many uses of plasma, there are several means for its generation, however, one principle is common to all of them: there must be energy input to produce and sustain it. For this case, plasma is generated when an electrical current is applied across a dielectric gas or fluid (an electrically non-conducting material) as can be seen in the image below, which shows a discharge tube as a simple example (DC used for simplicity).


High-Voltage DC
Power Supply
I
V
–
PLASMA
+
Cathode
Anode



Cascade process of ionization. Electrons are 'e−', neutral atoms 'o', and cations '+'.

The potential difference and subsequent electric field pulls the bound electrons (negative) toward the anode (positive electrode) while the (positive) cathode (negative electrode) pulls the nucleus. As the voltage increases, the current stresses the material (by electric polarization) beyond its dielectric limit (termed strength) into a stage of electrical breakdown, marked by an electric spark, where the material transforms from being an insulator into a conductor (as it becomes increasingly ionized). This is a stage of avalanching ionization, where collisions between electrons and neutral gas atoms, create more ions and electrons (as can be seen in the figure on the right). The first impact of an electron on an atom results in one ion and two electrons. Therefore, the number of charged particles increases rapidly (in the millions) only "after about 20 successive sets

of collisions", mainly due to a small mean free path (average distance travelled between collisions).

With ample current density and ionization, this forms a luminous electric arc (essentially lightning) between the electrodes. Electrical resistance along the continuous electric arc creates heat, which ionizes more gas molecules (where degree of ionization is determined by temperature), and as per the sequence: solid-liquid-gas-plasma, the gas is gradually turned into a thermal plasma. A thermal plasma is in thermal equilibrium, which is to say that the temperature is relatively homogeneous throughout the heavy particles (i.e. atoms, molecules and ions) and electrons. This is so because when thermal plasmas are generated, electrical energy is given to electrons, which, due to their great mobility and large numbers, are able to disperse it rapidly and by elastic collision (without energy loss) to the heavy particles.

Examples of industrial/commercial plasma

Because of their sizable temperature and density ranges, plasmas find applications in many fields of research, technology and industry. For example, in: industrial and extractive metallurgy, surface treatments such as thermal spraying (coating), etching in microelectronics, metal cutting and welding; as well as in everyday vehicle exhaust cleanup and fluorescent/luminescent lamps, while even playing a part in supersonic combustion engines for aerospace engineering.

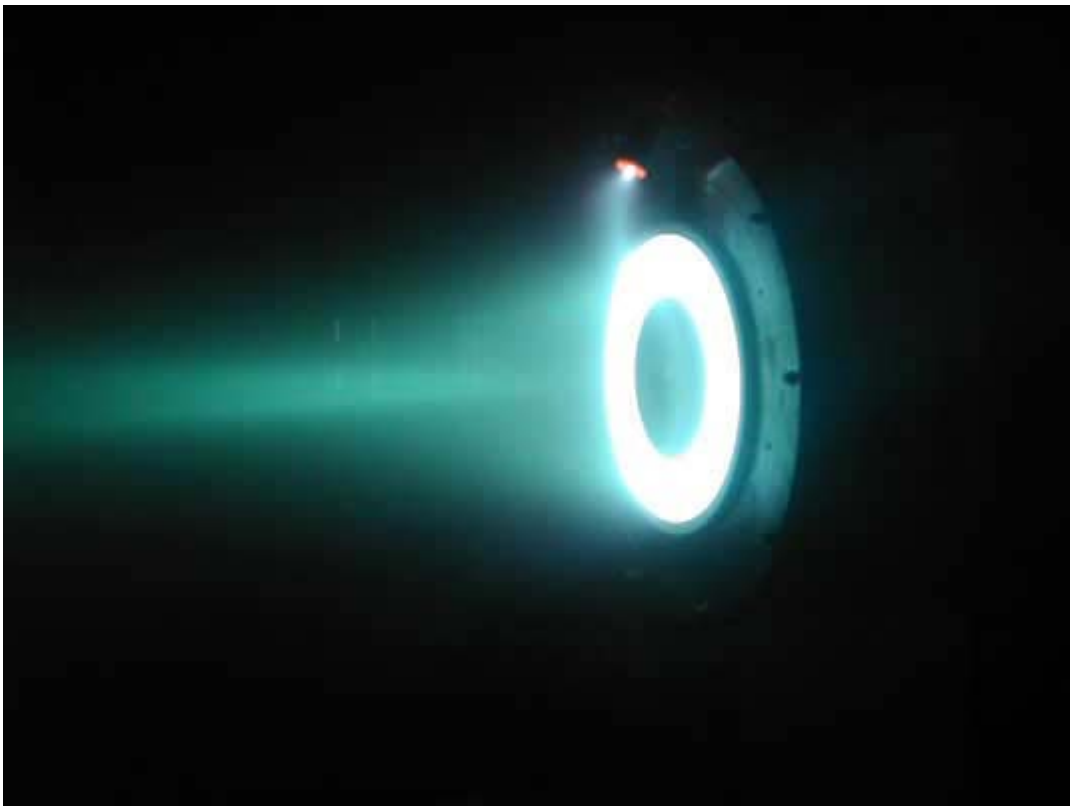
Low-pressure discharges

- *Glow discharge plasmas*: non-thermal plasmas generated by the application of DC or low frequency RF (<100 kHz) electric field to the gap between two metal electrodes. Probably the most common plasma; this is the type of plasma generated within fluorescent light tubes.
- *Capacitively coupled plasma (CCP)*: similar to glow discharge plasmas, but generated with high frequency RF electric fields, typically 13.56 MHz. These differ from glow discharges in that the sheaths are much less intense. These are widely used in the microfabrication and integrated circuit manufacturing industries for plasma etching and plasma enhanced chemical vapor deposition.
- *Inductively coupled plasma (ICP)*: similar to a CCP and with similar applications but the electrode consists of a coil wrapped around the discharge volume that inductively excites the plasma.
- *Wave heated plasma*: similar to CCP and ICP in that it is typically RF (or microwave), but is heated by both electrostatic and electromagnetic means. Examples are helicon discharge, electron cyclotron resonance (ECR), and ion cyclotron resonance (ICR). These typically require a coaxial magnetic field for wave propagation.

Atmospheric pressure

- *Arc discharge:* this is a high power thermal discharge of very high temperature (~10,000 K). It can be generated using various power supplies. It is commonly used in metallurgical processes. For example, it is used to melt rocks containing Al_2O_3 to produce aluminium.
- *Corona discharge:* this is a non-thermal discharge generated by the application of high voltage to sharp electrode tips. It is commonly used in ozone generators and particle precipitators.
- *Dielectric barrier discharge (DBD):* this is a non-thermal discharge generated by the application of high voltages across small gaps wherein a non-conducting coating prevents the transition of the plasma discharge into an arc. It is often mislabeled 'Corona' discharge in industry and has similar application to corona discharges. It is also widely used in the web treatment of fabrics. The application of the discharge to synthetic fabrics and plastics functionalizes the surface and allows for paints, glues and similar materials to adhere.
- *Capacitive discharge:* this is a nonthermal plasma generated by the application of RF power (e.g., 13.56 MHz) to one powered electrode, with a grounded electrode held at a small separation distance on the order of 1 cm. Such discharges are commonly stabilized using a noble gas such as helium or argon.

Fields of active research



Hall effect thruster. The electric field in a plasma double layer is so effective at accelerating ions that electric fields are used in ion drives.

This is just a partial list of topics. A more complete and organized list can be found on the web site Plasma science and technology.

- Plasma theory
  - Plasma equilibria and stability
  - Plasma interactions with waves and beams
  - Guiding center
  - Adiabatic invariant
  - Debye sheath
  - Coulomb collision
- Plasmas in nature
  - The Earth's ionosphere
  - Northern and southern (polar) lights
  - Space plasmas, e.g. Earth's plasmasphere (an inner portion of the magnetosphere dense with plasma)
  - Astrophysical plasma
- Industrial plasmas
  - Plasma chemistry
  - Plasma processing
  - Plasma spray
  - Plasma display
- Plasma sources
- Dusty plasmas
- Plasma diagnostics
  - Thomson scattering
  - Langmuir probe
  - Spectroscopy
  - Interferometry
  - Ionospheric heating
  - Incoherent scatter radar
- Plasma applications
  - Fusion power
    - Magnetic fusion energy (MFE) — tokamak, stellarator, reversed field pinch, magnetic mirror, dense plasma focus
    - Inertial fusion energy (IFE) (also Inertial confinement fusion — ICF)
    - Plasma-based weaponry
  - Ion implantation
  - Ion thruster
  - Plasma ashing
  - Food processing (nonthermal plasma, aka "cold plasma")
  - Plasma arc waste disposal, convert waste into reusable material with plasma.
  - Plasma acceleration
  - Plasma medicine (e. g. Dentistry)
  - Plasma window