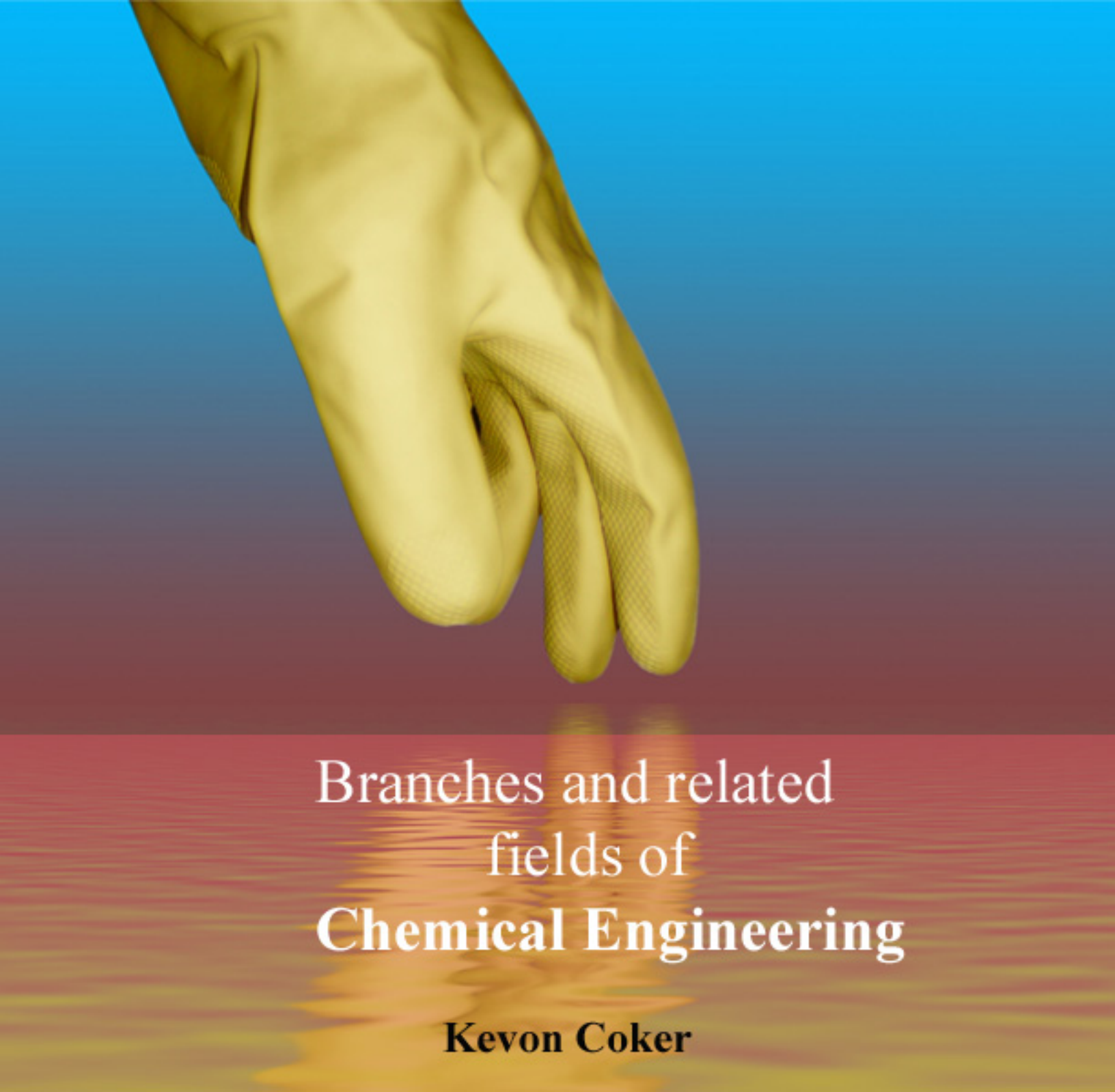

Branches and related
fields of
Chemical Engineering
Kevon Coker

First Edition, 2012

ISBN 978-81-323-3030-1

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

Table of Contents

Chapter 1 - Chemical Reactor

Chapter 2 - Process Design (Chemical Engineering) and Separation Process

Chapter 3 - Crystallization

Chapter 4 - Thermodynamics

Chapter 5 - Electrochemistry

Chapter 6 - Distillation

Chapter 1

Chemical Reactor

In chemical engineering, **chemical reactors** are vessels designed to contain chemical reactions. The design of a chemical reactor deals with multiple aspects of chemical engineering. Chemical engineers design reactors to maximize net present value for the given reaction. Designers ensure that the reaction proceeds with the highest efficiency towards the desired output product, producing the highest yield of product while requiring the least amount of money to purchase and operate. Normal operating expenses include energy input, energy removal, raw material costs, labor, etc. Energy changes can come in the form of heating or cooling, pumping to increase pressure, frictional pressure loss (such as pressure drop across a 90^{o} elbow or an orifice plate), agitation, etc.

Chemical reaction engineering is the branch of chemical engineering which deals with chemical reactors and their design, especially by application of chemical kinetics.

Overview

Cut-away view of a stirred-tank chemical reactor with a cooling jacket

Chemical reactor with half coils wrapped around it

There are a couple main basic vessel types:

- A tank
- A pipe or tubular reactor

Both types can be used as continuous reactors or batch reactors. Most commonly, reactors are run at steady-state, but can also be operated in a transient state. When a reactor is first brought back into operation (after maintenance or inoperation) it would be considered to be in a transient state, where key process variables change with time. Both types of reactors may also accommodate one or more solids (reagents, catalyst, or inert materials), but the reagents and products are typically liquids and gases.

There are three main basic models used to estimate the most important process variables of different chemical reactors:

- *batch reactor* model (batch),
- *continuous stirred-tank reactor* model (CSTR), and
- *plug flow reactor* model (PFR).

Furthermore, catalytic reactors require separate treatment, whether they are batch, CST, or PF reactors, as the many assumptions of the simpler models are not valid.

Key process variables include

- Residence time (τ, lower case Greek tau)
- Volume (V)
- Temperature (T)
- Pressure (P)
- Concentrations of chemical species (C_1, C_2, C_3, ... C_n)
- Heat transfer coefficients (h, U)

A chemical reactor, typically tubular reactor, could be a packed bed. The packing inside the bed may have catalyst to catalyze the chemical reaction.

Chemical reactions occurring in a reactor may be exothermic, meaning giving off heat, or endothermic, meaning absorbing heat. A chemical reactor vessel may have a cooling or heating jacket or cooling or heating coils (tubes) wrapped around the outside of its vessel wall to cool down or heat up the contents.

Types

CSTR (Continuous Stirred-Tank Reactor)

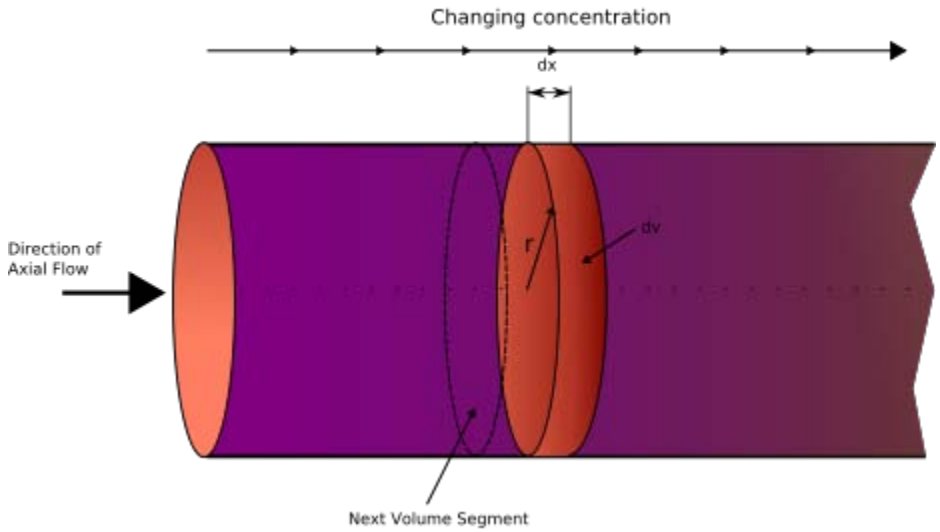
Reactor CSTR. Check of a condition of the case. Note the impeller (or agitator) blades on the shaft for mixing. Also note the baffle at the bottom of the image which also helps in mixing.

In a CSTR, one or more fluid reagents are introduced into a tank reactor equipped with an impeller while the reactor effluent is removed. The impeller stirs the reagents to ensure proper mixing. Simply dividing the volume of the tank by the average volumetric flow rate through the tank gives the *residence time*, or the average amount of time a discrete quantity of reagent spends inside the tank. Using chemical kinetics, the reaction's expected percent completion can be calculated. Some important aspects of the CSTR:

- At steady-state, the flow rate in must equal the mass flow rate out, otherwise the tank will overflow or go empty (transient state). While the reactor is in a transient state the model equation must be derived from the differential mass and energy balances.
- The reaction proceeds at the reaction rate associated with the final (output) concentration.
- Often, it is economically beneficial to operate several CSTRs in series. This allows, for example, the first CSTR to operate at a higher reagent concentration and therefore a higher reaction rate. In these cases, the sizes of the reactors may be varied in order to minimize the total capital investment required to implement the process.
- It can be seen that an infinite number of infinitely small CSTRs operating in series would be equivalent to a PFR.

The behavior of a CSTR is often approximated or modeled by that of a Continuous Ideally Stirred-Tank Reactor (CISTR). All calculations performed with CISTRs assume perfect mixing. If the residence time is 5-10 times the mixing time, this approximation is valid for engineering purposes. The CISTR model is often used to simplify engineering calculations and can be used to describe research reactors. In practice it can only be approached, in particular in industrial size reactors.

PFR (Plug Flow Reactor)



Simple diagram illustrating plug flow reactor model

In a PFR, one or more fluid reagents are pumped through a pipe or tube. The chemical reaction proceeds as the reagents travel through the PFR. In this type of reactor, the changing reaction rate creates a gradient with respect to distance traversed; at the inlet to the PFR the rate is very high, but as the concentrations of the reagents decrease and the concentration of the product(s) increases the reaction rate slows. Some important aspects of the PFR:

- All calculations performed with PFRs assume no upstream or downstream mixing, as implied by the term "plug flow".
- Reagents may be introduced into the PFR at locations in the reactor other than the inlet. In this way, a higher efficiency may be obtained, or the size and cost of the PFR may be reduced.
- A PFR typically has a higher efficiency than a CSTR of the same volume. That is, given the same space-time, a reaction will proceed to a higher percentage completion in a PFR than in a CSTR.

For most chemical reactions, it is impossible for the reaction to proceed to 100% completion. The rate of reaction decreases as the percent completion increases until the point where the system reaches dynamic equilibrium (no net reaction, or change in chemical species occurs). The equilibrium point for most systems is less than 100% complete. For this reason a separation process, such as distillation, often follows a chemical reactor in order to separate any remaining reagents or byproducts from the desired product. These reagents may sometimes be reused at the beginning of the process, such as in the Haber process.

Continuous oscillatory baffled reactor (COBR) is a tubular plug flow reactor. The mixing in COBR is achieved by the combination of fluid oscillation and orifice baffles, allowing plug flow to be achieved under laminar flow conditions with the net flow Reynolds number just about 100.

Semi-batch reactor

A semi-batch reactor is operated with both continuous and batch inputs and outputs. A fermenter, for example, is loaded with a batch, which constantly produces carbon dioxide, which has to be removed continuously. Analogously, driving a reaction of gas with a liquid is usually difficult, since the gas bubbles off. Therefore, a continuous feed of gas is injected into the batch of a liquid. An example of such a reaction is chlorination.

Catalytic reactor

Although catalytic reactors are often implemented as plug flow reactors, their analysis requires more complicated treatment. The rate of a catalytic reaction is proportional to the amount of catalyst the reagents contact. With a solid phase catalyst and fluid phase reagents, this is proportional to the exposed area, efficiency of diffusion of reagents in and products out, and turbulent mixing or lack thereof. Perfect mixing cannot be assumed. Furthermore, a catalytic reaction pathway is often multi-step with intermediates that are chemically bound to the catalyst; and as the chemical binding to the catalyst is also a chemical reaction, it may affect the kinetics.

The behavior of the catalyst is also a consideration. Particularly in high-temperature petrochemical processes, catalysts are deactivated by sintering, coking, and similar processes.

A common example of a catalytic reactor is the catalytic converter following an engine.

Chapter 2

Process Design (Chemical Engineering) and Separation Process

Process design (chemical engineering)

Process design is the design of processes for desired physical and/or chemical transformation of materials. Process design is central to chemical engineering and it can be considered to be the summit of chemical engineering, bringing together all of the components of that field.

Process design can be the design of new facilities or it can be the modification or expansion of existing facilities. The design starts at a conceptual level and ultimately ends in the form of fabrication and construction plans.

Process design is distinct from equipment design, which is closer in spirit to the design of unit operations. Processes often include many unit operations.

Documentation

Process design documents serve to define the design and they ensure that the design components fit together. They are useful in communicating ideas and plans to other engineers involved with the design, to external regulatory agencies, to equipment vendors and to construction contractors.

In order of increasing detail, process design documents include:

- **Block Flow Diagrams (BFD)**: Very simple diagrams composed of rectangles and lines indicating major material or energy flows.
- **Process Flow Diagrams (PFD's)**: Typically more complex diagrams of major unit operations as well as flow lines. They usually include a material balance, and sometimes an energy balance, showing typical or design flowrates, stream compositions, and stream and equipment pressures and temperatures.

- **Piping and Instrumentation Diagrams (P&ID's)**: Diagrams showing each and every pipeline with piping class (carbon steel or stainless steel) and pipe size (diameter). They also show valving along with instrument locations and process control schemes.

- **Specifications**: Written design requirements of all major equipment items.

Process designers also typically write operating manuals on how to start-up, operate and shut-down the process.

Documents are maintained after construction of the process facility for the operating personnel to refer to. The documents also are useful when modifications to the facility are planned.

A primary method of developing the process documents is process flowsheeting.

Design Considerations

Designs have objectives and constraints, and even a simple process requires a trade-off among such factors.

Objectives that a design may strive to include:

- Throughput rate
- Process yield
- Product purity

Constraints include:

- Capital cost
- Available space
- Safety concerns
- Environmental impact and projected effluents and emissions
- Waste production
- Operating and maintenance costs

Other factors that designers may include are:

- Reliability
- Redundancy
- Flexibility
- Anticipated variability in feedstock and allowable variability in product.

Sources of Design Information

Designers usually do not start from scratch, especially for complex projects. Often the engineers have pilot plant data available or data from full-scale operating facilities. Other sources of information include proprietary design criteria provided by process licensors, published scientific data, laboratory experiments, and input.

Computer Help

The advent of low cost powerful computers has aided complex mathematical simulation of processes, and simulation software is often used by design engineers. Simulations can identify weaknesses in designs and allow engineers to choose better alternatives.

However, engineers still rely on heuristics, intuition, and experience when designing a process. Human creativity is an element in complex designs.

Separation process

In chemistry and chemical engineering, a **separation process** is used to transform a mixture of substances into two or more distinct products. The separated products could differ in chemical properties or some physical property, such as size, or crystal modification or other separation into different components.

Barring a few exceptions, almost every element or compound is found naturally in an impure state such as a mixture of two or more substances. Many times the need to separate it into its individual components arises. Separation applications in the field of chemical engineering are very important. A good example is that of crude oil. Crude oil is a mixture of various hydrocarbons and is valuable in this natural form. Demand is greater, however, for the purified various hydrocarbons such as natural gases, gasoline, diesel, jet fuel, lubricating oils, asphalt, etc.

Separation processes can essentially be termed as mass transfer processes. The classification can be based on the means of separation, *mechanical* or *chemical*. The choice of separation depends on the pros and cons of each. *Mechanical* separations are usually favored if possible due to the lower cost of the operations as compared to *chemical* separations. Systems that can not be separated by purely mechanical means (e.g. crude oil), chemical separation is the remaining solution. The mixture at hand could exist as a combination of any two or more states: solid-solid, solid-liquid, solid-gas, liquid-liquid, liquid-gas, gas-gas, solid-liquid-gas mixture, etc.

Depending on the raw mixture, various processes can be employed to separate the mixtures. Many times two or more of these processes have to be used in combination to obtain the desired separation. In addition to *chemical* processes, *mechanical* processes can also be applied where possible. In the example of crude oil, one upstream distillation operation will feed its two or more product streams into multiple downstream distillation operations to further separate the crude, and so on until final products are purified.

Various types of separation processes

- Adsorption.
- Centrifugation and cyclonic separation, separates based on density differences.
- Chromatography separates dissolved substances by different interaction with (i.e., travel through) a material.
- Crystallization.
- Decantation.
- Demister (vapor), removes liquid droplets from gas streams.
- Distillation, used for mixtures of liquids with different boiling points.
- Drying, removes liquid from a solid by vaporisation.
- Electrophoresis, separates organic molecules based on their different interaction with a gel under an electric potential (i.e., different travel).

- Elutriation.
- Evaporation.
- Extraction.
    - Leaching.
    - Liquid-liquid extraction.
    - Solid phase extraction.
- Flotation.
    - Dissolved air flotation, removes suspended solids non-selectively from slurry by bubbles that are generated by air coming out of solution.
    - Froth flotation, recovers valuable, hydrophobic solids by attachment to air bubbles generated by mechanical agitation of an air-slurry mixture, which float, and are recovered.
    - Deinking, separating hydrophobic ink particles from hydrophilic paper pulp in paper recycling.
- Flocculation, separates a solid from a liquid in a colloid, by use of a flocculant, which promotes the solid clumping into flocs.
- Filtration, Mesh, bag and paper filters are used to remove large particulates suspended in fluids (e.g., fly ash) while membrane processes including microfiltration, ultrafiltration, nanofiltration, reverse osmosis, dialysis (biochemistry) utilising synthetic membranes, separates micrometre-sized or smaller species.
- Fractional distillation
- Fractional freezing
- Oil-water separation, gravimetrically separates suspended oil droplets from waste water in oil refineries, petrochemical and chemical plants, natural gas processing plants and similar industries.
- Magnetic separation.
- Precipitation.
- Recrystallization.
- Sedimentation, separates using density differences.
    - Gravity separation.
- Sieving.
- Stripping.
- Sublimation.
- Vapor-liquid separation, separates by gravity, based on the Souders-Brown equation.
- Winnowing.
- Zone refining.

Chapter 3

Crystallization

Crystallization is the (natural or artificial) process of formation of solid crystals precipitating from a solution, melt or more rarely deposited directly from a gas. Crystallization is also a chemical solid-liquid separation technique, in which mass transfer of a solute from the liquid solution to a pure solid crystalline phase occurs. In chemical engineering crystallization occurs in a crystallizator. Crystallization is therefore an aspect of precipitation, obtained through a variation of the solubility conditions of the solute in the solvent, as compared to precipitation due to chemical reaction.

Frost crystallization on a shrub.

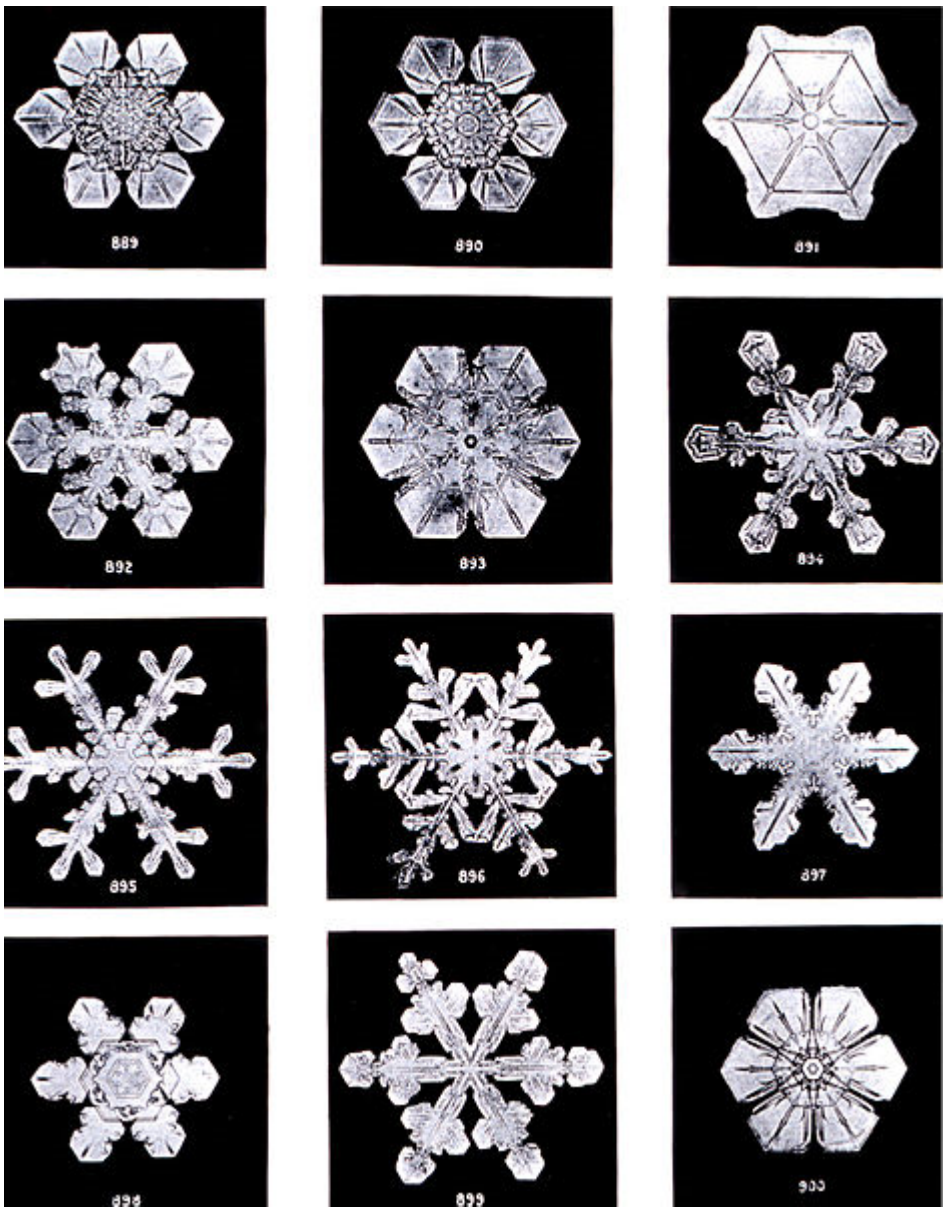
Process

The crystallization process consists of two major events, *nucleation* and *crystal growth*. *Nucleation* is the step where the solute molecules dispersed in the solvent start to gather into clusters, on the nanometer scale (elevating solute concentration in a small region), that become stable under the current operating conditions. These stable clusters constitute the nuclei. However, when the clusters are not stable, they redissolve. Therefore, the clusters need to reach a critical size in order to become stable nuclei. Such critical size is dictated by the operating conditions (temperature, supersaturation, etc.). It is at the stage of nucleation that the atoms arrange in a defined and periodic manner that defines the crystal structure — note that "crystal structure" is a special term that refers to the relative arrangement of the atoms, not the macroscopic properties of the crystal (size and shape), although those are a result of the internal crystal structure.

The *crystal growth* is the subsequent growth of the nuclei that succeed in achieving the critical cluster size. Nucleation and growth continue to occur simultaneously while the supersaturation exists. Supersaturation is the driving force of the crystallization, hence the rate of nucleation and growth is driven by the existing supersaturation in the solution. Depending upon the conditions, either nucleation or growth may be predominant over the other, and as a result, crystals with different sizes and shapes are obtained (control of crystal size and shape constitutes one of the main challenges in industrial manufacturing, such as for pharmaceuticals). Once the supersaturation is exhausted, the solid-liquid system reaches equilibrium and the crystallization is complete, unless the operating conditions are modified from equilibrium so as to supersaturate the solution again.

Many compounds have the ability to crystallize with different crystal structures, a phenomenon called polymorphism. Each polymorph is in fact a different thermodynamic solid state and crystal polymorphs of the same compound exhibit different physical properties, such as dissolution rate, shape (angles between facets and facet growth rates), melting point, etc. For this reason, polymorphism is of major importance in industrial manufacture of crystalline products.

Crystallization in nature



Snowflakes are a very well known example, where subtle differences in *crystal growth* conditions result in different geometries.

Crystallized honey

There are many examples of natural process that involve crystallization.

Geological time scale process examples include:

- Natural (mineral) crystal formation;
- Stalactite/stalagmite, rings formation.

Usual time scale process examples include:

- Snow flakes formation;
- Honey crystallization (nearly all types of honey crystallize).

Artificial methods

For crystallization to occur from a solution it must be supersaturated. This means that the solution has to contain more solute entities (molecules or ions) dissolved than it would contain under the equilibrium (saturated solution). This can be achieved by various methods, with (1) solution cooling, (2) addition of a second solvent to reduce the solubility of the solute (technique known as antisolvent or drown-out), (3) chemical reaction and (4) change in pH being the most common methods used in industrial practice. Other methods, such as solvent evaporation, can also be used. The spherical crystallization has some advantages (flowability and bioavailability) for the formulation of pharmaceutical drugs.

Applications

There are two major groups of applications for the *artificial crystallization* process: *crystal production* and purification.

Crystal production

From a material industry perspective:

- *Macroscopic crystal* production: for supply the demand of natural-like crystals with methods that "accelerate time-scale" for massive production and/or perfection.
    - *Ionic crystal* production;
    - *Covalent crystal* production.
- *Tiny size crystal*s:
    - *Powder, sand and smaller sizes*: using methods for powder and controlled (nanotechnology fruits) forms.
        - *Mass-production*: on chemical industry, like salt-powder production.
        - *Sample production*: small production of tiny crystals for material characterization. Controlled recrystallization is an important method to supply unusual crystals, that are needed to reveal the molecular structure and nuclear forces inside a typical molecule of a crystal. Many techniques, like X-ray crystallography and NMR spectroscopy, are widely used in chemistry and biochemistry to determine the structures of an immense variety of molecules, including inorganic compounds and bio-macromolecules.
    - Thin film production.

Massive production examples:

- "Powder salt for food" industry;
- Silicon crystal wafer production.
- Production of sucrose from sugar beet, where the sucrose is crystallized out from an aqueous solution.

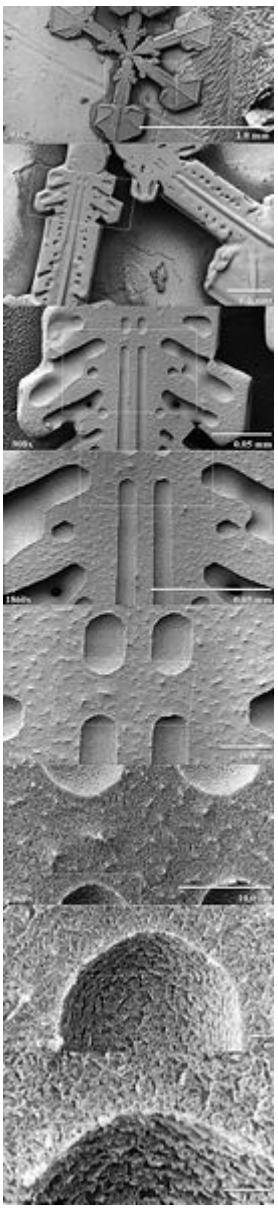
Purification

Used to improve (obtaining very pure substance) and/or verify their purity.

Crystallization separates a product from a liquid feedstream, often in extremely pure form, by cooling the feedstream or adding precipitants which lower the solubility of the desired product so that it forms crystals.

Well formed crystals are expected to be pure because each molecule or ion must fit perfectly into the lattice as it leaves the solution. Impurities would normally not fit as well in the lattice, and thus remain in solution preferentially. Hence, molecular recognition is the principle of purification in crystallization. However, there are instances when impurities incorporate into the lattice, hence, decreasing the level of purity of the final crystal product. Also, in some cases, the solvent may incorporate into the lattice forming a *solvate*. In addition, the solvent may be 'trapped' (in liquid state) within the crystal formed, and this phenomenon is known as *inclusion*.

Thermodynamic view



Low-temperature SEM magnification series for a snow crystal. The crystals are captured, stored, and sputter coated with platinum at cryo-temperatures for imaging.

The nature of a crystallization process is governed by both thermodynamic and kinetic factors, which can make it highly variable and difficult to control. Factors such as impurity level, mixing regime, vessel design, and cooling profile can have a major impact on the size, number, and shape of crystals produced.

Now put yourself in the place of a molecule within a pure and *perfect crystal*, being heated by an external source. At some sharply defined temperature, a bell rings, you must leave your neighbours, and the complicated architecture of the crystal collapses to that of a liquid. Textbook thermodynamics says that melting occurs because the entropy, S, gain in your system by spatial randomization of the molecules has overcome the enthalpy, H, loss due to breaking the crystal packing forces:

$$T(S_{liquid} - S_{solid}) > H_{liquid} - H_{solid}$$

$$G_{liquid} < G_{solid}$$

This rule suffers no exceptions when the temperature is rising. By the same token, on cooling the melt, at the very same temperature the bell should ring again, and molecules should click back into the very same crystalline form. The entropy decrease due to the ordering of molecules within the system is overcompensated by the thermal randomization of the surroundings, due to the release of the heat of fusion; the entropy of the universe increases.

But liquids that behave in this way on cooling are the exception rather than the rule; in spite of the second principle of thermodynamics, crystallization usually occurs at lower temperatures (supercooling). This can only mean that a crystal is more easily destroyed than it is formed. Similarly, it is usually much easier to dissolve a perfect crystal in a solvent than to grow again a good crystal from the resulting solution. The nucleation and growth of a crystal are under kinetic, rather than thermodynamic, control.

Equipment for crystallization

1. *Tank crystallizers*. Tank crystallization is an old method still used in some specialized cases. Saturated solutions, in tank crystallization, are allowed to cool in open tanks. After a period of time the mother liquid is drained and the crystals removed. Nucleation and size of crystals are difficult to control. Typically, labor costs are very high.
2. *Scraped surface crystallizers*. One type of scraped surface crystallizer is the Swenson-Walker crystallizer, which consists of an open trough 0.6 m wide with a semicircular bottom having a cooling jacket outside. A slow-speed spiral agitator rotates and suspends the growing crystals on turning. The blades pass close to the wall and break off any deposits of crystals on the cooled wall. The product generally has a somewhat wide crystal-size distribution.
3. *Double-pipe scraped surface crystallizer*. Also called a *votator*, this type of crystallizer is used in crystallizing ice cream and plasticizing margarine. Cooling water passes in the annular space. An internal agitator is fitted with spring-loaded scrapers that wipe the wall and provide good heat-transfer coefficients.
4. *Circulating-liquid evaporator-crystallizer*. Also called *Oslo crystallizer*. Here supersaturation is reached by evaporation. The circulating liquid is drawn by the screw pump down inside the tube side of the condensing stream heater. The heated liquid then flows into the vapor space, where flash evaporation occurs, giving some supersaturation.The vapor leaving is condensed. The supersaturated liquid flows down the downflow tube and then up through the bed of fluidized and agitated crystals, which are growing in size. The leaving saturated liquid then goes back as a recycle stream to the heater, where it is joined by the entering fluid. The larger crystals settle out and slurry of crystals and mother liquid is withdrawn as a product.
5. *Circulating-magma vacuum crystallizer*. The magma or suspension of crystals is circulated out of the main body through a circulating pipe by a screw pump. The magma flows though a heater, where its temperature is raised 2-6 K. The heated liquor then mixes with body slurry and boiling occurs at the liquid surface. This causes supersaturation in the swirling liquid near the surface, which deposits in the swirling suspended crystals until they leave again via the circulating pipe. The vapors leave through the top. A steam-jet ejector provides vacuum.

6. *Continuous oscillatory baffled crystallizer* (*COBC*). The COBC is a tubular baffled crystallizer that offers plug flow under laminar flow conditions (low flow rates) with superior heat transfer coefficient, allowing controlled cooling profiles, e.g. linear, parabolic, discontinued, step-wise or any type, to be achieved. This gives much better control over crystal size, morphology and consistent crystal products.

History

Crystallization is one of the pristine unit processes. It may be assumed that our ancestors used sodium chloride found in crevices of the surface rocks after drying caused by the sun: this process is still in use in modern solar ponds.

Other crystallization processes, for example sucrose production (this is the crystalline product with the largest world production, followed by sodium chloride), or in pigment manufacturing, were used in ancient times. These substances were sometimes produced by crystallizing the solutes of some more or less natural brine.

In more recent times, the fast expansion of the chemical industry has required a thorough study of the dynamics of crystallization, and this unit operation is now used in many industrial manufacturing areas: table salt, sugar, sodium sulfate, urea, just to name a few, are produced by crystallization from solutions.

Crystallizer technology has progressed alongside with the new processes. Once simple tanks in which, through cooling, evaporation or maybe through pH variation a crystal was obtained, nowadays continuous machines ensure a remarkable consistency in the product characteristics. Among the first models of modern crystallizers were probably the *calandria* type, being today the standard crystallizer for sucrose, and the *Oslo*, named after the Norwegian capital, since it was developed to produce salt in a climate not particularly fit for solar ponds, salt being widely used in Norway in stockfish production. The Oslo type was probably the first crystallizer designed specifically for the control of crystal growth.

Crystallization dynamics

As mentioned above, a crystal is formed following a well-defined pattern, or structure, dictated by forces acting at the molecular level. As a consequence, during its formation process the crystal is in an environment where the solute concentration reaches a certain critical value, before changing status. Solid formation, impossible below the solubility threshold at the given temperature and pressure conditions, may then take place at a concentration higher than the theoretical solubility level. The difference between the actual value of the solute concentration at the crystallization limit and the theoretical (static) solubility threshold is called supersaturation and is a fundamental factor in crystallization dynamics. Supersaturation is the driving force for both the initial nucleation step and the following crystal growth, both of which could not occur in saturated or undersaturated conditions.

Nucleation

Nucleation is the initiation of a phase change in a small region, such as the formation of a solid crystal from a liquid solution. It is a consequence of rapid local fluctuations on a

molecular scale in a homogeneous phase that is in a state of metastable equilibrium. Total nucleation is the sum effect of two categories of nucleation - primary and secondary.

Primary nucleation

Primary nucleation is the initial formation of a crystal where there are no other crystals present or where, if there are crystals present in the system, they do not have any influence on the process. This can occur in two conditions. The first is homogeneous nucleation, which is nucleation that is not influenced in any way by solids. These solids include the walls of the crystallizer vessel and particles of any foreign substance. The second category, then, is heterogeneous nucleation. This occurs when solid particles of foreign substances cause an increase in the rate of nucleation that would otherwise not be seen without the existence of these foreign particles. Homogeneous nucleation rarely occurs in practice due to the high energy necessary to begin nucleation without a solid surface to catalyse the nucleation.

Primary nucleation (both homogeneous and heterogeneous) has been modelled with the following:

$$B = \frac{dN}{dt} = k_n(c - c^*)^n$$

- B is the number of nuclei formed per unit volume per unit time.
- N is the number of nuclei per unit volume.
- k_{np} is a rate constant.
- c is the instantaneous solute concentration.
- c^* is the solute concentration at saturation.
- $(c-c^*)$ is also known as supersaturation.
- n is an empirical exponent that can be as large as 10, but generally ranges between 3 and 4.

Secondary nucleation

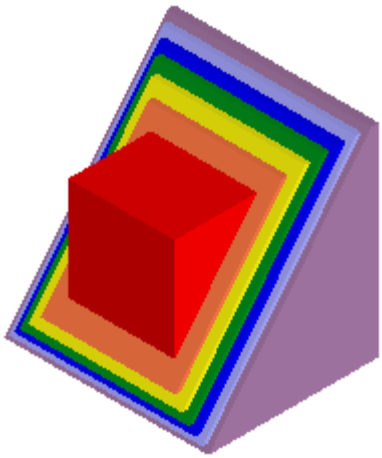
Secondary nucleation is the formation of nuclei attributable to the influence of the existing microscopic crystals in the magma. The first type of known secondary crystallization is attributable to fluid shear, the other due to collisions between already existing crystals with either a solid surface of the crystallizer or with other crystals themselves. Fluid shear nucleation occurs when liquid travels across a Crystal at a high speed, sweeping away nuclei that would otherwise be incorporated into a Crystal, causing the swept-away nuclei to become new crystals. Contact nucleation has been found to be the most effective and common method for nucleation. The benefits include the following

- Low kinetic order and rate-proportional to supersaturation, allowing easy control without unstable operation.
- Occurs at low supersaturation, where growth rate is optimum for good quality.
- Low necessary energy at which crystals strike avoids the breaking of existing crystals into new crystals.
- The quantitative fundamentals have already been isolated and are being incorporated into practice.

The following model, although somewhat simplified, is often used to model secondary nucleation:

$$B = \frac{dN}{dt} = k_1 M_T^j (c - c^*)^b$$

- k_1 is a rate constant.
- M_T is the suspension density.
- j is an empirical exponent that can range up to 1.5, but is generally 1.
- b is an empirical exponent that can range up to 5, but is generally 2.



Crystal growth

Crystal growth

Once the first small crystal, the nucleus, forms it acts as a convergence point (if unstable due to supersaturation) for molecules of solute touching - or adjacent to - the crystal so that it increases its own dimension in successive layers. The pattern of growth resembles the rings of an onion, as shown in the picture, where each colour indicates the same mass of solute; this mass creates increasingly thin layers due to the increasing surface area of the growing crystal. The supersaturated solute mass the original nucleus may *capture* in a time unit is called the *growth rate* expressed in kg/(m^2*h), and is a constant specific to the process. Growth rate is influenced by several physical factors, such as surface tension of solution, pressure, temperature, relative crystal velocity in the solution, Reynolds number, and so forth.

The main values to control are therefore:

- Supersaturation value, as an index of the quantity of solute available for the growth of the crystal;
- Total crystal surface in unit fluid mass, as an index of the capability of the solute to fix onto the crystal;
- Retention time, as an index of the probability of a molecule of solute to come into contact with an existing crystal;
- Flow pattern, again as an index of the probability of a molecule of solute to come into contact with an existing crystal (higher in laminar flow, lower in turbulent flow, but the reverse applies to the probability of contact).

The first value is a consequence of the physical characteristics of the solution, while the others define a difference between a well- and poorly designed crystallizer.

Crystal size distribution

The appearance and size range of a crystalline product is extremely important in crystallization. If further processing of the crystals is desired, large crystals with uniform size are important for washing, filtering, transportation, and storage. The importance lies in the fact that large crystals are easier to filter out of a solution than small crystals. Also, larger crystals have a smaller surface area to volume ratio, leading to a higher purity. This higher purity is due to less retention of mother liquor which contains impurities, and a smaller loss of yield when the crystals are washed to remove the mother liquor. The theoretical crystal size distribution can be estimated as a function of operating conditions with a fairly complicated mathematical process called population balance theory (using population balance equations).

Main crystallization processes

The main factors influencing solubility are, as we saw above:

- Concentration
- Temperature

So we may identify two main families of crystallization processes:

- Cooling crystallization
- Evaporative crystallization

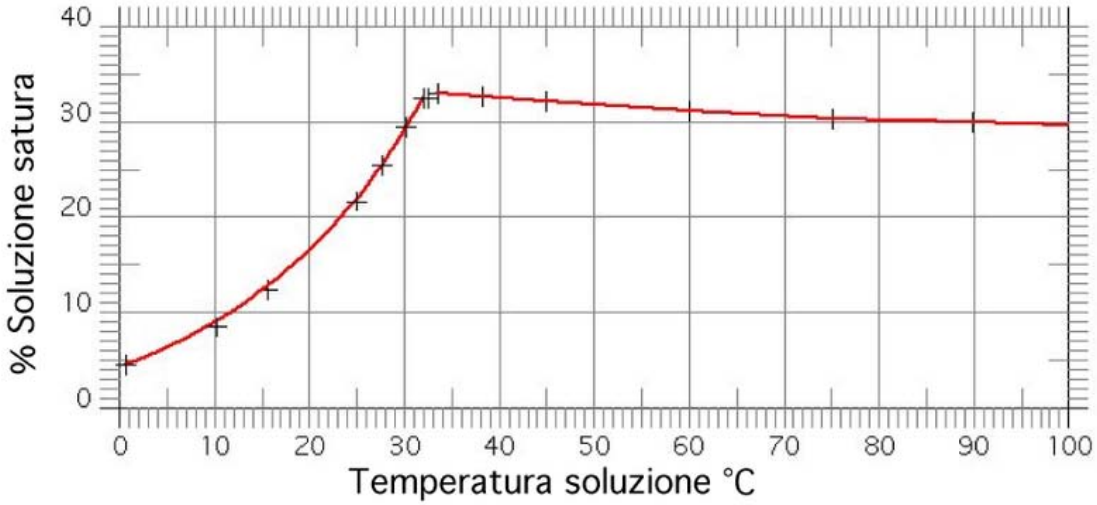
This division is not really clear-cut, since hybrid systems exist, where cooling is performed through evaporation, thus obtaining at the same time a concentration of the solution.

A crystallization process often referred to in chemical engineering is the Fractional crystallization. This is not a different process, rather a special application of one (or both) of the above.

Cooling crystallization

Application

Most chemical compounds, dissolved in most solvents, show the so-called *direct* solubility that is, the solubility threshold increases with temperature.



Solubility of the system Na_2SO_4 - H_2O

So, whenever the conditions are favourable, crystal formation results from simply cooling the solution. Here *cooling* is a relative term: austenite crystals in a steel form well above 1000 °C. An example of this crystallization process is the production of Glauber's salt, a crystalline form of sodium sulphate. In the picture, where equilibrium temperature is on the x-axis and equilibrium concentration (as mass percent of solute in saturated solution) in y-axis, it is clear that sulphate solubility quickly decreases below 32.5 °C. Assuming a saturated solution at 30 °C, by cooling it to 0 °C (note that this is possible thanks to the freezing-point depression), the precipitation of a mass of sulphate occurs corresponding to the change in solubility from 29% (equilibrium value at 30°C) to approximately 4.5% (at 0°C) - actually a larger crystal mass is precipitated, since sulphate entrains hydration water, and this has the side effect of increasing the final concentration.

There are of course limitation in the use of cooling crystallization:

- Many solutes precipitate in hydrate form at low temperatures: in the previous example this is acceptable, and even useful, but it may be detrimental when, for example, the mass of water of hydration to reach a stable hydrate crystallization form is more than the available water: a single block of hydrate solute will be formed - this occurs in the case of calcium chloride);
- Maximum supersaturation will take place in the coldest points. These may be the heat exchanger tubes which are sensitive to scaling, and heat exchange may be greatly reduced or discontinued;
- A decrease in temperature usually implies an increase of the viscosity of a solution. Too high a viscosity may give hydraulic problems, and the laminar flow thus created may affect the crystallization dynamics.
- It is of course not applicable to compounds having *reverse* solubility, a term to indicate that solubility increases with temperature decrease (an example occurs with sodium sulphate where solubility is reversed above 32.5 °C).

Cooling crystallizers

The simplest cooling crystallizers are tanks provided with a mixer for internal circulation, where temperature decrease is obtained by heat exchange with an intermediate fluid circulating in a jacket. These simple machines are used in batch processes, as in

processing of pharmaceuticals and are prone to scaling. Batch processes normally provide a relatively variable quality of product along the batch.

The *Swenson-Walker* crystallizer is a model, specifically conceived by Swenson Co. around 1920, having a semicylindric horizontal hollow trough in which a hollow screw conveyor or some hollow discs, in which a refrigerating fluid is circulated, plunge during rotation on a longitudinal axis. The refrigerating fluid is sometimes also circulated in a jacket around the trough. Crystals precipitate on the cold surfaces of the screw/discs, from which they are removed by scrapers and settle on the bottom of the trough. The screw, if provided, pushes the slurry towards a discharge port.

A common practice is to cool the solutions by flash evaporation: when a liquid at a given T_0 temperature is transferred in a chamber at a pressure P_1 such that the liquid saturation temperature T_1 at P_1 is lower than T_0, the liquid will release heat according to the temperature difference and a quantity of solvent, whose total latent heat of vaporization equals the difference in enthalpy. In simple words, the liquid is cooled by evaporating a part of it.

Evaporative crystallization

Another option is to obtain, at an approximately constant temperature, the precipitation of the crystals by increasing the solute concentration above the solubility threshold. To obtain this, the solute/solvent mass ratio is increased using the technique of evaporation. This process is of course insensitive to change in temperature (as long as hydration state remains unchanged).

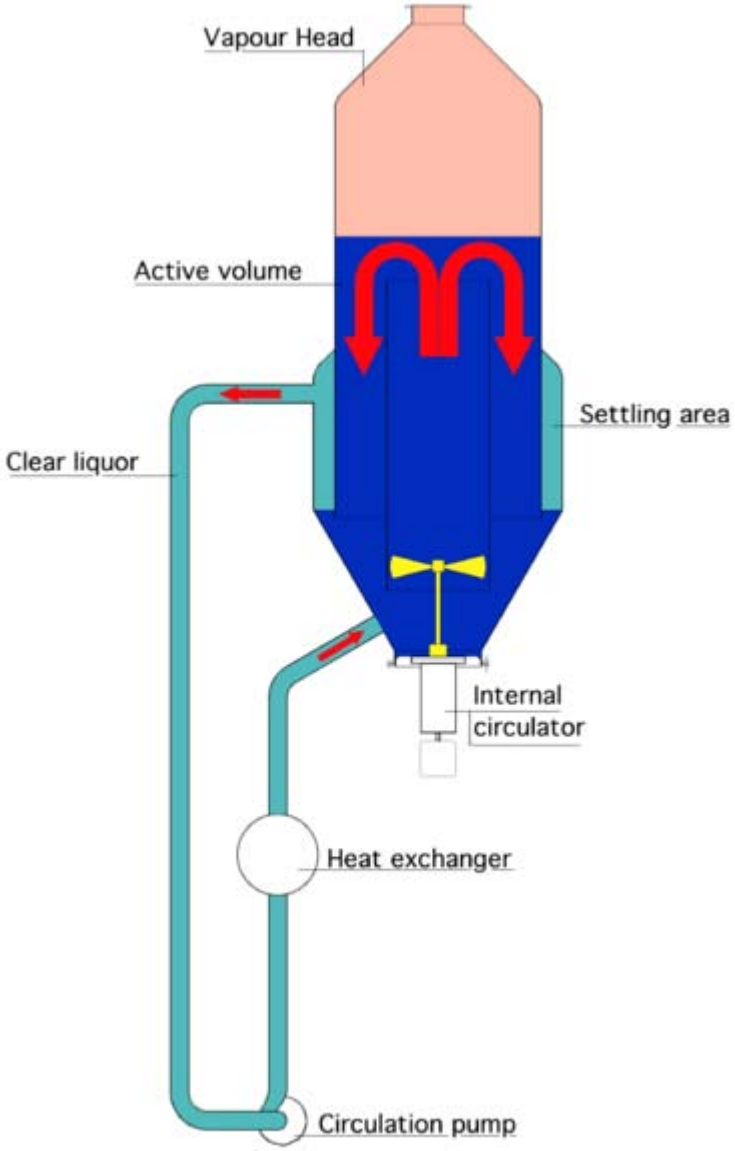
All considerations on control of crystallization parameters are the same as for the cooling models.

Evaporative crystallizers

Most industrial crystallizers are of the evaporative type, such as the very large sodium chloride and sucrose units, whose production accounts for more than 50% of the total world production of crystals. The most common type is the *forced circulation* (FC) model. A pumping device (a pump or an axial flow mixer) keeps the crystal slurry in homogeneous suspension throughout the tank, including the exchange surfaces; by controlling pump flow, control of the contact time of the crystal mass with the supersaturated solution is achieved, together with reasonable velocities at the exchange surfaces. The Oslo, mentioned above, is a refining of the evaporative forced circulation crystallizer, now equipped with a large crystals settling zone to increase the retention time (usually low in the FC) and to roughly separate heavy slurry zones from clear liquid.

The DTB crystallizer

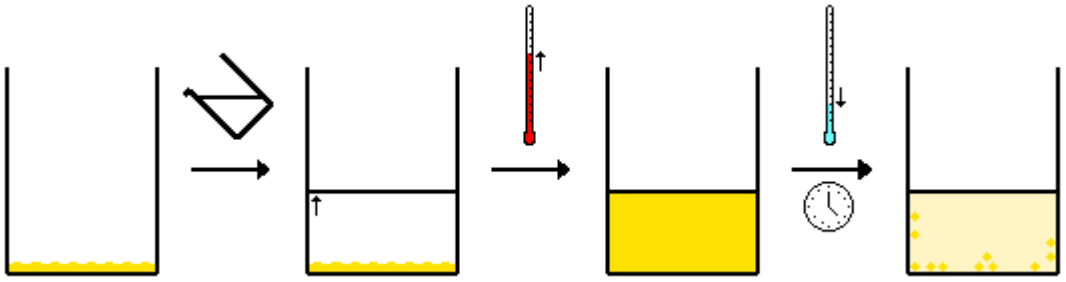
DTB Crystallizer



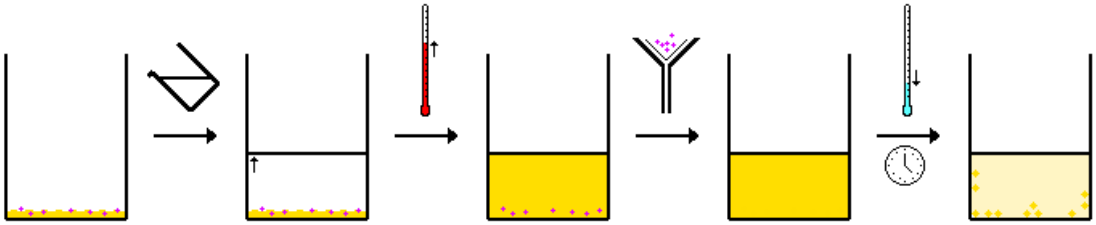
Schematic of DTB

Whichever the form of the crystallizer, to achieve an effective process control it is important to control the retention time and the crystal mass, to obtain the optimum

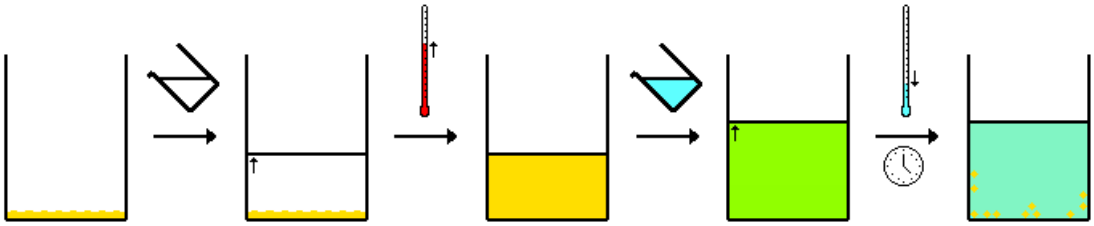
conditions in terms of crystal specific surface and the fastest possible growth. This is achieved by a separation - to put it simply - of the crystals from the liquid mass, in order to manage the two flows in a different way. The practical way is to perform a gravity settling to be able to extract (and possibly recycle separately) the (almost) clear liquid, while managing the mass flow around the crystallizer to obtain a precise slurry density elsewhere. A typical example is the DTB (*Draft Tube and Baffle*) crystallizer, an idea of Richard Chisum Bennett (a Swenson engineer and later President of Swenson) at the end of the 1950s. The DTB crystallizer has an internal circulator, typically an axial flow mixer - yellow - pushing upwards in a draft tube while outside the crystallizer there is a settling area in an annulus; in it the exhaust solution moves upwards at a very low velocity, so that large crystals settle - and return to the main circulation - while only the fines, below a given grain size are extracted and eventually destroyed by increasing or decreasing temperature, thus creating additional supersaturation. A quasi-perfect control of all parameters is achieved. This crystallizer, and the derivative models (Krystal, CSC, etc.) could be the ultimate solution if not for a major limitation in the evaporative capacity, due to the limited diameter of the vapour head and the relatively low external circulation not allowing large amounts of energy to be supplied to the system.



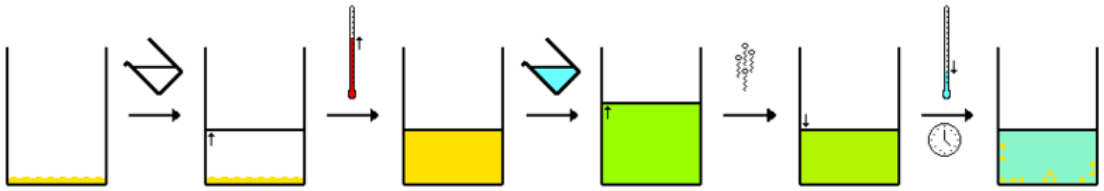
The process



Hot-filtration, 1 solvent

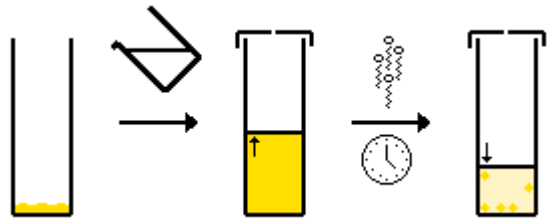


2 solvent

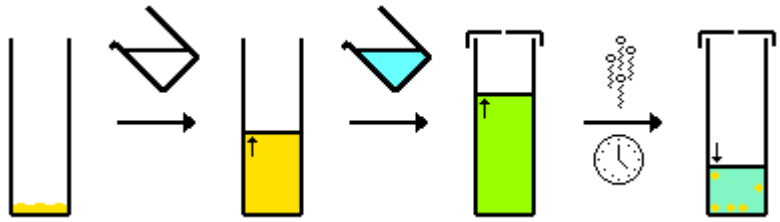


2 solvent, with evaporation

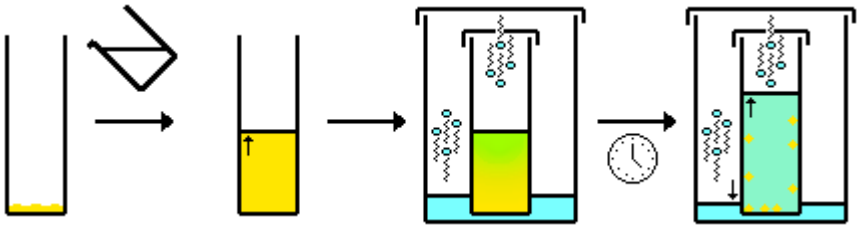
X-ray crystals



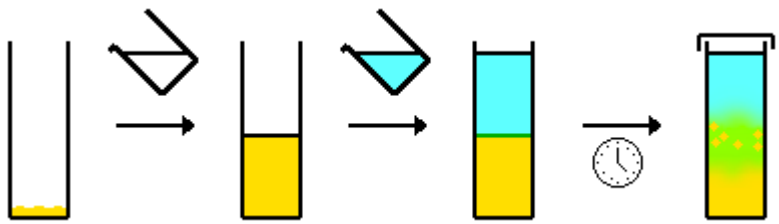
slow evaporation 1 solvent



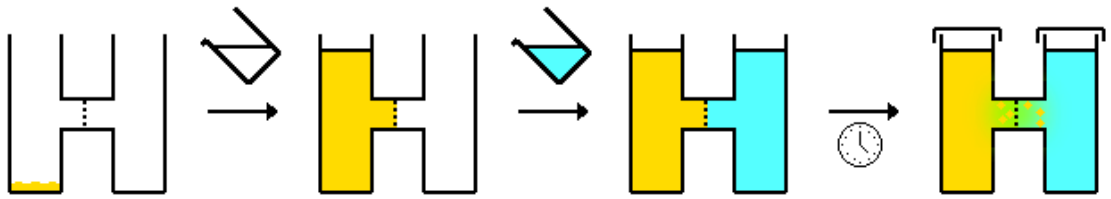
slow evaporation 2 solvent



slow gas diffusion 2 solvent



slow liquid diffusion



slow liquid diffusion - H Tube

Chpater – 5

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.

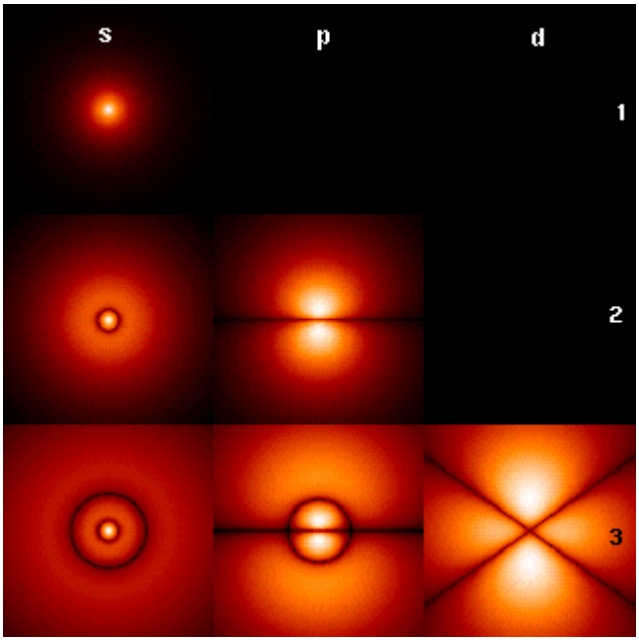
Silicon is used to create most semiconductors commercially. Dozens of other materials are used, including germanium, gallium arsenide, and silicon carbide. A pure semiconductor is often called an “intrinsic” semiconductor. The electronic properties and the conductivity of a semiconductor can be changed 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

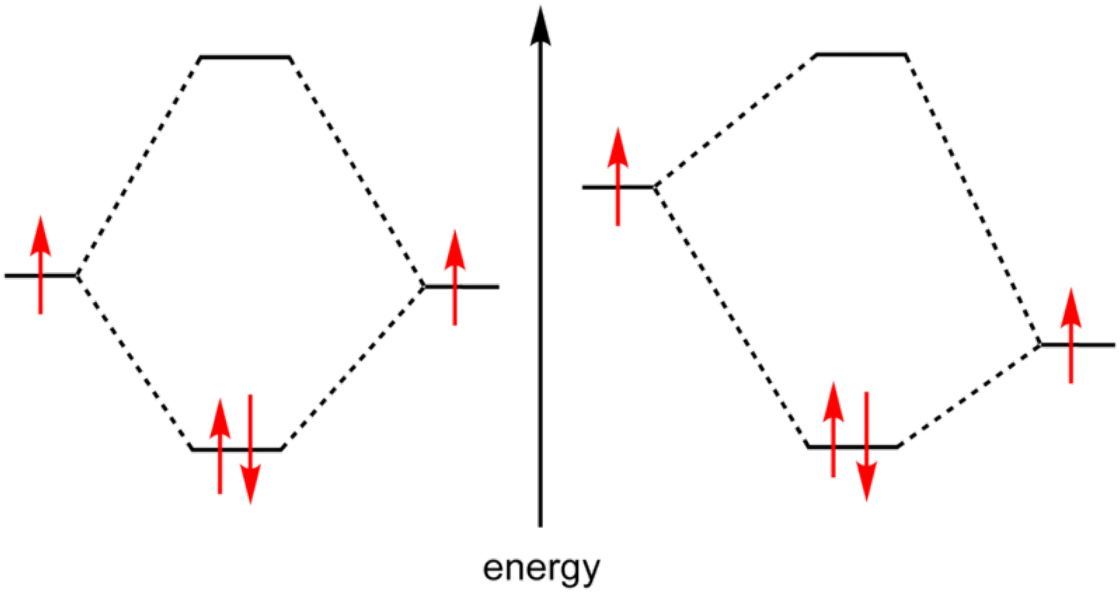
There are three popular ways to classify the electronic structure of a crystal.

Band structure

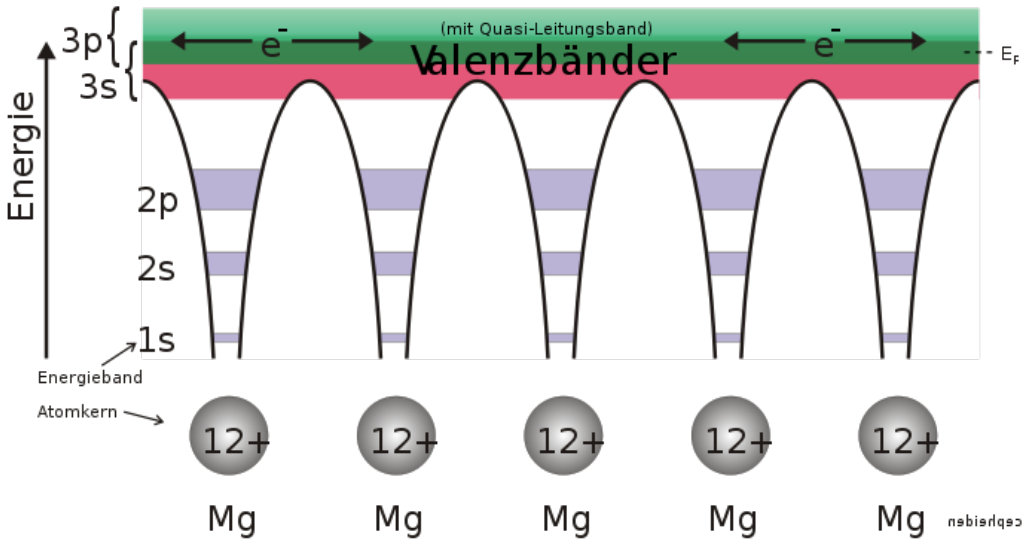
atoms – crystal – vacuum



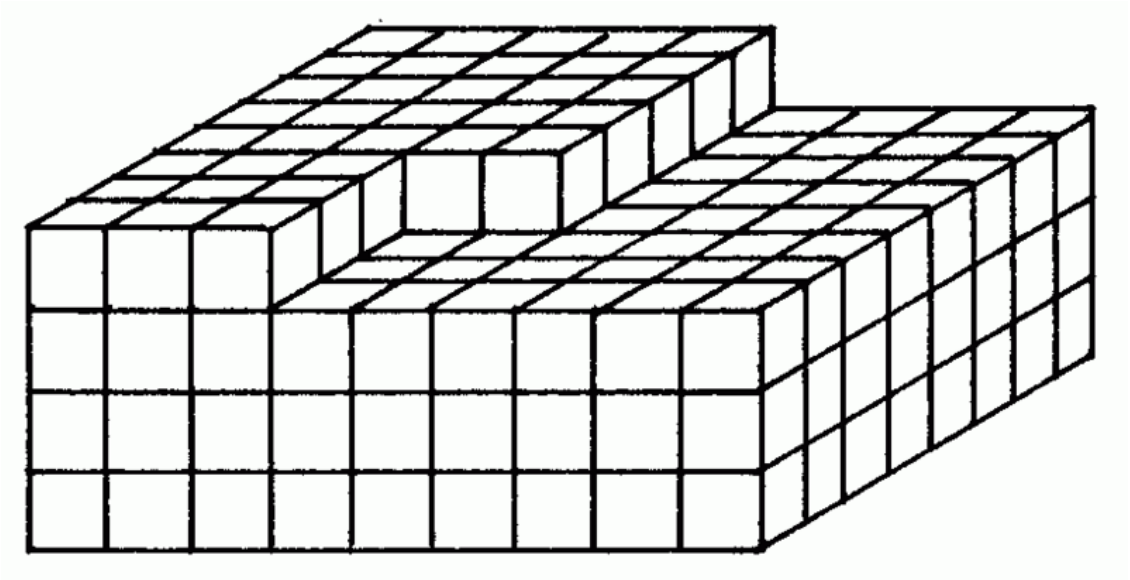
In a single H-atom an electron resides in well known orbitals. Note that the orbitals are called s,p,d in order of increasing circular current.



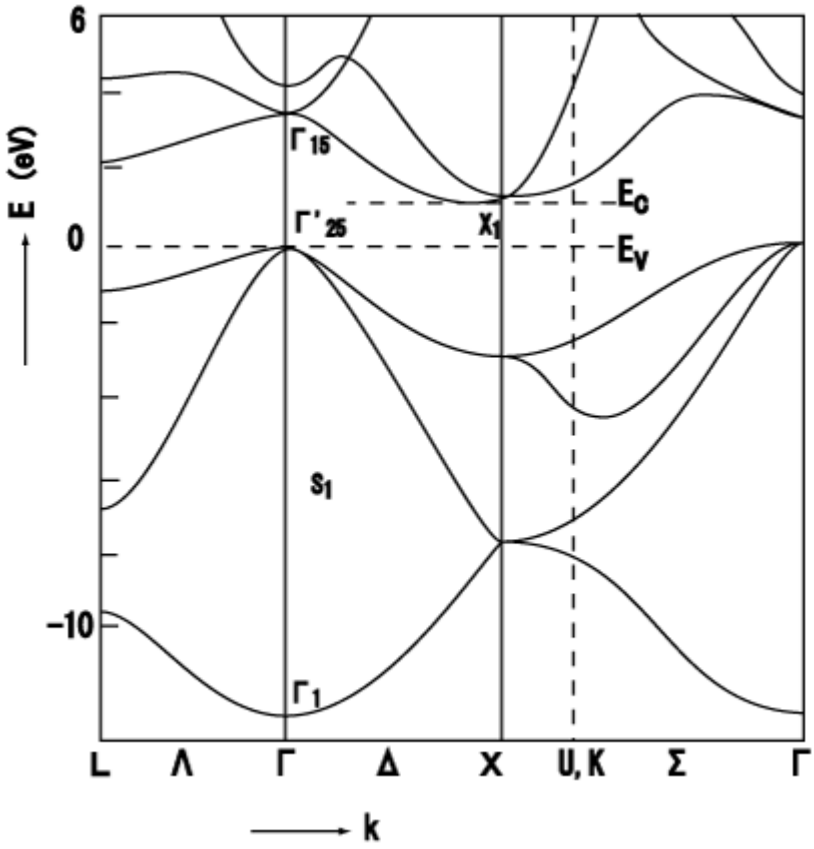
Putting two atoms together leads to delocalized orbitals across two atoms, yielding a covalent bond. Due to the Pauli exclusion principle, every state can contain only one electron.



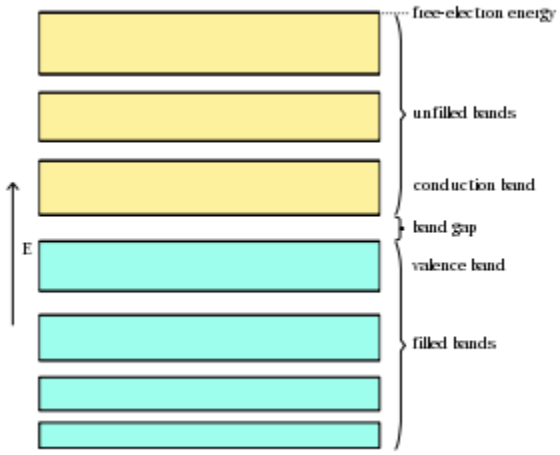
This can be continued with more atoms. Note: This picture shows a metal, not an actual semiconductor.



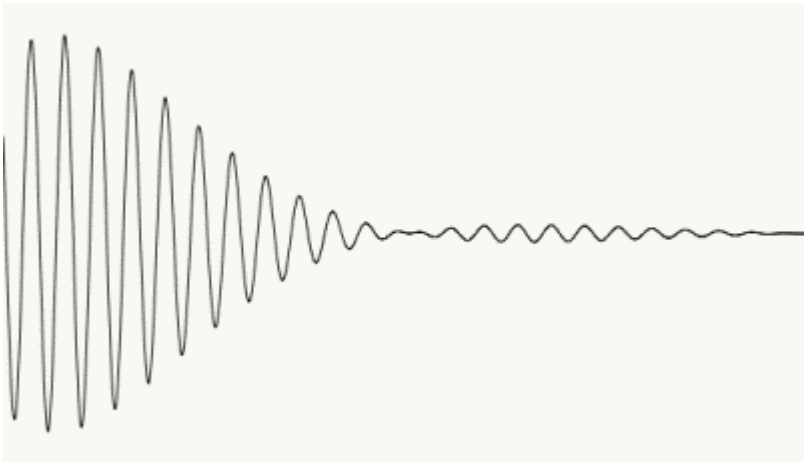
Continuing to add creates a crystal, which may then be cut into a tape and fused together at the ends to allow circular currents.



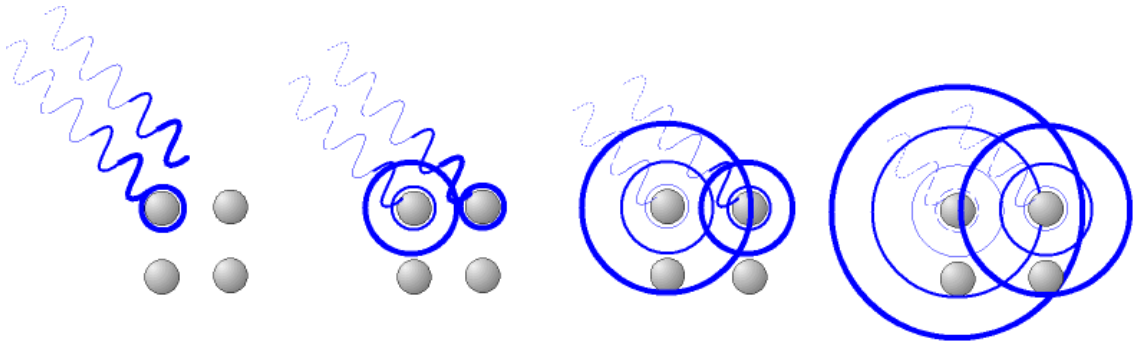
For this regular solid the band structure can be calculated or measured.



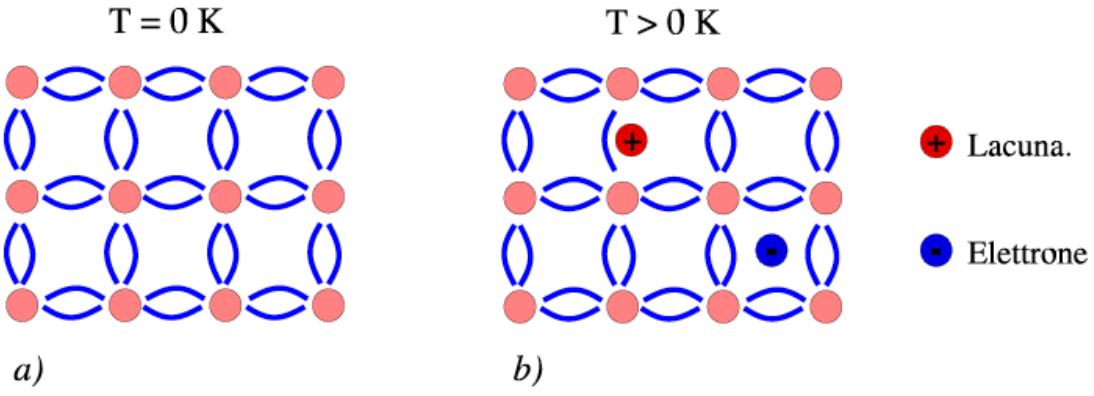
Integrating over the k axis gives the bands of a semiconductor showing a full valence band and an empty conduction band. Generally stopping at the vacuum level is undesirable, because some people want to calculate: photoemission, inverse photoemission



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.

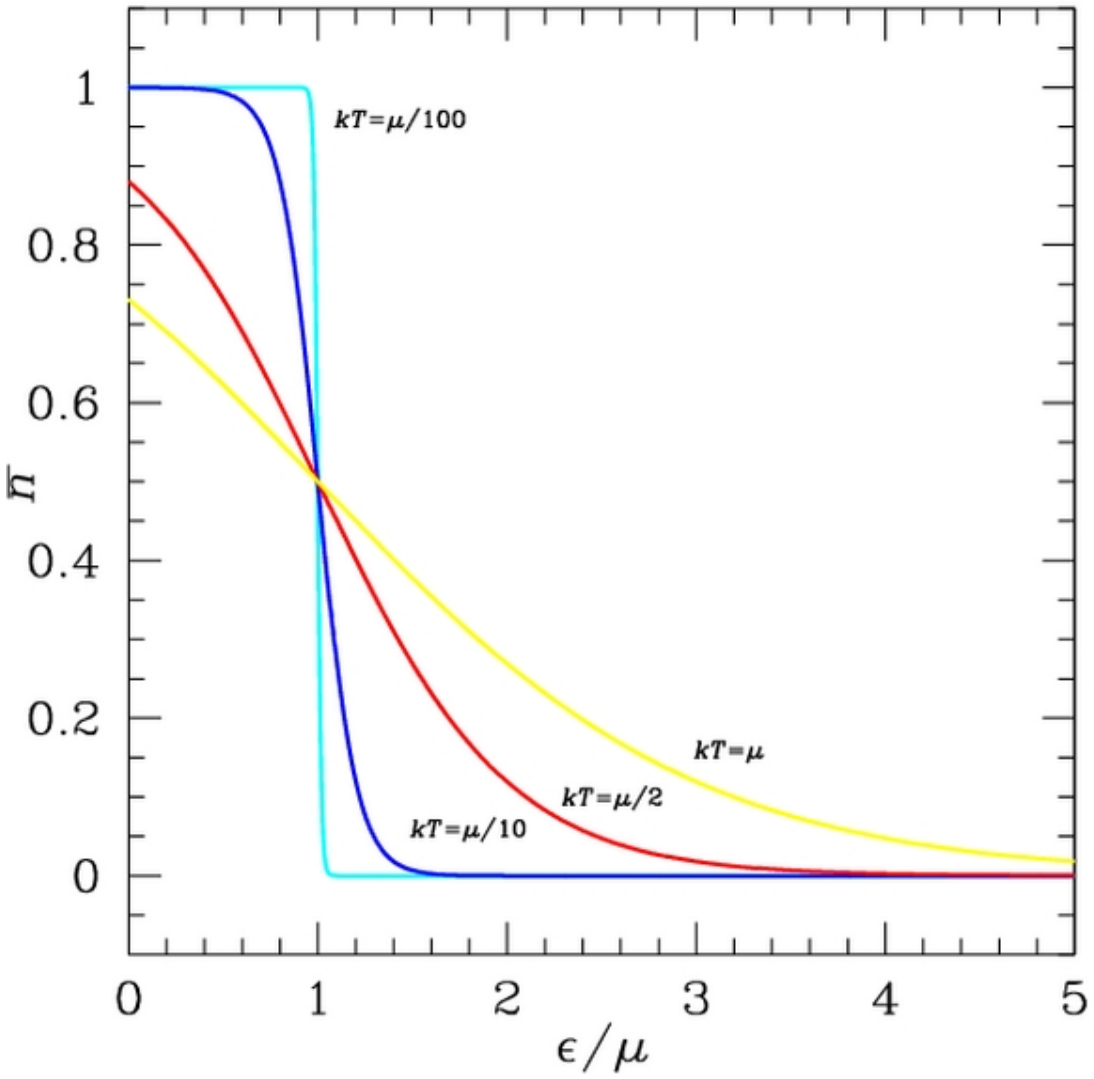
Electrons excited to the conduction band also leave behind electron holes, or unoccupied states in the valence band. Both the conduction band electrons and the valence band holes contribute to electrical conductivity. The holes themselves don't actually move, but a neighboring electron can move to fill 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,

$$n = p = n_i.$$

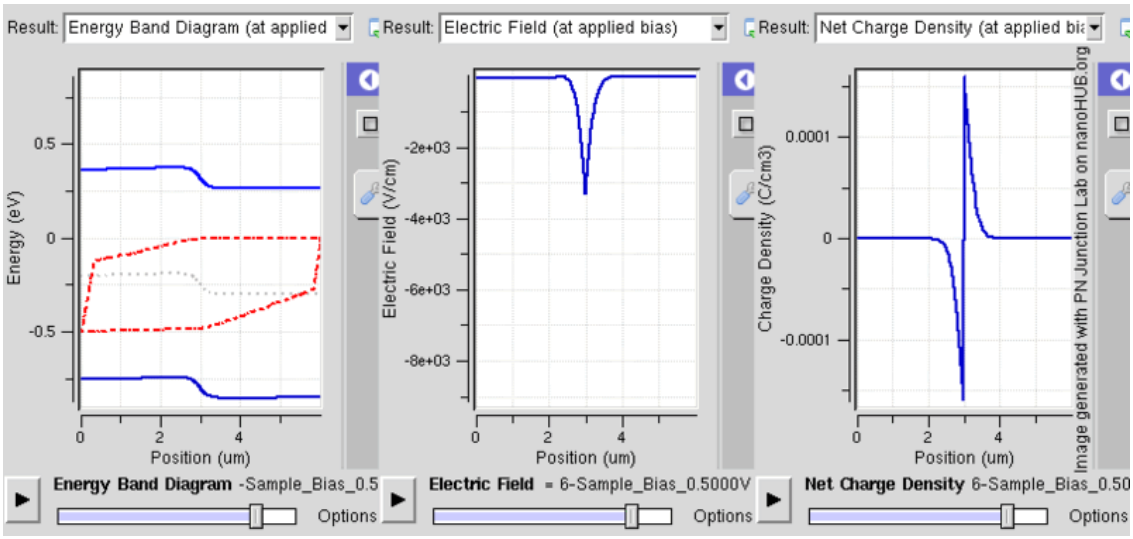
If we have a non-intrinsic semiconductor in thermal equilibrium the relation becomes:

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

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/cm^3. 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

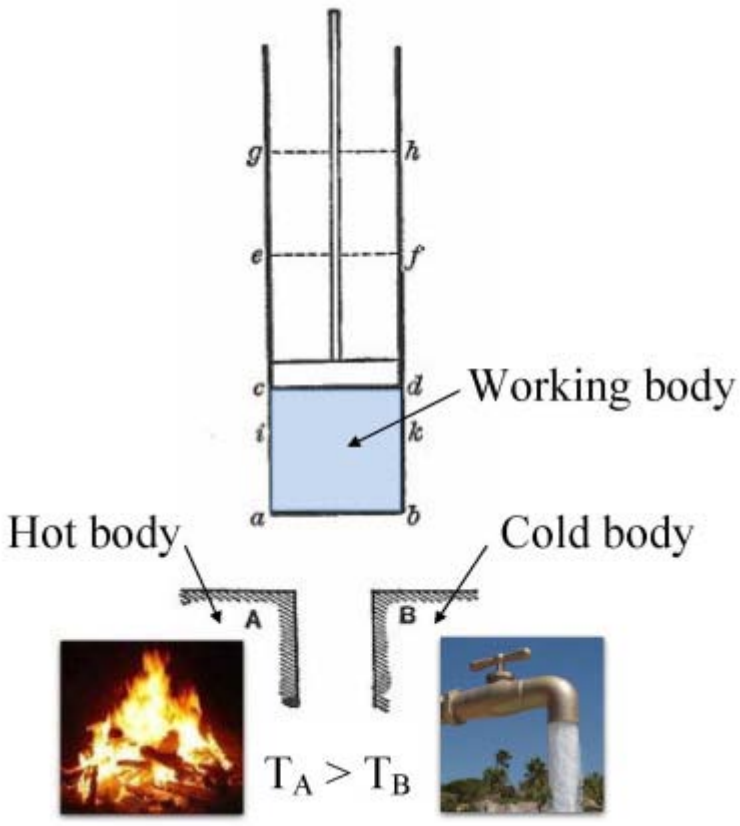
the material. A high degree of crystalline perfection is also required, since faults in crystal structure (such as dislocations, twins, and stacking faults) interfere with the semiconducting properties of the material. Crystalline faults are a major cause of defective semiconductor devices. The larger 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 4

Thermodynamics



Annotated color version of the original 1824 Carnot heat engine showing the hot body (boiler), working body (system, steam), and cold body (water), the letters labeled according to the stopping points in Carnot cycle.

Thermodynamics is the science of energy conversion involving heat and other forms of energy, most notably mechanical work. It studies and interrelates the macroscopic variables, such as temperature, volume and pressure, which describe physical, thermodynamic systems.

Historically, thermodynamics developed out of a desire to increase the efficiency of early steam engines, particularly through the work of French physicist Nicolas Léonard Sadi Carnot (1824) who believed that engine efficiency was the key that could help France win the Napoleonic Wars. Scottish physicist Lord Kelvin was the first to formulate a concise definition of thermodynamics in 1854:

Thermo-dynamics is the subject of the relation of heat to forces acting between contiguous parts of bodies, and the relation of heat to electrical agency.

Two fields of thermodynamics emerged in the following decades. Statistical thermodynamics, or statistical mechanics, (1860) concerned itself with statistical predictions of the collective motion of particles from their microscopic behavior, while chemical thermodynamics (1873) studies the nature of the role of entropy in the process of chemical reaction.

Introduction

The starting point for most thermodynamic considerations are the laws of thermodynamics, which postulate that energy can be exchanged between physical systems as heat or work. They also postulate the existence of a quantity named entropy, which can be defined for any isolated system that is in thermodynamic equilibrium.

In thermodynamics, interactions between large ensembles of objects are studied and categorized. Central to this are the concepts of *system* and *surroundings*. A system is composed of particles, whose average motions define its properties, which in turn are related to one another through equations of state. Properties can be combined to express internal energy and thermodynamic potentials, which are useful for determining conditions for equilibrium and spontaneous processes.

With these tools, thermodynamics can be used to describe how systems respond to changes in their environment. This can be applied to a wide variety of topics in science and engineering, such as engines, phase transitions, chemical reactions, transport phenomena, and even black holes. The results of thermodynamics are essential for other fields of physics and for chemistry, chemical engineering, aerospace engineering, mechanical engineering, cell biology, biomedical engineering, materials science, and economics, to name a few.

The present article is focused mainly on classical thermodynamics, which is concerned with systems in thermodynamic equilibrium. It is wise to distinguish classical thermodynamics from non-equilibrium thermodynamics, which is concerned with systems that are not in thermodynamic equilibrium.

École Polytechnique	Glasgow school	Berlin school	Edinburgh school
			
Sadi Carnot (1796-1832)	William Thomson (1824-1907)	Rudolf Clausius (1822-1888)	James Maxwell (1831-1879)
Vienna school	**Gibbsian school**	**Dresden school**	**Dutch school**
			
Ludwig Boltzmann (1844-1906)	Willard Gibbs (1839-1903)	Gustav Zeuner (1828-1907)	Johannes der Waals (1837-1923)

The thermodynamicists representative of the original eight founding schools of thermodynamics. The schools with the most-lasting effect in founding the modern versions of thermodynamics are the Berlin school, particularly as established in Rudolf Clausius's 1865 textbook *The Mechanical Theory of Heat*, the Vienna school, with the statistical mechanics of Ludwig Boltzmann, and the Gibbsian school at Yale University, American engineer Willard Gibbs' 1876 *On the Equilibrium of Heterogeneous Substances* launching chemical thermodynamics.

History

The history of thermodynamics as a scientific discipline generally begins with Otto von Guericke who, in 1650, built and designed the world's first vacuum pump and demonstrated a vacuum using his Magdeburg hemispheres. Guericke was driven to make a vacuum in order to disprove Aristotle's long-held supposition that 'nature abhors a vacuum'. Shortly after Guericke, the English physicist and chemist Robert Boyle had learned of Guericke's designs and, in 1656, in coordination with English scientist Robert Hooke, built an air pump. Using this pump, Boyle and Hooke noticed a correlation between pressure, temperature, and volume. In time, Boyle's Law was formulated, which states that pressure and volume are inversely proportional. Then, in 1679, based on these concepts, an associate of Boyle's named Denis Papin built a steam digester, which was a closed vessel with a tightly fitting lid that confined steam until a high pressure was generated.

Later designs implemented a steam release valve that kept the machine from exploding. By watching the valve rhythmically move up and down, Papin conceived of the idea of a piston and a cylinder engine. He did not, however, follow through with his design.

Nevertheless, in 1697, based on Papin's designs, engineer Thomas Savery built the first engine, followed by Thomas Newcomen in 1712. Although these early engines were crude and inefficient, they attracted the attention of the leading scientists of the time.

The fundamental concepts of heat capacity and latent heat, which were necessary for the development of thermodynamics, were developed by Professor Joseph Black at the University of Glasgow, where James Watt was employed as an instrument maker. Black and Watt performed experiments together, but it was Watt who conceived the idea of the external condenser which resulted in a large increase in steam engine efficiency. Drawing on all the previous work led Sadi Carnot, the "father of thermodynamics", to publish *Reflections on the Motive Power of Fire* (1824), a discourse on heat, power, energy and engine efficiency. The paper outlined the basic energetic relations between the Carnot engine, the Carnot cycle, and motive power. It marked the start of thermodynamics as a modern science.

The first thermodynamic textbook was written in 1859 by William Rankine, originally trained as a physicist and a civil and mechanical engineering professor at the University of Glasgow. The first and second laws of thermodynamics emerged simultaneously in the 1850s, primarily out of the works of William Rankine, Rudolf Clausius, and William Thomson (Lord Kelvin).

The foundations of statistical thermodynamics were set out by physicists such as James Clerk Maxwell, Ludwig Boltzmann, Max Planck, Rudolf Clausius and J. Willard Gibbs.

During the years 1873-76 the American mathematical physicist Josiah Willard Gibbs published a series of three papers, the most famous being *On the Equilibrium of Heterogeneous Substances*, in which he showed how thermodynamic processes could be graphically analyzed, by studying the energy, entropy, volume, temperature and pressure of the thermodynamic system in such a manner, one can determine if a process would occur spontaneously. During the early 20th century, chemists such as Gilbert N. Lewis, Merle Randall, and E. A. Guggenheim began to apply the mathematical methods of Gibbs to the analysis of chemical processes.

Etymology

The etymology of *thermodynamics* has an intricate history.It was first spelled in a hyphenated form as an adjective (*thermo-dynamic*) and from 1854 to 1868 as the noun *thermo-dynamics* to represent the science of generalized heat engines.

American biophysisist Donald Haynie claims that *thermodynamics* was coined in 1840 from the Greek root θέρμη *therme,* meaning heat and δύναμις,*dynamis,* meaning power. However, this etymology has been cited as unlikely.

Pierre Perrot claims that the term *thermodynamics* was coined by James Joule in 1858 to designate the science of relations between heat and power., however, Joule never used that term, but used instead the term *perfect thermo-dynamic engine* in reference to Thomson’s 1849 phraseology.

By 1858, *thermo-dynamics*, as a functional term, was used in William Thomson's paper *An Account of Carnot's Theory of the Motive Power of Heat.*

Interpretations

Thermodynamics has developed into several related branches of science, each with a different focus.

Classical thermodynamics

Classical thermodynamics is the description of the states of thermodynamical systems at near-equilibrium, using macroscopic, empirical properties directly measurable in the laboratory. It is used to model exchanges of energy, work and heat based on the laws of thermodynamics. The qualifier *classical* reflects the fact that it represents the level of knowledge in the late 19th century. An atomic interpretation of these laws was gradually provided by the development of statistical mechanics.

Statistical mechanics

Statistical mechanics (or statistical thermodynamics) emerged only with the development of atomic and molecular theories in the late 19th century and early 20th century, giving thermodynamics a molecular interpretation. This field relates the microscopic properties of individual atoms and molecules to the macroscopic or bulk properties of materials that can be observed in everyday life, thereby explaining thermodynamics as a natural result of statistics and mechanics (classical and quantum) at the microscopic level. This statistical approach is in contrast to classical thermodynamics, which is a more phenomenological approach.

Chemical thermodynamics

Chemical thermodynamics is the study of the interrelation of energy with chemical reactions or with a physical change of state within the confines of the laws of thermodynamics.

Treatment of equilibrium

Equilibrium thermodynamics is the systematic study of transformations of matter and energy in systems as they approach equilibrium. The word equilibrium implies a state of balance. In an equilibrium state there are no unbalanced potentials, or driving forces, within the system. A central aim in equilibrium thermodynamics is: given a system in a well-defined initial state, subject to accurately specified constraints, to calculate what the state of the system will be once it has reached equilibrium.

Non-equilibrium thermodynamics is a branch of thermodynamics that deals with systems that are not in thermodynamic equilibrium. Most systems found in nature are not in thermodynamic equilibrium because they are not in stationary states, and are continuously and discontinuously subject to flux of matter and energy to and from other systems. The thermodynamic study of non-equilibrium systems requires more general concepts than are dealt with by equilibrium thermodynamics. Many natural systems still today remain beyond the scope of currently known macroscopic thermodynamic methods.

Laws of thermodynamics

Thermodynamics defines four laws which do not depend on the details of the systems under study or how they interact. Hence these laws are generally valid and can be applied to systems about which one knows nothing other than the balance of energy and matter transfer. Examples of such systems include Einstein's prediction of spontaneous emission, and ongoing research into the thermodynamics of black holes.

These four laws are:

- Zeroth law of thermodynamics: *If two systems are in thermal equilibrium with a third, they are also in thermal equilibrium with each other.*

This statement implies that thermal equilibrium is an equivalence relation on the set of thermodynamic systems under consideration. Systems are said to be in equilibrium if the small, random exchanges between them (eg. Brownian motion) do not lead to a net change in energy. This law is tacitly assumed in every measurement of temperature. Thus, if one seeks to decide if two bodies are at the same temperature, it is not necessary to bring them into contact and measure any changes of their observable properties in time. The law provides a fundamental definition of temperature and justification for the construction of practical thermometers.

It is interesting to note that the zeroth law was not initially recognized as a law. The need to for the zeroth law was not initially realized, so the first, second, and third laws were explicitly stated and found common acceptance in the physics community first. Once the importance of the zeroth law was realized, it was impracticable to renumber the other laws, hence the *zeroth.*

- First law of thermodynamics: *The internal energy of an isolated system is constant.*

The first law of thermodynamics is an expression of the principle of conservation of energy. It states that energy can be transformed (changed from one form to another), but cannot be created or destroyed.

The first law is usually formulated by saying that the change in the internal energy of a closed thermodynamic system is equal to the difference between the heat supplied to the system and the amount of work done by the system on its surroundings. It is important to note that internal energy is a state of the system whereas heat and work modify the state of the system. In other words, a specific internal energy of a system may be achieved by any combination of heat and work; the manner by which a system achieves a specific internal energy is path independent.

- Second law of thermodynamics: *Heat cannot spontaneously flow from a colder location to a hotter location.*

The second law of thermodynamics is an expression of the universal principle of decay observable in nature. The second law is an observation of the fact that over time, differences in temperature, pressure, and chemical potential tend to even out in a physical system that is isolated from the outside world. Entropy is a measure of how much this process has progressed. The entropy of an isolated system which is not in equilibrium will tend to increase over time, approaching a maximum value at equilibrium.

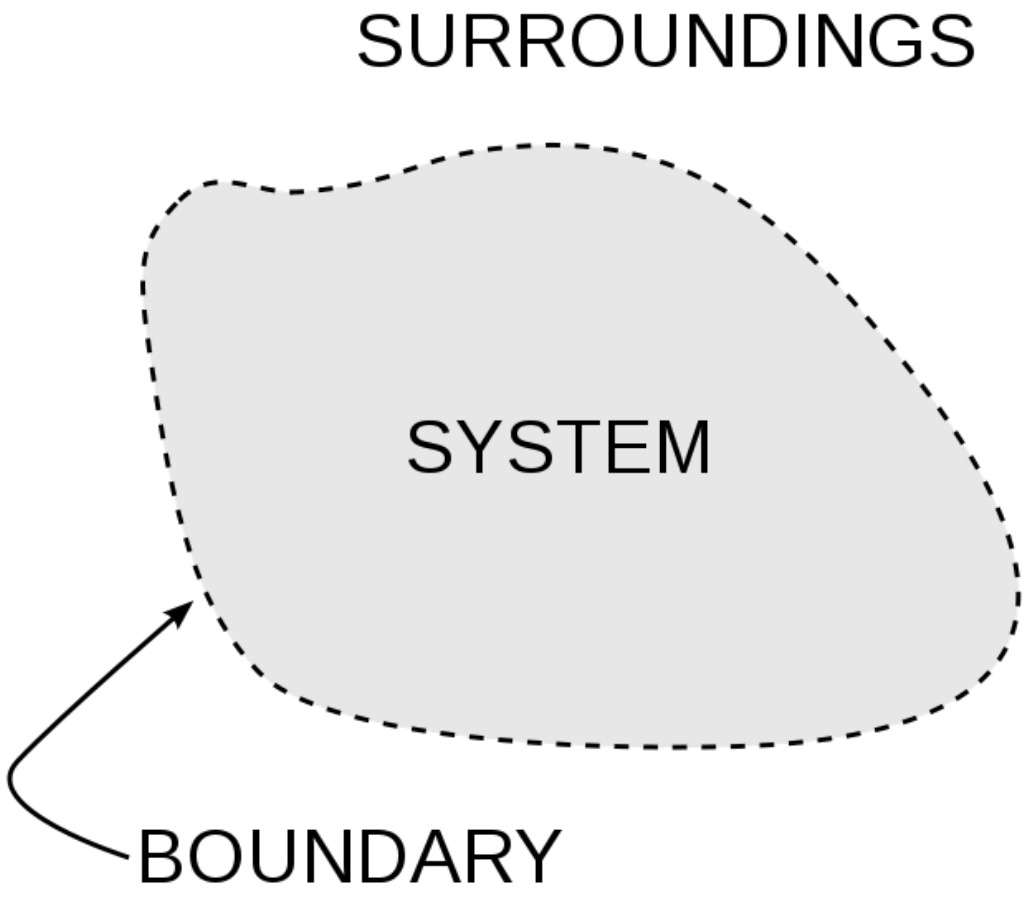
In classical thermodynamics, the second law is a basic postulate applicable to any system involving heat energy transfer; in statistical thermodynamics, the second law is a consequence of the assumed randomness of molecular chaos. There are many versions of the second law, but they all have the same effect, which is to explain the phenomenon of irreversibility in nature.

- Third law of thermodynamics: *As a system approaches absolute zero, all processes cease and the entropy of the system approaches a minimum value.*

The third law of thermodynamics is a statistical law of nature regarding entropy and the impossibility of reaching absolute zero of temperature. This law provides an absolute reference point for the determination of entropy. The entropy determined relative to this point is the absolute entropy. Alternate definitions are, "the entropy of all systems and of all states of a system is smallest at absolute zero," or equivalently "it is impossible to reach the absolute zero of temperature by any finite number of processes".

Absolute zero, at which all activity would stop if it were possible to happen, is −273.15 °C (degrees Celsius), or −459.67 °F (degrees Fahrenheit) or 0 K (kelvin).

System models



A diagram of a generic thermodynamic system

An important concept in thermodynamics is the thermodynamic system, a precisely defined region of the universe under study. Everything in the universe except the system is known as the *surroundings*. A system is separated from the remainder of the universe by a *boundary* which may be notional or not, but which by convention delimits a finite volume. Exchanges of work, heat, or matter between the system and the surroundings take place across this boundary.

In practice, the boundary is simply an imaginary dotted line drawn around a volume when there is going to be a change in the internal energy of that volume. Anything that passes across the boundary that effects a change in the internal energy needs to be accounted for in the energy balance equation. The volume can be the region surrounding a single atom resonating energy, such as Max Planck defined in 1900; it can be a body of steam or air in a steam engine, such as Sadi Carnot defined in 1824; it can be the body of a tropical cyclone, such as Kerry Emanuel theorized in 1986 in the field of atmospheric thermodynamics; it could also be just one nuclide (i.e. a system of quarks) as hypothesized in quantum thermodynamics.

Boundaries are of four types: fixed, moveable, real, and imaginary. For example, in an engine, a fixed boundary means the piston is locked at its position; as such, a constant volume process occurs. In that same engine, a moveable boundary allows the piston to move in and out. For closed systems, boundaries are real while for open system boundaries are often imaginary.

Generally, thermodynamics distinguishes three classes of systems, defined in terms of what is allowed to cross their boundaries:

Interactions of thermodynamic systems

Type of system	**Mass flow**	**Work**	**Heat**
Open	✓	✓	✓
Closed	✗	✓	✓
Isolated	✗	✗	✗

As time passes in an isolated system, internal differences in the system tend to even out and pressures and temperatures tend to equalize, as do density differences. A system in which all equalizing processes have gone to completion is considered to be in a state of thermodynamic equilibrium.

In thermodynamic equilibrium, a system's properties are, by definition, unchanging in time. Systems in equilibrium are much simpler and easier to understand than systems which are not in equilibrium. Often, when analysing a thermodynamic process, it can be assumed that each intermediate state in the process is at equilibrium. This will also considerably simplify the situation. Thermodynamic processes which develop so slowly as to allow each intermediate step to be an equilibrium state are said to be reversible processes.

States and processes

When a system is at equilibrium under a given set of conditions, it is said to be in a definite thermodynamic state. The state of the system can be described by a number of intensive variables and extensive variables. The properties of the system can be described by an equation of state which specifies the relationship between these variables. State

may be thought of as the instantaneous quantitative description of a system with a set number of variables held constant.

A thermodynamic process may be defined as the energetic evolution of a thermodynamic system proceeding from an initial state to a final state. Typically, each thermodynamic process is distinguished from other processes in energetic character according to what parameters, such as temperature, pressure, or volume, etc., are held fixed. Furthermore, it is useful to group these processes into pairs, in which each variable held constant is one member of a conjugate pair.

Several commonly studied thermodynamic processes are:

- Isobaric process: occurs at constant pressure
- Isochoric process: occurs at constant volume (also called isometric/isovolumetric)
- Isothermal process: occurs at a constant temperature
- Adiabatic process: occurs without loss or gain of energy by heat
- Isentropic process: a reversible adiabatic process, occurs at a constant entropy
- Isenthalpic process: occurs at a constant enthalpy
- Steady state process: occurs without a change in the internal energy

Instrumentation

There are two types of thermodynamic instruments, the **meter** and the **reservoir**. A thermodynamic meter is any device which measures any parameter of a thermodynamic system. In some cases, the thermodynamic parameter is actually defined in terms of an idealized measuring instrument. For example, the zeroth law states that if two bodies are in thermal equilibrium with a third body, they are also in thermal equilibrium with each other. This principle, as noted by James Maxwell in 1872, asserts that it is possible to measure temperature. An idealized thermometer is a sample of an ideal gas at constant pressure. From the ideal gas law $pV=nRT$, the volume of such a sample can be used as an indicator of temperature; in this manner it defines temperature. Although pressure is defined mechanically, a pressure-measuring device, called a barometer may also be constructed from a sample of an ideal gas held at a constant temperature. A calorimeter is a device which is used to measure and define the internal energy of a system.

A thermodynamic reservoir is a system which is so large that it does not appreciably alter its state parameters when brought into contact with the test system. It is used to impose a particular value of a state parameter upon the system. For example, a pressure reservoir is a system at a particular pressure, which imposes that pressure upon any test system that it is mechanically connected to. The Earth's atmosphere is often used as a pressure reservoir.

Conjugate variables

The central concept of thermodynamics is that of energy, the ability to do work. By the First Law, the total energy of a system and its surroundings is conserved. Energy may be transferred into a system by heating, compression, or addition of matter, and extracted from a system by cooling, expansion, or extraction of matter. In mechanics, for example, energy transfer equals the product of the force applied to a body and the resulting displacement.

Conjugate variables are pairs of thermodynamic concepts, with the first being akin to a "force" applied to some thermodynamic system, the second being akin to the resulting "displacement," and the product of the two equalling the amount of energy transferred. The common conjugate variables are:

- Pressure-volume (the mechanical parameters);
- Temperature-entropy (thermal parameters);
- Chemical potential-particle number (material parameters).

Potentials

Thermodynamic potentials are different quantitative measures of the stored energy in a system. Potentials are used to measure energy changes in systems as they evolve from an initial state to a final state. The potential used depends on the constraints of the system, such as constant temperature or pressure. For example, the Helmholtz and Gibbs energies are the energies available in a system to do useful work when the temperature and volume or the pressure and temperature are fixed, respectively.

The five most well known potentials are:

Name	Symbol	Formula	Natural variables
Internal energy	U	$\int (TdS - pdV + \sum_i \mu_i dN_i)$	$S,V,\{N_i\}$
Helmholtz free energy	F, A	$U - TS$	$T,V,\{N_i\}$
Enthalpy	H	$U + pV$	$S,p,\{N_i\}$
Gibbs free energy	G	$U + pV - TS$	$T,p,\{N_i\}$
Landau Potential (Grand potential)	Ω, Φ_G	$U - TS - \sum_i \mu_i N_i$	$T,V,\{\mu_i\}$

where T is the temperature, S the entropy, p the pressure, V the volume, μ the chemical potential, N the number of particles in the system, and i is the count of particles types in the system.

Thermodynamic potentials can be derived from the energy balance equation applied to a thermodynamic system. Other thermodynamic potentials can also be obtained through Legendre transformation.

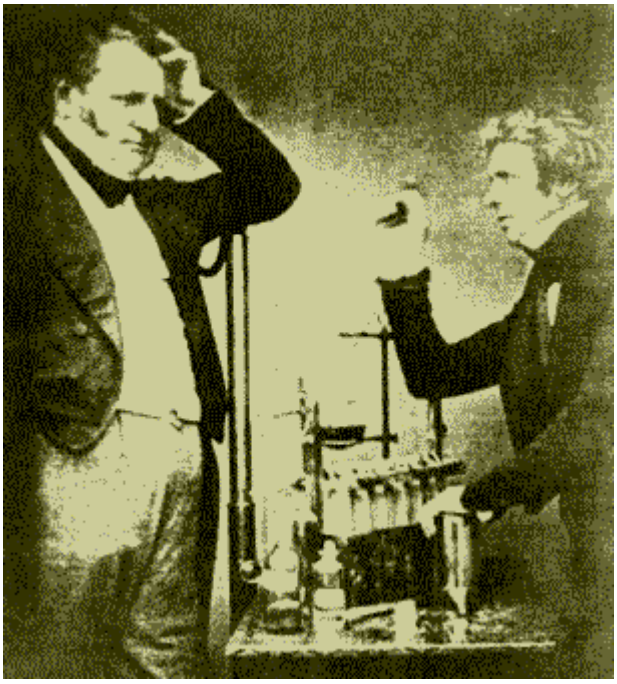
Applied fields

- Atmospheric thermodynamics
- Biological thermodynamics
- Black hole thermodynamics
- Chemical thermodynamics
- Classical thermodynamics
- Equilibrium thermodynamics
- Industrial ecology (re: Exergy)
- Maximum entropy thermodynamics
- Non-equilibrium thermodynamics
- Philosophy of thermal and statistical physics

- Psychrometrics
- Quantum thermodynamics
- Statistical thermodynamics
- Thermoeconomics

Chapter 5

Electrochemistry



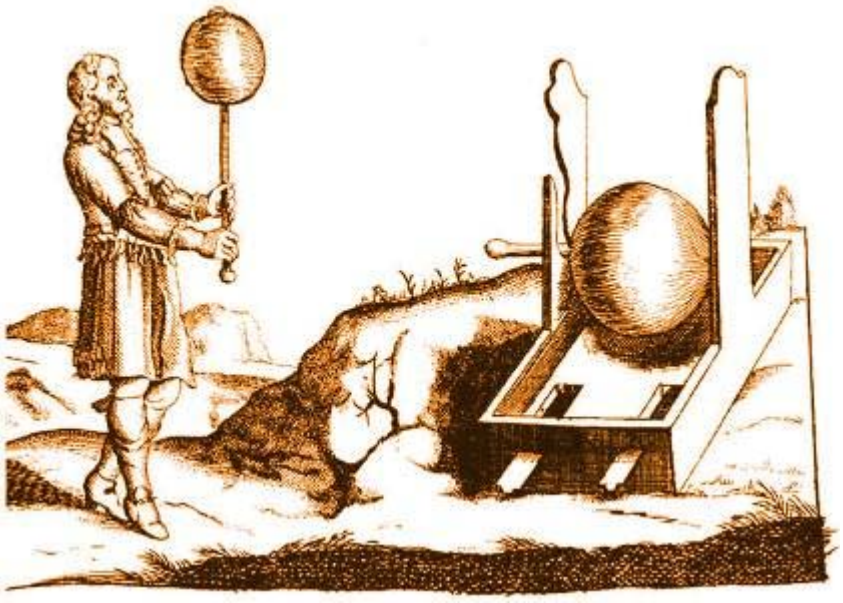
English chemists John Daniell (left) and Michael Faraday (right), both credited as founders of electrochemistry today.

Electrochemistry is a branch of chemistry that studies chemical reactions which take place in a solution at the interface of an electron conductor (a metal or a semiconductor) and an ionic conductor (the electrolyte), and which involve electron transfer between the electrode and the electrolyte or species in solution.

If a chemical reaction is driven by an external applied voltage, as in electrolysis, or if a voltage is created by a chemical reaction as in a battery, it is an *electrochemical* reaction. In contrast, chemical reactions where electrons are transferred between molecules are called oxidation/reduction (redox) reactions. In general, electrochemistry deals with situations where oxidation and reduction reactions are separated in space or time, connected by an external electric circuit to understand each process.

History

16th to 18th century developments

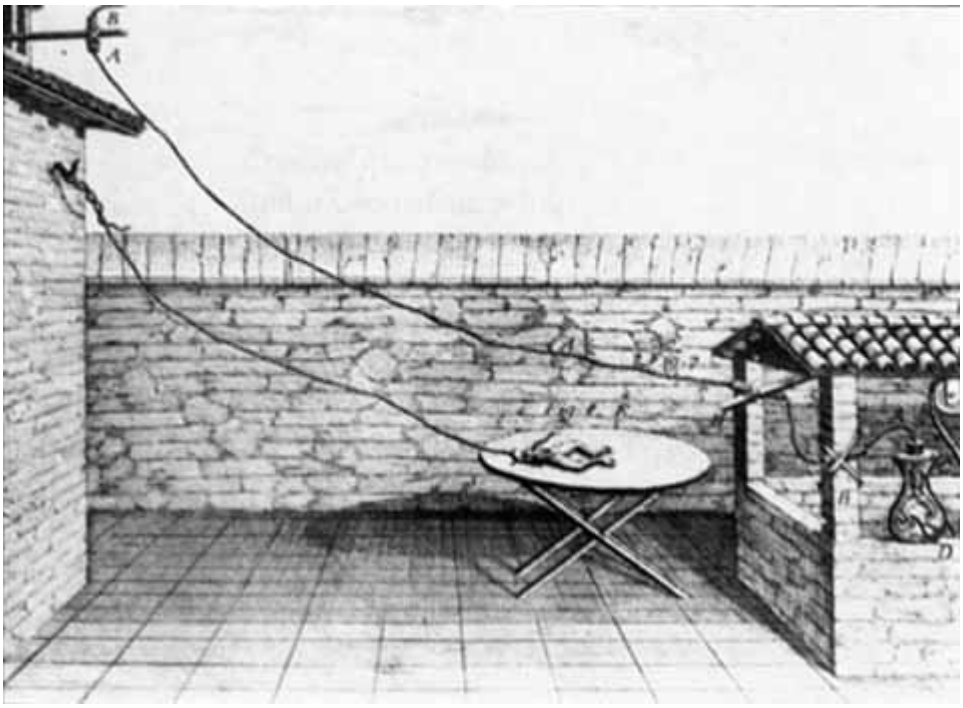


German physicist Otto von Guericke beside his electrical generator while conducting an experiment.

Understanding of electrical matters began in the sixteenth century. During this century the English scientist William Gilbert spent 17 years experimenting with magnetism and, to a lesser extent, electricity. For his work on magnets, Gilbert became known as the *"Father of Magnetism."* He discovered various methods for producing and strengthening magnets.

In 1663 the German physicist Otto von Guericke created the first electric generator, which produced static electricity by applying friction in the machine. The generator was made of a large sulfur ball cast inside a glass globe, mounted on a shaft. The ball was rotated by means of a crank and a static electric spark was produced when a pad was rubbed against the ball as it rotated. The globe could be removed and used as source for experiments with electricity.

By the mid—18th century the French chemist Charles François de Cisternay du Fay had discovered two types of static electricity, and that like charges repel each other whilst unlike charges attract. Du Fay announced that electricity consisted of two fluids: *"vitreous"* (from the Latin for *"glass"*), or positive, electricity; and *"resinous,"* or negative, electricity. This was the *two-fluid theory* of electricity, which was to be opposed by Benjamin Franklin's *one-fluid theory* later in the century.



Late 1780s diagram of Galvani's experiment on frog legs.

Charles-Augustin de Coulomb developed the law of electrostatic attraction in 1785 as an outgrowth of his attempt to investigate the law of electrical repulsions as stated by Joseph Priestley in England.

Italian physicist Alessandro Volta showing his *"battery"* to French emperor Napoleon Bonaparte in the early 19th century.

In the late 18th century the Italian physician and anatomist Luigi Galvani marked the birth of electrochemistry by establishing a bridge between chemical reactions and electricity on his essay *"De Viribus Electricitatis in Motu Musculari Commentarius"* (Latin for Commentary on the Effect of Electricity on Muscular Motion) in 1791 where he proposed a *"nerveo-electrical substance"* on biological life forms.

In his essay Galvani concluded that animal tissue contained a here-to-fore neglected innate, vital force, which he termed *"animal electricity,"* which activated nerves and

muscles spanned by metal probes. He believed that this new force was a form of electricity in addition to the *"natural"* form produced by lightning or by the electric eel and torpedo ray as well as the *"artificial"* form produced by friction (i.e., static electricity).

Galvani's scientific colleagues generally accepted his views, but Alessandro Volta rejected the idea of an *"animal electric fluid,"* replying that the frog's legs responded to differences in metal temper, composition, and bulk. Galvani refuted this by obtaining muscular action with two pieces of the same material.

19th century

Sir Humphry Davy's portrait in the 19th century.

In 1800, William Nicholson and Johann Wilhelm Ritter succeeded in decomposing water into hydrogen and oxygen by electrolysis. Soon thereafter Ritter discovered the process of electroplating. He also observed that the amount of metal deposited and the amount of oxygen produced during an electrolytic process depended on the distance between the electrodes. By 1801 Ritter observed thermoelectric currents and anticipated the discovery of thermoelectricity by Thomas Johann Seebeck.

By the 1810s William Hyde Wollaston made improvements to the galvanic cell. Sir Humphry Davy's work with electrolysis led to the conclusion that the production of electricity in simple electrolytic cells resulted from chemical action and that chemical combination occurred between substances of opposite charge. This work led directly to the isolation of sodium and potassium from their compounds and of the alkaline earth metals from theirs in 1808.

Hans Christian Ørsted's discovery of the magnetic effect of electrical currents in 1820 was immediately recognized as an epoch-making advance, although he left further work on electromagnetism to others. André-Marie Ampère quickly repeated Oestred's experiment, and formulated them mathematically.

In 1821, Estonian-German physicist Thomas Johann Seebeck demonstrated the electrical potential in the juncture points of two dissimilar metals when there is a heat difference between the joints.

In 1827, the German scientist Georg Ohm expressed his law in this famous book *"Die galvanische Kette, mathematisch bearbeitet"* (The Galvanic Circuit Investigated Mathematically) in which he gave his complete theory of electricity.

In 1832, Michael Faraday's experiments led him to state his two laws of electrochemistry. In 1836, John Daniell invented a primary cell in which hydrogen was eliminated in the generation of the electricity. Daniell had solved the problem of polarization. Later results revealed that alloying the amalgamated zinc with mercury would produce a better voltage.

Swedish chemist Svante Arrhenius portrait circa 1880s.

William Grove produced the first fuel cell in 1839. In 1846, Wilhelm Weber developed the electrodynamometer. In 1868, Georges Leclanché patented a new cell which eventually became the forerunner to the world's first widely used battery, the zinc carbon cell.

Svante Arrhenius published his thesis in 1884 on *Recherches sur la conductibilité galvanique des électrolytes* (Investigations on the galvanic conductivity of electrolytes).

From his results the author concluded that electrolytes, when dissolved in water, become to varying degrees split or dissociated into electrically opposite positive and negative ions.

In 1886, Paul Héroult and Charles M. Hall developed an efficient method (the Hall–Héroult process) to obtain aluminium using electrolysis of molten alumina.

In 1894, Friedrich Ostwald concluded important studies of the conductivity and electrolytic dissociation of organic acids.

German scientist Walther Nernst portrait in the 1910s.

Walther Hermann Nernst developed the theory of the electromotive force of the voltaic cell in 1888. In 1889, he showed how the characteristics of the current produced could be used to calculate the free energy change in the chemical reaction producing the current. He constructed an equation, known as Nernst equation, which related the voltage of a cell to its properties.

In 1898, Fritz Haber showed that definite reduction products can result from electrolytic processes if the potential at the cathode is kept constant. In 1898, he explained the reduction of nitrobenzene in stages at the cathode and this became the model for other similar reduction processes.

The 20th century and recent developments

In 1902, The Electrochemical Society (ECS) was founded.

In 1909, Robert Andrews Millikan began a series of experiments to determine the electric charge carried by a single electron.

In 1923, Johannes Nicolaus Brønsted and Martin Lowry published essentially the same theory about how acids and bases behave, using an electrochemical basis.

Arne Tiselius developed the first sophisticated electrophoretic apparatus in 1937 and some years later he was awarded the 1948 Nobel Prize for his work in protein electrophoresis.

A year later, in 1949, the International Society of Electrochemistry (ISE) was founded.

By the 1960s–1970s quantum electrochemistry was developed by Revaz Dogonadze and his pupils.

Principles

Redox reactions

Redox stands for **reduction-oxidation**, and are electrochemical processes involving electron transfer to or from a molecule or ion changing its oxidation state. This reaction can occur through the application of an external voltage or through the release of chemical energy.

Oxidation and reduction

Oxidation and reduction describe the change of oxidation state that takes place in the atoms, ions or molecules involved in an electrochemical reaction. Formally, oxidation state is the hypothetical charge that an atom would have if all bonds to atoms of different elements were 100% ionic. An atom or ion that gives up an electron to another atom or ion has its oxidation state increase, and the recipient of the negatively charged electron has its oxidation state decrease. Oxidation and reduction always occur in a paired fashion such that one species is oxidized when another is reduced. This paired electron transfer is called a redox reaction.

For example, when atomic sodium reacts with atomic chlorine, sodium donates one electron and attains an oxidation state of +1. Chlorine accepts the electron and its oxidation state is reduced to −1. The sign of the oxidation state (positive/negative) actually corresponds to the value of each ion's electronic charge. The attraction of the differently charged sodium and chlorine ions is the reason they then form an ionic bond.

The loss of electrons from an atom or molecule is called oxidation, and the gain of electrons is reduction. This can be easily remembered through the use of mnemonic devices. Two of the most popular are *"OIL RIG"* (Oxidation Is Loss, Reduction Is Gain) and *"LEO"* the lion says *"GER"* (Lose Electrons: Oxidization, Gain Electrons: Reduction). For cases where electrons are shared (covalent bonds) between atoms with large differences in electronegativity, the electron is assigned to the atom with the largest electronegativity in determining the oxidation state.

The atom or molecule which loses electrons is known as the *reducing agent*, or *reductant*, and the substance which accepts the electrons is called the *oxidizing agent*, or *oxidant*. The oxidizing agent is always being reduced in a reaction; the reducing agent is always being oxidized. Oxygen is a common oxidizing agent, but not the only one. Despite the name, an oxidation reaction does not necessarily need to involve oxygen. In fact, a fire can be fed by an oxidant other than oxygen; fluorine fires are often unquenchable, as fluorine is an even stronger oxidant (it has a higher electronegativity) than oxygen.

For reactions involving oxygen, the gain of oxygen implies the oxidation of the atom or molecule to which the oxygen is added (and the oxygen is reduced). In organic compounds, such as butane or ethanol, the loss of hydrogen implies oxidation of the molecule from which it is lost (and the hydrogen is reduced). This follows because the hydrogen donates its electron in covalent bonds with non-metals but it takes the electron along when it is lost. Conversely, loss of oxygen or gain of hydrogen implies reduction.

Balancing redox reactions

Electrochemical reactions in water are better understood by balancing redox reactions using the ion-electron method where H^+ , OH^- ion, H_2O and electrons (to compensate the oxidation changes) are added to cell's half-reactions for oxidation and reduction.

Acidic medium

In acid medium H^+ ions and water are added to half-reactions to balance the overall reaction. For example, when manganese reacts with sodium bismuthate.

Unbalanced reaction: $Mn^{2+}(aq) + NaBiO_3(s) \rightarrow Bi^{3+}(aq) + MnO_4^-(aq)$
Oxidation: $4\ H_2O(l) + Mn^{2+}(aq) \rightarrow MnO_4^-(aq) + 8\ H^+(aq) + 5\ e^-$
Reduction: $2\ e^- + 6\ H^+(aq) + BiO_3^-(s) \rightarrow Bi^{3+}(aq) + 3\ H_2O(l)$

Finally, the reaction is balanced by multiplying the number of electrons from the reduction half reaction to oxidation half reaction and vice versa and adding both half reactions, thus solving the equation.

$8\ H_2O(l) + 2\ Mn^{2+}(aq) \rightarrow 2\ MnO_4^-(aq) + 16\ H^+(aq) + 10\ e^-$
$10\ e^- + 30\ H^+(aq) + 5\ BiO_3^-(s) \rightarrow 5\ Bi^{3+}(aq) + 15\ H_2O(l)$

Reaction balanced:

$14\ H^+(aq) + 2\ Mn^{2+}(aq) + 5\ NaBiO_3(s) \rightarrow 7\ H_2O(l) + 2\ MnO_4^-(aq) + 5\ Bi^{3+}(aq) + 5\ Na^+(aq)$

Basic medium

In basic medium OH^- ions and water are added to half reactions to balance the overall reaction. For example, on reaction between potassium permanganate and sodium sulfite.

Unbalanced reaction: $KMnO_4 + Na_2SO_3 + H_2O \rightarrow MnO_2 + Na_2SO_4 + KOH$
Reduction: $3\ e^- + 2\ H_2O + MnO_4^- \rightarrow MnO_2 + 4\ OH^-$
Oxidation: $2\ OH^- + SO_3^{2-} \rightarrow SO_4^{2-} + H_2O + 2\ e^-$

The same procedure as followed on acid medium by multiplying electrons to opposite half reactions solve the equation thus balancing the overall reaction.

$6\ e^- + 4\ H_2O + 2\ MnO_4^- \rightarrow 2\ MnO_2 + 8\ OH^-$
$6\ OH^- + 3\ SO_3^{2-} \rightarrow 3\ SO_4^{2-} + 3\ H_2O + 6e^-$

Equation balanced:

$2\ KMnO_4 + 3\ Na_2SO_3 + H_2O \rightarrow 2\ MnO_2 + 3\ Na_2SO_4 + 2\ KOH$

Neutral medium

The same procedure as used on acid medium is applied, for example on balancing using electron ion method to complete combustion of propane.

Unbalanced reaction: $C_3H_8 + O_2 \rightarrow CO_2 + H_2O$
Reduction: $4\ H^+ + O_2 + 4\ e^- \rightarrow 2\ H_2O$
Oxidation: $6\ H_2O + C_3H_8 \rightarrow 3\ CO_2 + 20\ e^- + 20\ H^+$

As in acid and basic medium, electrons which were used to compensate oxidation changes are multiplied to opposite half reactions, thus solving the equation.

$20\ H^+ + 5\ O_2 + 20\ e^- \rightarrow 10\ H_2O$
$6\ H_2O + C_3H_8 \rightarrow 3\ CO_2 + 20\ e^- + 20\ H^+$

Equation balanced:

$C_3H_8 + 5\ O_2 \rightarrow 3\ CO_2 + 4\ H_2O$

Electrochemical cells

An electrochemical cell is a device that produces an electric current from energy released by a spontaneous redox reaction. This kind of cell includes the Galvanic cell or Voltaic cell, named after Luigi Galvani and Alessandro Volta, both scientists who conducted several experiments on chemical reactions and electric current during the late 18th century.

Electrochemical cells have two conductive electrodes (the anode and the cathode). The anode is defined as the electrode where oxidation occurs and the cathode is the electrode where the reduction takes place. Electrodes can be made from any sufficiently conductive materials, such as metals, semiconductors, graphite, and even conductive polymers. In between these electrodes is the electrolyte, which contains ions that can freely move.

The Galvanic cell uses two different metal electrodes, each in an electrolyte where the positively charged ions are the oxidized form of the electrode metal. One electrode will undergo oxidation (the anode) and the other will undergo reduction (the cathode). The metal of the anode will oxidize, going from an oxidation state of 0 (in the solid form) to a positive oxidation state and become an ion. At the cathode, the metal ion in solution will accept one or more electrons from the cathode and the ion's oxidation state is reduced to 0. This forms a solid metal that electrodeposits on the cathode. The two electrodes must be electrically connected to each other, allowing for a flow of electrons that leave the metal of the anode and flow through this connection to the ions at the surface of the cathode. This flow of electrons is an electrical current that can be used to do work, such as turn a motor or power a light.

A Galvanic cell whose electrodes are zinc and copper submerged in zinc sulfate and copper sulfate, respectively, is known as a Daniell cell.

Half reactions for a Daniell cell are these:

Zinc electrode (anode): $Zn(s) \rightarrow Zn^{2+}(aq) + 2\ e^-$
Copper electrode (cathode): $Cu^{2+}(aq) + 2\ e^- \rightarrow Cu(s)$

A modern cell stand for electrochemical research. The electrodes attach to high-quality metallic wires, and the stand is attached to a potentiostat/galvanostat (not pictured). A shot glass-shaped container is aerated with a noble gas and sealed with the Teflon block.

In this example, the anode is zinc metal which oxidizes (loses electrons) to form zinc ions in solution, and copper ions accept electrons from the copper metal electrode and the ions deposit at the copper cathode as an electrodeposit. This cell forms a simple battery as it will spontaneously generate a flow of electrical current from the anode to the cathode

through the external connection. This reaction can be driven in reverse by applying a voltage, resulting in the deposition of zinc metal at the anode and formation of copper ions at the cathode.

To provide a complete electric circuit, there must also be an ionic conduction path between the anode and cathode electrolytes in addition to the electron conduction path. The simplest ionic conduction path is to provide a liquid junction. To avoid mixing between the two electrolytes, the liquid junction can be provided through a porous plug that allows ion flow while reducing electrolyte mixing. To further minimize mixing of the electrolytes, a salt bridge can be used which consists of an electrolyte saturated gel in an inverted U-tube. As the negatively charged electrons flow in one direction around this circuit, the positively charged metal ions flow in the opposite direction in the electrolyte.

A voltmeter is capable of measuring the change of electrical potential between the anode and the cathode.

Electrochemical cell voltage is also referred to as electromotive force or emf.

A cell diagram can be used to trace the path of the electrons in the electrochemical cell. For example, here is a cell diagram of a Daniell cell:

$$Zn(s) \mid Zn^{2+}\,(1M) \parallel Cu^{2+}\,(1M) \mid Cu(s)$$

First, the reduced form of the metal to be oxidized at the anode (Zn) is written. This is separated from its oxidized form by a vertical line, which represents the limit between the phases (oxidation changes). The double vertical lines represent the saline bridge on the cell. Finally, the oxidized form of the metal to be reduced at the cathode, is written, separated from its reduced form by the vertical line. The electrolyte concentration is given as it is an important variable in determining the cell potential.

Standard electrode potential

To allow prediction of the cell potential, tabulations of standard electrode potential are available. Such tabulations are referenced to the standard hydrogen electrode (SHE). The standard hydrogen electrode undergoes the reaction

$$2\,H^{+}(aq) + 2\,e^{-} \rightarrow H_2$$

which is shown as reduction but, in fact, the SHE can act as either the anode or the cathode, depending on the relative oxidation/reduction potential of the other electrode/electrolyte combination. The term standard in SHE requires a supply of hydrogen gas bubbled through the electrolyte at a pressure of 1 atm and an acidic electrolyte with H^{+} activity equal to 1 (usually assumed to be $[H^{+}] = 1$ mol/liter).

The SHE electrode can be connected to any other electrode by a salt bridge to form a cell. If the second electrode is also at standard conditions, then the measured cell potential is called the standard electrode potential for the electrode. The standard electrode potential for the SHE is zero, by definition. The polarity of the standard electrode potential provides information about the relative reduction potential of the electrode compared to the SHE. If the electrode has a positive potential with respect to the SHE, then that means it is a strongly reducing electrode which forces the SHE to be the anode (an example is Cu in aqueous $CuSO_4$ with a standard electrode potential of 0.337 V). Conversely, if the

measured potential is negative, the electrode is more oxidizing than the SHE (such as Zn in $ZnSO_4$ where the standard electrode potential is −0.76 V).

Standard electrode potentials are usually tabulated as reduction potentials. However, the reactions are reversible and the role of a particular electrode in a cell depends on the relative oxidation/reduction potential of both electrodes. The oxidation potential for a particular electrode is just the negative of the reduction potential. A standard cell potential can be determined by looking up the standard electrode potentials for both electrodes (sometimes called half cell potentials). The one that is smaller will be the anode and will undergo oxidation. The cell potential is then calculated as the sum of the reduction potential for the cathode and the oxidation potential for the anode.

$$E°_{cell} = E°_{red}(cathode) – E°_{red}(anode) = E°_{red}(cathode) + E°_{oxi}(anode)$$

For example, the standard electrode potential for a copper electrode is:

Cell diagram
$Pt(s) \mid H_2(1\ atm) \mid H^+(1\ M) \parallel Cu^{2+}\ (1\ M) \mid Cu(s)$
$E°_{cell} = E°_{red}(cathode) – E°_{red}(anode)$

At standard temperature, pressure and concentration conditions, the cell's emf (measured by a multimeter) is 0.34 V. By definition, the electrode potential for the SHE is zero. Thus, the Cu is the cathode and the SHE is the anode giving

$$E_{cell} = E°(Cu^{2+}/Cu) – E°(H^+/H_2)$$

Or,

$$E°(Cu^{2+}/Cu) = 0.34\ V$$

Changes in the stoichiometric coefficients of a balanced cell equation will not change $E°_{red}$ value because the standard electrode potential is an intensive property.

Spontaneity of redox reaction

During operation of electrochemical cells, chemical energy is transformed into electrical energy and is expressed mathematically as the product of the cell's emf and the electrical charge transferred through the external circuit.

$$\text{Electrical energy} = E_{cell}C_{trans}$$

where E_{cell} is the cell potential measured in volts (V) and C_{trans} is the cell current integrated over time and measured in coulombs (C); C_{trans} can also be determined by multiplying the total number of electrons transferred (measured in moles) times Faraday's constant (F).

The emf of the cell at zero current is the maximum possible emf. It is used to calculate the maximum possible electrical energy that could be obtained from a chemical reaction. This energy is referred to as electrical work and is expressed by the following equation:

$$W_{max} = W_{electrical} = –nF \cdot E_{cell}$$

where work is defined as positive into the system.

Since the free energy is the maximum amount of work that can be extracted from a system, one can write:

$$\Delta G = -nF \cdot E_{cell}$$

A positive cell potential gives a negative change in Gibbs free energy. This is consistent with the cell production of an electric current flowing from the cathode to the anode through the external circuit. If the current is driven in the opposite direction by imposing an external potential, then work is done on the cell to drive electrolysis.

A spontaneous electrochemical reaction (change in Gibbs free energy less than zero) can be used to generate an electric current, in electrochemical cells. This is the basis of all batteries and fuel cells. For example, gaseous oxygen (O_2) and hydrogen (H_2) can be combined in a fuel cell to form water and energy, typically a combination of heat and electrical energy.

Conversely, non-spontaneous electrochemical reactions can be driven forward by the application of a current at sufficient voltage. The electrolysis of water into gaseous oxygen and hydrogen is a typical example.

The relation between the equilibrium constant, *K*, and the Gibbs free energy for an electrochemical cell is expressed as follows:

$$\Delta G^\circ = -RT \ln(K) = -nF \cdot E^\circ_{cell}$$

Rearranging to express the relation between standard potential and equilibrium constant yields

$$E^o_{cell} = \frac{RT}{nF} \ln K$$

Previous equation can use Briggsian logarithm as shown below:

$$E^o_{cell} = \frac{0.0591V}{n} \log K$$

Cell emf dependency on changes in concentration

Nernst equation

The standard potential of an electrochemical cell requires standard conditions for all of the reactants. When reactant concentrations differ from standard conditions, the cell potential will deviate from the standard potential. In the 20th century German chemist Walther Nernst proposed a mathematical model to determine the effect of reactant concentration on electrochemical cell potential.

In the late 19th century, Josiah Willard Gibbs had formulated a theory to predict whether a chemical reaction is spontaneous based on the free energy

$$\Delta G = \Delta G^\circ + RT \cdot \ln(Q)$$

Here ΔG is change in Gibbs free energy, T is absolute temperature, R is the gas constant and Q is reaction quotient.

Gibbs' key contribution was to formalize the understanding of the effect of reactant concentration on spontaneity.

Based on Gibbs' work, Nernst extended the theory to include the contribution from electric potential on charged species. As shown in the previous section, the change in Gibbs free energy for an electrochemical cell can be related to the cell potential. Thus, Gibbs' theory becomes

$$nF\Delta E = nF\Delta E^\circ - RT \ln(Q)$$

Here n is the number of electrons/mole product, F is the Faraday constant (coulombs/mole), and ΔE is cell potential.

Finally, Nernst divided through by the amount of charge transferred to arrive at a new equation which now bears his name:

$$\Delta E = \Delta E^\circ - (RT/nF)\ln(Q)$$

Assuming standard conditions (T = 25 °C) and R = 8.3145 J/(K·mol), the equation above can be expressed on base—10 logarithm as shown below:

$$\Delta E = \Delta E^o - \frac{0.05916\ \mathrm{V}}{\mathrm{n}} \log\ \mathrm{Q}$$

Concentration cells

A concentration cell is an electrochemical cell where the two electrodes are the same material, the electrolytes on the two half-cells involve the same ions, but the electrolyte concentration differs between the two half-cells.

For example an electrochemical cell, where two copper electrodes are submerged in two copper(II) sulfate solutions, whose concentrations are 0.05 M and 2.0 M, connected through a salt bridge. This type of cell will generate a potential that can be predicted by the Nernst equation. Both electrodes undergo the same chemistry (although the reaction proceeds in reverse at the cathode)

$$Cu^{2+}(aq) + 2\ e^- \rightarrow Cu(s)$$

Le Chatelier's principle indicates that the reaction is more favorable to reduction as the concentration of Cu^{2+} ions increases. Reduction will take place in the cell's compartment where concentration is higher and oxidation will occur on the more dilute side.

The following cell diagram describes the cell mentioned above:

$$Cu(s) \mid Cu^{2+}\ (0.05\ M) \parallel Cu^{2+}\ (2.0\ M) \mid Cu(s)$$

Where the half cell reactions for oxidation and reduction are:

Oxidation: $Cu(s) \rightarrow Cu^{2+}$ (0.05 M) + 2 e^-
Reduction: Cu^{2+} (2.0 M) + 2 $e^- \rightarrow Cu(s)$
Overall reaction: Cu^{2+} (2.0 M) $\rightarrow Cu^{2+}$ (0.05 M)

The cell's emf is calculated through Nernst equation as follows:

$$E = E^o - \frac{0.05916V}{2} log \frac{[Cu^{2+}]_{diluted}}{[Cu^{2+}]_{concentrated}}$$

The value of E° in this kind of cell is zero, as electrodes and ions are the same in both half-cells.

After replacing values from the case mentioned, it is possible to calculate cell's potential:

$$E = 0 - \frac{0.05916V}{2} log \frac{0.05}{2.0} = 0.0474V$$

or by:

$$E = 0 - \frac{0.0257V}{2} ln \frac{0.05}{2.0} = 0.0474V$$

However, this value is only approximate, as reaction quotient is defined in terms of ion activities which can be approximated with the concentrations as calculated here.

The Nernst equation plays an important role in understanding electrical effects in cells and organelles. Such effects include nerve synapses and cardiac beat as well as the resting potential of a somatic cell.

Battery

A battery is a number of cells combined and is used to supply electrical energy that is stored chemically. The term battery is often, and incorrectly, used to describe a single cell. In a battery, cells are usually wired in series to increase the supply voltage but sometimes wired in parallel to allow greater current to be supplied. Batteries are optimized to produce a constant electric current for as long as possible. Although the cells discussed previously are useful for theoretical purposes and some laboratory experiments, the large internal resistance of the salt bridge make them inappropriate battery technologies. Various alternative battery technologies have been commercialized as discussed next.

Dry cell

Zinc carbon battery diagram.

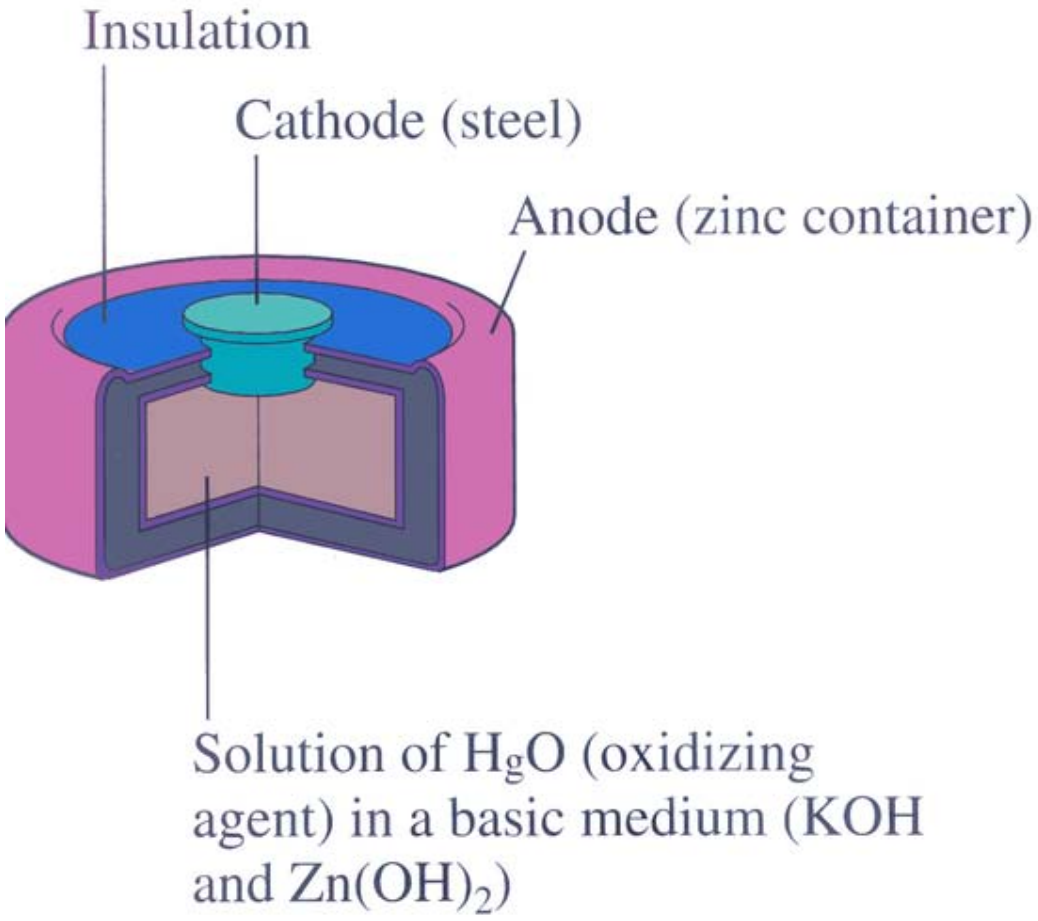
Dry cells do not have a fluid electrolyte. Instead, they use a moist electrolyte paste. Leclanché's cell is a good example of this, where the anode is a zinc container surrounded by a thin layer of manganese dioxide and a moist electrolyte paste of ammonium chloride and zinc chloride mixed with starch. The cell's cathode is represented by a carbon bar inserted on the cell's electrolyte, usually placed in the middle.

Leclanché's simplified half reactions are shown below:

Anode: $Zn(s) \rightarrow Zn^{2+}(aq) + 2\ e^-$
Cathode: $2\ NH_4^+(aq) + 2\ MnO_2(s) + 2\ e^- \rightarrow Mn_2O_3(s) + 2\ NH_3(aq) + H_2O(l)$
Overall reaction: $Zn(s) + 2\ NH_4^+(aq) + 2\ MnO_2(s) \rightarrow Zn^{2+}(aq) + Mn_2O_3(s) + 2\ NH_3(aq) + H_2O(l)$

The voltage obtained from the zinc-carbon battery is around 1.5 V.

Mercury battery



Cutaway view of a mercury battery.

This battery first appeared in the early 1940s. The mercury battery has many applications in medicine and electronics. The battery consists of a steel—made container in the shape of a cylinder acting as the cathode, where an amalgamated anode of mercury and zinc is surrounded by a stronger alkaline electrolyte and a paste of zinc oxide and mercury(II) oxide.

Mercury battery half reactions are shown below:

Anode: $Zn(Hg) + 2\ OH^-(aq) \rightarrow ZnO(s) + H_2O(l) + 2\ e^-$
Cathode: $HgO(s) + H_2O(l) + 2\ e^- \rightarrow Hg(l) + 2\ OH^-(aq)$
Overall reaction: $Zn(Hg) + HgO(s) \rightarrow ZnO(s) + Hg(l)$

There are no changes in the electrolyte's composition when the cell works. Such batteries provide 1.35 V of direct current.

Lead-acid battery

A sealed lead-acid battery.

The lead-acid battery used in automobiles, consists of a series of six identical cells assembled in series. Each cell has a lead anode and a cathode made from lead dioxide packed in a metal plaque. Cathode and anode are submerged in a solution of sulfuric acid acting as the electrolyte.

Lead-acid battery half cell reactions are shown below:

Anode: $Pb(s) + SO_4^{2-}(aq) \rightarrow PbSO_4(s) + 2\ e^-$
Cathode: $PbO_2(s) + 4\ H^+(aq) + SO_4^{2-}(aq) + 2\ e^- \rightarrow PbSO_4(s) + 2\ H_2O(l)$
Overall reaction: $Pb(s) + PbO_2(s) + 4\ H^+(aq) + 2\ SO_4^{2-}(aq) \rightarrow 2\ PbSO_4(s) + 2\ H_2O(l)$

At standard conditions, each cell may produce a potential of 2 V, hence overall voltage produced is 12 V. Differing from mercury and zinc-carbon batteries, lead-acid batteries are rechargeable. If an external voltage is supplied to the battery it will produce an electrolysis of the products in the overall reaction (discharge), thus recovering initial components which made the battery work.

Lithium rechargeable battery

Instead of an aqueous electrolyte or a moist electrolyte paste, a solid state battery operates using a solid electrolyte. Lithium polymer batteries are an example of this; a graphite bar acts as the anode, a bar of lithium cobaltate acts as the cathode, and a polymer, swollen with a lithium salt, allows the passage of ions and serves as the electrolyte. In this cell, the carbon in the anode can reversibly form a lithium-carbon alloy. Upon discharging, lithium ions spontaneously leave the lithium cobaltate cathode and travel through the polymer and into the carbon anode forming the alloy. This flow of positive lithium ions is the electrical current that the battery provides. By charging the cell, the lithium dealloys and travels back into the cathode. The advantage of this kind of battery is that Lithium possess the highest negative value of standard reduction potential. It is also a light metal and therefore less mass is required to generate 1 mole of electrons. Lithium ion battery technologies are widely used in portable electronic devices because they have high energy storage density and are rechargeable. These technologies show promise for future automotive applications, with new materials such as iron phosphates and lithium vanadates.

Flow battery

Most batteries have all of the electrolyte and electrodes within a single housing. A flow battery is unusual in that the majority of the electrolyte, including dissolved reactive species, is stored in separate tanks. The electrolytes are pumped through a reactor, which houses the electrodes, when the battery is charged or discharged. Because the electrodes are preserved, and the electrolyte combusted, a "flow battery" is better described as a reversible fuel cell.

These types of batteries are typically used for large-scale energy storage (kWh – multi MWh). Of the several different types that have been developed, some are of current commercial interest, including the iron/chromium flow battery, vanadium redox battery and zinc-bromine flow battery.

Fuel cells

Fossil fuels are used in power plants to supply electrical needs, however their conversion into electricity is an inefficient process. The most efficient electrical power plant may only convert about 40% of the original chemical energy into electricity when burned or processed.

To enhance electrical production, scientists have developed fuel cells where combustion is replaced by electrochemical methods, similar to a battery but requiring continuous replenishment of the reactants consumed.

The most popular is the oxygen-hydrogen fuel cell, where two inert electrodes (porous electrodes of nickel and nickel oxide) are placed in an electrolytic solution such as hot caustic potash, in both compartments (anode and cathode) gaseous hydrogen and oxygen are bubbled into solution.

Oxygen-hydrogen fuel cell reactions are shown bellow:

Anode: $2\ H_2(g) \rightarrow 4\ H^+ + 4\ e^-$
Cathode: $O_2(g) + 4\ e^- + 4\ H^+ \rightarrow 2\ H_2O(l)$
Overall reaction: $2\ H_2(g) + O_2(g) \rightarrow 2\ H_2O(l)$

The overall reaction is identical to hydrogen combustion. Oxidation and reduction take place in the anode and cathode separately. This is similar to the electrode used in the cell for measuring standard reduction potential which has a double function acting as electrical conductors providing a surface required to decomposition of the molecules into atoms before electron transferring, thus named electrocatalysts. Platinum, nickel, and rhodium are good electrocatalysts.

Corrosion

Corrosion is the term applied to steel rust caused by an electrochemical process. Most people are likely familiar with the corrosion of iron, in the form of reddish rust. Other examples include the black tarnish on silver, and red or green corrosion that may appear on copper and its alloys, such as brass. The cost of replacing metals lost to corrosion is in the multi-billions of dollars per year.

Iron corrosion

For iron rust to occur the metal has to be in contact with oxygen and water, although chemical reactions for this process are relatively complex and not all of them are completely understood, it is believed the causes are the following: Electron transferring (reduction-oxidation)

> One area on the surface of the metal acts as the anode, which is where the oxidation (corrosion) occurs. At the anode, the metal gives up electrons.
> $Fe(s) \rightarrow Fe^{2+}(aq) + 2\ e^-$
> Electrons are transferred from iron reducing oxygen in the atmosphere into water on the cathode, which is placed in another region of the metal.
> $O_2(g) + 4\ H^+(aq) + 4\ e^- \rightarrow 2\ H_2O(l)$
> Global reaction for the process:
> $2\ Fe(s) + O_2(g) + 4\ H^+(aq) \rightarrow 2\ Fe^{2+}(aq) + 2\ H_2O(l)$
> Standard emf for iron rusting:
> $E° = E°_{cathode} - E°_{anode}$
> $E° = 1.23V - (-0.44\ V) = 1.67\ V$

Iron corrosion takes place on acid medium; H^+ ions come from reaction between carbon dioxide in the atmosphere and water, forming carbonic acid. Fe^{2+} ions oxides, following this equation:

$$4\ Fe^{2+}(aq) + O_2(g) + (4+2x)H_2O(l) \rightarrow 2\ Fe_2O_3{\cdot}xH_2O + 8\ H^+(aq)$$

Iron(III) oxide hydrated is known as rust. The concentration of water associated with iron oxide varies, thus chemical representation is presented as $Fe_2O_3{\cdot}xH_2O$. The electric circuit works as passage of electrons and ions occurs, thus if an electrolyte is present it will facilitate oxidation, this explains why rusting is quicker on salt water.

Corrosion of common metals

Coinage metals, such as copper and silver, slowly corrode through use. A patina of green-blue copper carbonate forms on the surface of copper with exposure to the water and carbon dioxide in the air. Silver coins or cutlery that are exposed to high sulfur foods such as eggs or the low levels of sulfur species in the air develop a layer of black Silver sulfide.

Gold and platinum are extremely difficult to oxidize under normal circumstances, and require exposure to a powerful chemical oxidizing agent such as aqua regia.

Some common metals oxidize extremely rapidly in air. Titanium and aluminium oxidize instantaneously in contact with the oxygen in the air. These metals form an extremely thin layer of oxidized metal on the surface. This thin layer of oxide protects the underlying layers of the metal from the air preventing the entire metal from oxidizing. These metals are used in applications where corrosion resistance is important. Iron, in contrast, has an oxide that forms in air and water, called rust, that does not stop the further oxidation of the iron. Thus iron left exposed to air and water will continue to rust until all of the iron is oxidized.

Prevention of corrosion

Attempts to save a metal from becoming anodic are of two general types. Anodic regions dissolve and destroy the structural integrity of the metal.

While it is almost impossible to prevent anode/cathode formation, if a non-conducting material covers the metal, contact with the electrolyte is not possible and corrosion will not occur.

Coating

Metals can be coated with paint or other less conductive metals (*passivation*). This prevents the metal surface from being exposed to electrolytes. Scratches exposing the metal substrate will result in corrosion. The region under the coating adjacent to the scratch acts as the anode of the reaction.

Sacrificial anodes

A method commonly used to protect a structural metal is to attach a metal which is more anodic than the metal to be protected. This forces the structural metal to be cathodic, thus spared corrosion. It is called *"sacrificial"* because the anode dissolves and has to be replaced periodically.

Zinc bars are attached to various locations on steel ship hulls to render the ship hull cathodic. The zinc bars are replaced periodically. Other metals, such as magnesium, would work very well but zinc is the least expensive useful metal.

To protect pipelines, an ingot of buried or exposed magnesium (or zinc) is buried beside the pipeline and is connected electrically to the pipe above ground. The pipeline is forced to be a cathode and is protected from being oxidized and rusting. The magnesium anode is sacrificed. At intervals new ingots are buried to replace those lost.

Electrolysis

The spontaneous redox reactions of a conventional battery produce electricity through the different chemical potentials of the cathode and anode in the electrolyte. However, electrolysis requires an external source of electrical energy to induce a chemical reaction, and this process takes place in a compartment called an electrolytic cell.

Electrolysis of molten sodium chloride

When molten, the salt sodium chloride can be electrolyzed to yield metallic sodium and gaseous chlorine. Industrially this process takes place in a special cell named Down's cell. The cell is connected to an electrical power supply, allowing electrons to migrate from the power supply to the electrolytic cell.

Reactions that take place at Down's cell are the following:

Anode (oxidation): $2\ Cl^- \rightarrow Cl_2(g) + 2\ e^-$
Cathode (reduction): $2\ Na^+(l) + 2\ e^- \rightarrow 2\ Na(l)$
Overall reaction: $2\ Na^+ + 2\ Cl^-(l) \rightarrow 2\ Na(l) + Cl_2(g)$

This process can yield large amounts of metallic sodium and gaseous chlorine, and is widely used on mineral dressing and metallurgy industries.

The emf for this process is approximately −4 V indicating a (very) non-spontaneous process. In order for this reaction to occur the power supply should provide at least a potential of 4 V. However, larger voltages must be used for this reaction to occur at a high rate.

Electrolysis of water

Water can be converted to its component elemental gasses, H_2 and O_2 through the application of an external voltage. Water doesn't decompose into hydrogen and oxygen spontaneously as the Gibbs free energy for the process at standard conditions is about 474.4 kJ. The decomposition of water into hydrogen and oxygen can be performed in an electrolytic cell. In it, a pair of inert electrodes usually made of platinum immersed in water act as anode and cathode in the electrolytic process. The electrolysis starts with the application of an external voltage between the electrodes. This process will not occur except at extremely high voltages without an electrolyte such as sodium chloride or sulfuric acid (most used 0.1 M).

Bubbles from the gases will be seen near both electrodes. The following half reactions describe the process mentioned above:

Anode (oxidation): $2\ H_2O(l) \rightarrow O_2(g) + 4\ H^+(aq) + 4\ e^-$
Cathode (reduction): $2\ H_2O(g) + 2\ e^- \rightarrow H_2(g) + 2\ OH^-(aq)$
Overall reaction: $2\ H_2O(l) \rightarrow 2\ H_2(g) + O_2(g)$

Although strong acids may be used in the apparatus, the reaction will not net consume the acid. While this reaction will work at any conductive electrode at a sufficiently large potential, platinum catalyzes both hydrogen and oxygen formation, allowing for relatively mild voltages (~2 V depending on the pH).

Electrolysis of aqueous solutions

Electrolysis in an aqueous is a similar process as mentioned in electrolysis of water. However, it is considered to be a complex process because the contents in solution have to be analyzed in half reactions, whether reduced or oxidized.

Electrolysis of a solution of sodium chloride

The presence of water in a solution of sodium chloride must be examined in respect to its reduction and oxidation in both electrodes. Usually, water is electrolysed as mentioned in electrolysis of water yielding *gaseous oxygen in the anode* and gaseous hydrogen in the cathode. On the other hand, sodium chloride in water dissociates in Na^+ and Cl^- ions, cation, which is the positive ion, will be attracted to the cathode (+), thus reducing the sodium ion. The anion will then be attracted to the anode (–) oxidizing chloride ion.

The following half reactions describes the process mentioned:

1. Cathode: $Na^+(aq) + e^- \rightarrow Na(s)$ $\quad E°_{red} = –2.71$ V
2. Anode: $2\ Cl^-(aq) \rightarrow Cl_2(g) + 2\ e^-$ $\quad E°_{red} = +1.36$ V
3. Cathode: $2\ H_2O(l) + 2\ e^- \rightarrow H_2(g) + 2\ OH^-(aq)$ $\quad E°_{red} = –0.83$ V

4. Anode: $2\ H_2O(l) \rightarrow O_2(g) + 4\ H^+(aq) + 4\ e^-$ $E°_{red} = +1.23$ V

Reaction 1 is discarded as it has the most negative value on standard reduction potential thus making it less thermodynamically favorable in the process.

When comparing the reduction potentials in reactions 2 and 4, the reduction of chloride ion is favored. Thus, if the Cl^- ion is favored for reduction, then the water reaction is favored for oxidation producing gaseous oxygen, however experiments show gaseous chlorine is produced and not oxygen.

Although the initial analysis is correct, there is another effect that can happen, known as the overvoltage effect. Additional voltage is sometimes required, beyond the voltage predicted by the $E°_{cell}$. This may be due to kinetic rather than thermodynamic considerations. In fact, it has been proven that the activation energy for the chloride ion is very low, hence favorable in kinetic terms. In other words, although the voltage applied is thermodynamically sufficient to drive electrolysis, the rate is so slow that to make the process proceed in a reasonable time frame, the voltage of the external source has to be increased (hence, overvoltage).

Finally, reaction 3 is favorable because it describes the proliferation of OH^- ions thus letting a probable reduction of H^+ ions less favorable an option.

The overall reaction for the process according to the analysis would be the following:

Anode (oxidation): $2\ Cl^-(aq) \rightarrow Cl_2(g) + 2\ e^-$
Cathode (reduction): $2\ H_2O(l) + 2\ e^- \rightarrow H_2(g) + 2\ OH^-(aq)$
Overall reaction: $2\ H_2O + 2\ Cl^-(aq) \rightarrow H_2(g) + Cl_2(g) + 2\ OH^-(aq)$

As the overall reaction indicates, the concentration of chloride ions is reduced in comparison to OH^- ions (whose concentration increases). The reaction also shows the production of gaseous hydrogen, chlorine and aqueous sodium hydroxide.

Quantitative electrolysis and Faraday's Laws

Quantitative aspects of electrolysis were originally developed by Michael Faraday in 1834. Faraday is also credited to have coined the terms *electrolyte*, electrolysis, among many others while he studied quantitative analysis of electrochemical reactions. Also he was an advocate of the law of conservation of energy.

First law

Faraday concluded after several experiments on electrical current in non-spontaneous process, the mass of the products yielded on the electrodes was proportional to the value of current supplied to the cell, the length of time the current existed, and the molar mass of the substance analyzed. In other words, the amount of a substance deposited on each electrode of an electrolytic cell is directly proportional to the quantity of electricity passed through the cell.

Below a simplified equation of Faraday's first law:

$$m = \frac{1}{96485\ (\mathrm{C} \cdot \mathrm{mol}^{-1})} \cdot \frac{QM}{n}$$

Where

m is the mass of the substance produced at the electrode (in grams),
Q is the total electric charge that passed through the solution (in coulombs),
n is the valence number of the substance as an ion in solution (electrons per ion),
M is the molar mass of the substance (in grams per mole).

Second law

Faraday devised the laws of chemical electrodeposition of metals from solutions in 1857. He formulated the second law of electrolysis stating *"the amounts of bodies which are equivalent to each other in their ordinary chemical action have equal quantities of electricity naturally associated with them."* In other terms, the quantities of different elements deposited by a given amount of electricity are in the ratio of their chemical equivalent weights.

An important aspect of the second law of electrolysis is electroplating which together with the first law of electrolysis, has a significant number of applications in the industry, as when used to protect metals to avoid corrosion.

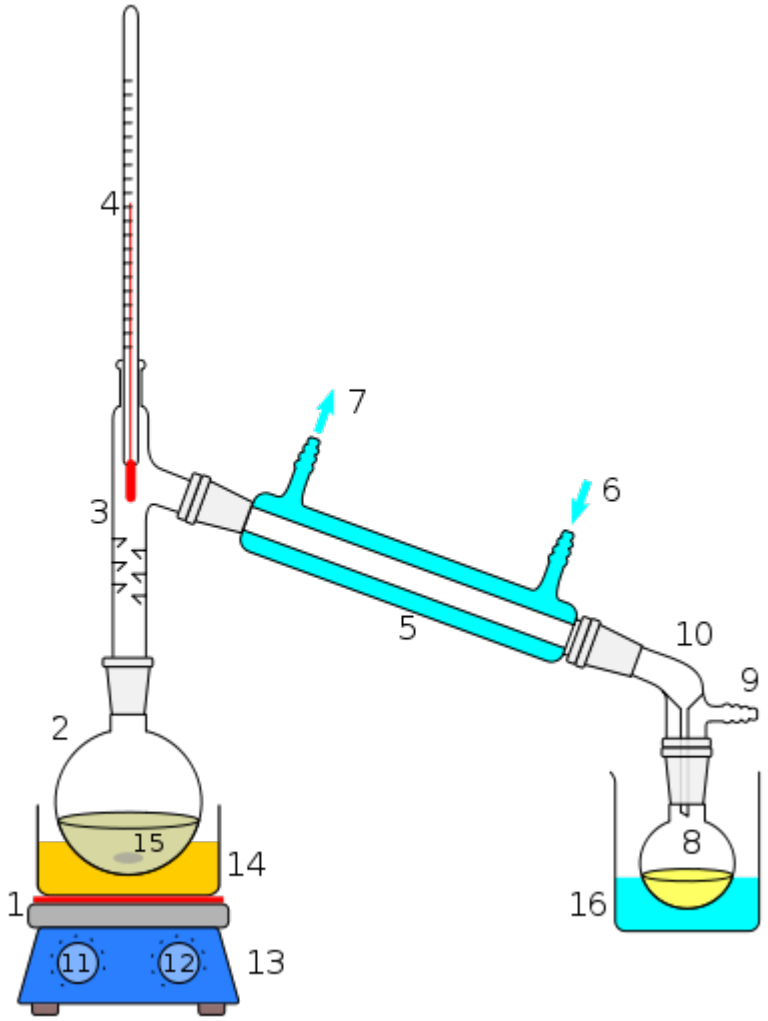
Applications

There are various extremely important electrochemical processes in both nature and industry, like the coating of objects with metals or metal oxides through electrodeposition and the detection of alcohol in drunken drivers through the redox reaction of ethanol. The generation of chemical energy through photosynthesis is inherently an electrochemical process, as is production of metals like aluminum and titanium from their ores. Certain diabetes blood sugar meters measure the amount of glucose in the blood through its redox potential.

The nervous impulses in neurons are based on electric current generated by the movement of sodium and potassium ions into and out of cells, and certain animals like eels can generate a powerful voltage from certain cells that can disable much larger animals.

Chapter 6

Distillation

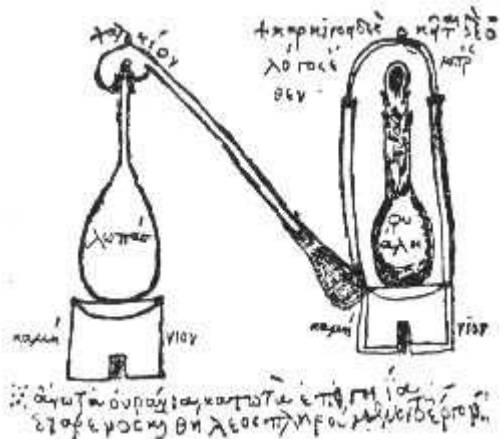


Laboratory display of distillation: **1:** A heating device **2:** Still pot **3:** Still head **4:** Thermometer/Boiling point temperature **5:** Condenser **6:** Cooling water in **7:** Cooling water out **8:** Distillate/receiving flask **9:** Vacuum/gas inlet **10:** Still receiver **11:** Heat control **12:** Stirrer speed control **13:** Stirrer/heat plate **14:** Heating (Oil/sand) bath **15:** Stirring means e.g.(shown), boiling chips or mechanical stirrer **16:** Cooling bath.

Distillation is a method of separating mixtures based on differences in their volatilities in a boiling liquid mixture. Distillation is a unit operation, or a physical separation process, and not a chemical reaction.

Commercially, distillation has a number of applications. It is used to separate crude oil into more fractions for specific uses such as transport, power generation and heating. Water is distilled to remove impurities, such as salt from seawater. Air is distilled to separate its components—notably oxygen, nitrogen, and argon—for industrial use. Distillation of fermented solutions has been used since ancient times to produce distilled beverages with a higher alcohol content. The premises where distillation is carried out, especially distillation of alcohol, are known as a **distillery**.

History



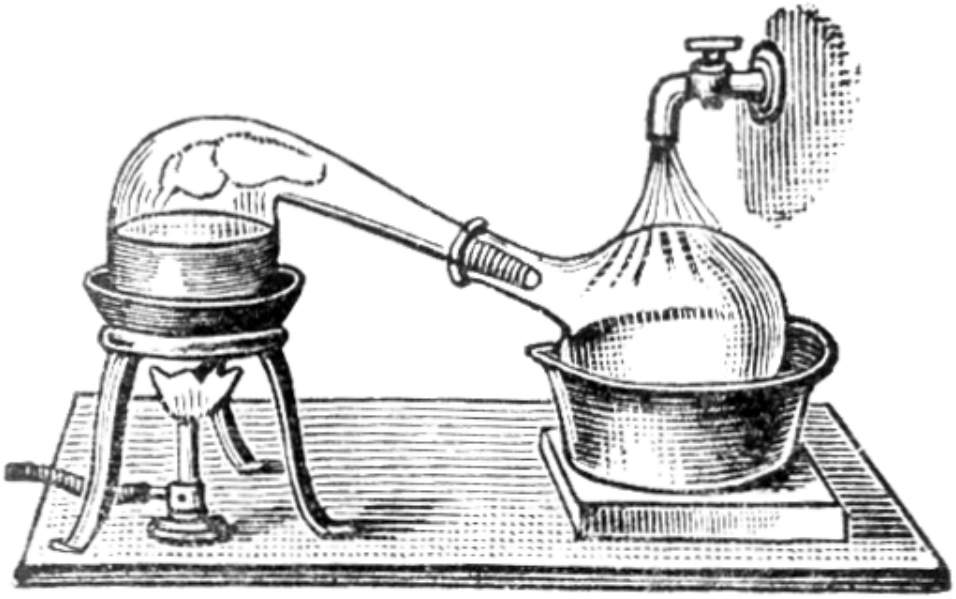
Distillation apparatus of Zosimus, from Marcelin Berthelot, *Collection des anciens alchimistes grecs* (3 vol., Paris, 1887-1888).

The first clear evidence of distillation comes from Greek alchemists working in Alexandria in the first century AD. Distilled water has been known since at least ca. 200 AD, when Alexander of Aphrodisias described the process. Arabs learned the process from the Egyptians and used it extensively in their chemical experiments. They introduced the apparatuses (such as the alembic, still, and retort) which were able to fully purify chemical substances.

Clear evidence of the distillation of alcohol comes from the School of Salerno in the 12th century. Fractional distillation was developed by Tadeo Alderotti in the 13th century.

In 1500, German alchemist Hieronymus Braunschweig published *Liber de arte destillandi* (The Book of the Art of Distillation) the first book solely dedicated to the subject of distillation, followed in 1512 by a much expanded version. In 1651, John French published The Art of Distillation the first major English compendium of practice, though it has been claimed that much of it derives from Braunschweig's work. This includes diagrams with people in them showing the industrial rather than bench scale of the operation.

A retort.



Distillation

Old Ukrainian vodka still

As alchemy evolved into the science of chemistry, vessels called retorts became used for distillations. Both alembics and retorts are forms of glassware with long necks pointing to the side at a downward angle which acted as air-cooled condensers to condense the distillate and let it drip downward for collection. Later, copper alembics were invented. Riveted joints were often kept tight by using various mixtures, for instance a dough made of rye flour. These alembics often featured a cooling system around the beak, using cold water for instance, which made the condensation of alcohol more efficient. These were called pot stills. Today, the retorts and pot stills have been largely supplanted by more efficient distillation methods in most industrial processes. However, the pot still is still widely used for the elaboration of some fine alcohols such as cognac, Scotch whisky, tequila and some vodkas. Pot stills made of various materials (wood, clay, stainless steel) are also used by bootleggers in various countries. Small pot stills are also sold for the domestic production of flower water or essential oils.

Early forms of distillation were batch processes using one vaporization and one condensation. Purity was improved by further distillation of the condensate. Greater volumes were processed by simply repeating the distillation. Chemists were reported to carry out as many as 500 to 600 distillations in order to obtain a pure compound.

In the early 19th century the basics of modern techniques including pre-heating and reflux were developed, particularly by the French, then in 1830 a British Patent was issued to Aeneas Coffey for a whiskey distillation column, which worked continuously and may be regarded as the archetype of modern petrochemical units. In 1877, Ernest Solvay was granted a U.S. Patent for a tray column for ammonia distillation and the same and subsequent years saw developments of this theme for oil and spirits.

With the emergence of chemical engineering as a discipline at the end of the 19th century, scientific rather than empirical methods could be applied. The developing petroleum industry in the early 20th century provided the impetus for the development of accurate design methods such as the McCabe-Thiele method and the Fenske equation. The availability of powerful computers has also allowed direct computer simulation of distillation columns.

Applications of distillation

The application of distillation can roughly be divided in four groups: laboratory scale, industrial distillation, distillation of herbs for perfumery and medicinals (herbal distillate), and food processing. The latter two are distinctively different from the former two in that in the processing of beverages, the distillation is not used as a true purification method but more to transfer all volatiles from the source materials to the distillate.

The main difference between laboratory scale distillation and industrial distillation is that laboratory scale distillation is often performed batch-wise, whereas industrial distillation often occurs continuously. In batch distillation, the composition of the source material, the vapors of the distilling compounds and the distillate change during the distillation. In batch distillation, a still is charged (supplied) with a batch of feed mixture, which is then separated into its component fractions which are collected sequentially from most volatile to less volatile, with the bottoms (remaining least or non-volatile fraction) removed at the end. The still can then be recharged and the process repeated.

In continuous distillation, the source materials, vapors, and distillate are kept at a constant composition by carefully replenishing the source material and removing fractions from

both vapor and liquid in the system. This results in a better control of the separation process.

Idealized distillation model

The boiling point of a liquid is the temperature at which the vapor pressure of the liquid equals the pressure in the liquid, enabling bubbles to form without being crushed. A special case is the normal boiling point, where the vapor pressure of the liquid equals the ambient atmospheric pressure.

It is a common misconception that in a liquid mixture at a given pressure, each component boils at the boiling point corresponding to the given pressure and the vapors of each component will collect separately and purely. This, however, does not occur even in an idealized system. Idealized models of distillation are essentially governed by Raoult's law and Dalton's law, and assume that vapor-liquid equilibria are attained.

Raoult's law assumes that a component contributes to the total vapor pressure of the mixture in proportion to its percentage of the mixture and its vapor pressure when pure, or succinctly: partial pressure equals mole fraction multiplied by vapor pressure when pure. If one component changes another component's vapor pressure, or if the volatility of a component is dependent on its percentage in the mixture, the law will fail.

Dalton's law states that the total vapor pressure is the sum of the vapor pressures of each individual component in the mixture. When a multi-component liquid is heated, the vapor pressure of each component will rise, thus causing the total vapor pressure to rise. When the total vapor pressure reaches the pressure surrounding the liquid, boiling occurs and liquid turns to gas throughout the bulk of the liquid. Note that a mixture with a given composition has one boiling point at a given pressure, when the components are mutually soluble.

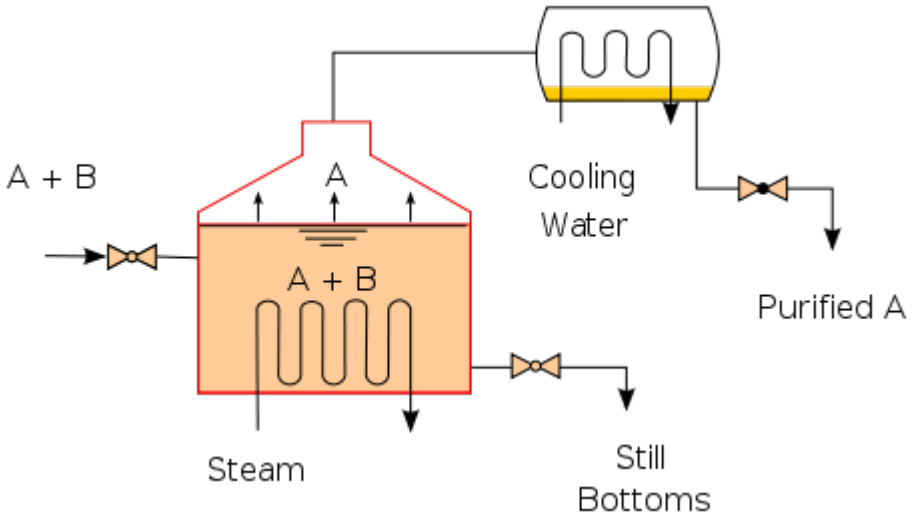
An implication of one boiling point is that lighter components never cleanly "boil first". At boiling point, all volatile components boil, but for a component, its percentage in the vapor is the same as its percentage of the total vapor pressure. Lighter components have a higher partial pressure and thus are concentrated in the vapor, but heavier volatile components also have a (smaller) partial pressure and necessarily evaporate also, albeit being less concentrated in the vapor. Indeed, batch distillation and fractionation succeed by varying the composition of the mixture. In batch distillation, the batch evaporates, which changes its composition; in fractionation, liquid higher in the fractionation column contains more lights and boils at lower temperatures.

The idealized model is accurate in the case of chemically similar liquids, such as benzene and toluene. In other cases, severe deviations from Raoult's law and Dalton's law are observed, most famously in the mixture of ethanol and water. These compounds, when heated together, form an azeotrope, which is a composition with a boiling point higher or lower than the boiling point of each separate liquid. Virtually all liquids, when mixed and heated, will display azeotropic behaviour. Although there are computational methods that can be used to estimate the behavior of a mixture of arbitrary components, the only way to obtain accurate vapor-liquid equilibrium data is by measurement.

It is not possible to *completely* purify a mixture of components by distillation, as this would require each component in the mixture to have a zero partial pressure. If ultra-pure products are the goal, then further chemical separation must be applied. When a binary

mixture is evaporated and the other component, e.g. a salt, has zero partial pressure for practical purposes, the process is simpler and is called evaporation in engineering.

Batch distillation



A batch still showing the separation of A and B.

Heating an ideal mixture of two volatile substances A and B (with A having the higher volatility, or lower boiling point) in a batch distillation setup (such as in an apparatus depicted in the opening figure) until the mixture is boiling results in a vapor above the liquid which contains a mixture of A and B. The ratio between A and B in the vapor will be different from the ratio in the liquid: the ratio in the liquid will be determined by how the original mixture was prepared, while the ratio in the vapor will be enriched in the more volatile compound, A. The vapor goes through the condenser and is removed from the system. This in turn means that the ratio of compounds in the remaining liquid is now different from the initial ratio (i.e. more enriched in B than the starting liquid).

The result is that the ratio in the liquid mixture is changing, becoming richer in component B. This causes the boiling point of the mixture to rise, which in turn results in a rise in the temperature in the vapor, which results in a changing ratio of A : B in the gas phase (as distillation continues, there is an increasing proportion of B in the gas phase). This results in a slowly changing ratio A : B in the distillate.

If the difference in vapor pressure between the two components A and B is large (generally expressed as the difference in boiling points), the mixture in the beginning of the distillation is highly enriched in component A, and when component A has distilled off, the boiling liquid is enriched in component B.

Continuous distillation

Continuous distillation is an ongoing distillation in which a liquid mixture is continuously (without interruption) fed into the process and separated fractions are removed continuously as output streams as time passes during the operation. Continuous distillation produces at least two output fractions, including at least one volatile distillate fraction, which has boiled and been separately captured as a vapor condensed to a liquid. There is always a bottoms (or residue) fraction, which is the least volatile residue that has not been separately captured as a condensed vapor.

Continuous distillation differs from batch distillation in the respect that concentrations should not change over time. Continuous distillation can be run at a steady state for an arbitrary amount of time. For any source material of specific composition, the main variables that affect the purity of products in continuous distillation are the reflux ratio and the number of theoretical equilibrium stages (practically, the number of trays or the height of packing). Reflux is a flow from the condenser back to the column, which generates a recycle that allows a better separation with a given number of trays. Equilibrium stages are ideal steps where compositions achieve vapor-liquid equilibrium, repeating the separation process and allowing better separation given a reflux ratio. A column with a high reflux ratio may have fewer stages, but it refluxes a large amount of liquid, giving a wide column with a large holdup. Conversely, a column with a low reflux ratio must have a large number of stages, thus requiring a taller column.

General improvements

Both batch and continuous distillations can be improved by making use of a fractionating column on top of the distillation flask. The column improves separation by providing a larger surface area for the vapor and condensate to come into contact. This helps it remain at equilibrium for as long as possible. The column can even consist of small subsystems ('trays' or 'dishes') which all contain an enriched, boiling liquid mixture, all with their own vapor-liquid equilibrium.

There are differences between laboratory-scale and industrial-scale fractionating columns, but the principles are the same. Examples of laboratory-scale fractionating columns (in increasing efficiency) include:

- Air condenser
- Vigreux column (usually laboratory scale only)
- Packed column (packed with glass beads, metal pieces, or other chemically inert material)
- Spinning band distillation system.

Laboratory scale distillation

Laboratory scale distillations are almost exclusively run as batch distillations. The device used in distillation, sometimes referred to as a *still*, consists at a minimum of a **reboiler** or *pot* in which the source material is heated, a **condenser** in which the heated vapour is cooled back to the liquid state, and a **receiver** in which the concentrated or purified liquid, called the **distillate**, is collected. Several laboratory scale techniques for distillation exist.

Simple distillation

In **simple distillation**, all the hot vapors produced are immediately channeled into a condenser that cools and condenses the vapors. Therefore, the distillate will not be pure - its composition will be identical to the composition of the vapors at the given temperature and pressure, and can be computed from Raoult's law.

As a result, simple distillation is usually used only to separate liquids whose boiling points differ greatly (rule of thumb is 25 °C), or to separate liquids from involatile solids or oils. For these cases, the vapor pressures of the components are usually sufficiently different that Raoult's law may be neglected due to the insignificant contribution of the

less volatile component. In this case, the distillate may be sufficiently pure for its intended purpose.

Fractional distillation

For many cases, the boiling points of the components in the mixture will be sufficiently close that Raoult's law must be taken into consideration. Therefore, **fractional distillation** must be used in order to separate the components well by repeated vaporization-condensation cycles within a packed fractionating column. This separation, by successive distillations, is also referred to as **rectification** .

As the solution to be purified is heated, its vapors rise to the fractionating column. As it rises, it cools, condensing on the condenser walls and the surfaces of the packing material. Here, the condensate continues to be heated by the rising hot vapors; it vaporizes once more. However, the composition of the fresh vapors are determined once again by Raoult's law. Each vaporization-condensation cycle (called a *theoretical plate*) will yield a purer solution of the more volatile component. In reality, each cycle at a given temperature does not occur at exactly the same position in the fractionating column; *theoretical plate* is thus a concept rather than an accurate description.

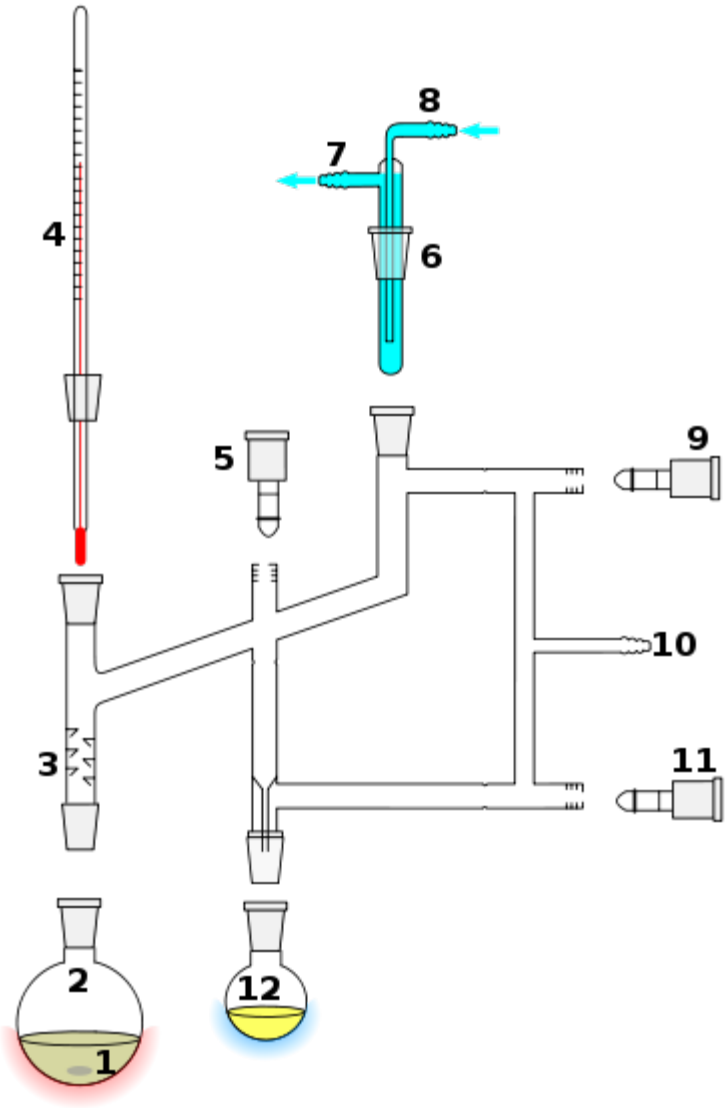
More theoretical plates lead to better separations. A spinning band distillation system uses a spinning band of Teflon or metal to force the rising vapors into close contact with the descending condensate, increasing the number of theoretical plates.

Steam distillation

Like vacuum distillation, **steam distillation** is a method for distilling compounds which are heat-sensitive. This process involves bubbling steam through a heated mixture of the raw material. By Raoult's law, some of the target compound will vaporize (in accordance with its partial pressure). The vapor mixture is cooled and condensed, usually yielding a layer of oil and a layer of water.

Steam distillation of various aromatic herbs and flowers can result in two products; an essential oil as well as a watery herbal distillate. The essential oils are often used in perfumery and aromatherapy while the watery distillates have many applications in aromatherapy, food processing and skin care.

Dimethyl sulfoxide usually boils at 189 °C. Under a vacuum, it distills off into the receiver at only 70 °C.



Perkin triangle distillation setup
1: Stirrer bar/anti-bumping granules **2:** Still pot **3:** Fractionating column **4:** Thermometer/Boiling point temperature **5:** Teflon tap 1 **6:** Cold finger **7:** Cooling water out **8:** Cooling water in **9:** Teflon tap 2 **10:** Vacuum/gas inlet **11:** Teflon tap 3 **12:** Still receiver

Vacuum distillation

Some compounds have very high boiling points. To boil such compounds, it is often better to lower the pressure at which such compounds are boiled instead of increasing the temperature. Once the pressure is lowered to the vapor pressure of the compound (at the given temperature), boiling and the rest of the distillation process can commence. This technique is referred to as **vacuum distillation** and it is commonly found in the laboratory in the form of the rotary evaporator.

This technique is also very useful for compounds which boil beyond their decomposition temperature at atmospheric pressure and which would therefore be decomposed by any attempt to boil them under atmospheric pressure.

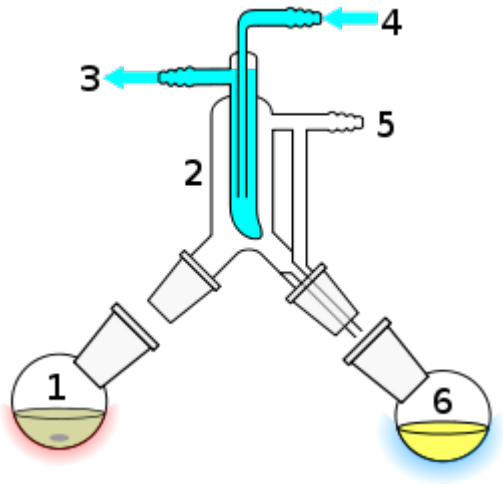
Molecular distillation is vacuum distillation below the pressure of 0.01 torr. 0.01 torr is one order of magnitude above high vacuum, where fluids are in the free molecular flow regime, i.e. the mean free path of molecules is comparable to the size of the equipment. The gaseous phase no longer exerts significant pressure on the substance to be evaporated, and consequently, rate of evaporation no longer depends on pressure. That is, because the continuum assumptions of fluid dynamics no longer apply, mass transport is governed by molecular dynamics rather than fluid dynamics. Thus, a short path between the hot surface and the cold surface is necessary, typically by suspending a hot plate covered with a film of feed next to a cold plate with a line of sight in between. Molecular distillation is used industrially for purification of oils.

Air-sensitive vacuum distillation

Some compounds have high boiling points as well as being air sensitive. A simple vacuum distillation system as exemplified above can be used, whereby the vacuum is replaced with an inert gas after the distillation is complete. However, this is a less satisfactory system if one desires to collect fractions under a reduced pressure. To do this a "cow" or "pig" adaptor can be added to the end of the condenser, or for better results or for very air sensitive compounds a Perkin triangle apparatus can be used.

The Perkin triangle, has means via a series of glass or Teflon taps to allows fractions to be isolated from the rest of the still, without the main body of the distillation being removed from either the vacuum or heat source, and thus can remain in a state of reflux. To do this, the sample is first isolated from the vacuum by means of the taps, the vacuum over the sample is then replaced with an inert gas (such as nitrogen or argon) and can then be stoppered and removed. A fresh collection vessel can then be added to the system, evacuated and linked back into the distillation system via the taps to collect a second fraction, and so on, until all fractions have been collected.

Short path distillation



Short path vacuum distillation apparatus with vertical condenser (cold finger), to minimize the distillation path; **1:** Still pot with stirrer bar/anti-bumping granules **2:** Cold finger - bent to direct condensate **3:** Cooling water out **4:** cooling water in **5:** Vacuum/gas inlet **6:** Distillate flask/distillate.

Short path distillation is a distillation technique that involves the distillate travelling a short distance, often only a few centimeters, and is normally done at reduced pressure. A

classic example would be a distillation involving the distillate travelling from one glass bulb to another, without the need for a condenser separating the two chambers. This technique is often used for compounds which are unstable at high temperatures or to purify small amounts of compound. The advantage is that the heating temperature can be considerably lower (at reduced pressure) than the boiling point of the liquid at standard pressure, and the distillate only has to travel a short distance before condensing. A short path ensures that little compound is lost on the sides of the apparatus. The Kugelrohr is a kind of a short path distillation apparatus which often contain multiple chambers to collect distillate fractions.

Other types

- The process of reactive distillation involves using the reaction vessel as the still. In this process, the product is usually significantly lower-boiling than its reactants. As the product is formed from the reactants, it is vaporized and removed from the reaction mixture. This technique is an example of a continuous vs. a batch process; advantages include less downtime to charge the reaction vessel with starting material, and less workup.
- Pervaporation is a method for the separation of mixtures of liquids by partial vaporization through a non-porous membrane.
- Extractive distillation is defined as distillation in the presence of a miscible, high boiling, relatively non-volatile component, the solvent, that forms no azeotrope with the other components in the mixture.
- Flash evaporation (or partial evaporation) is the partial vaporization that occurs when a saturated liquid stream undergoes a reduction in pressure by passing through a throttling valve or other throttling device. This process is one of the simplest unit operations, being equivalent to a distillation with only one equilibrium stage.
- Codistillation is distillation which is performed on mixtures in which the two compounds are not miscible.

The unit process of evaporation may also be called "distillation":

- In rotary evaporation a vacuum distillation apparatus is used to remove bulk solvents from a sample. Typically the vacuum is generated by a water aspirator or a membrane pump.
- In a kugelrohr a short path distillation apparatus is typically used (generally in combination with a (high) vacuum) to distill high boiling (> 300 °C) compounds. The apparatus consists of an oven in which the compound to be distilled is placed, a receiving portion which is outside of the oven, and a means of rotating the sample. The vacuum is normally generated by using a high vacuum pump.

Other uses:

- Dry distillation or destructive distillation, despite the name, is not truly distillation, but rather a chemical reaction known as pyrolysis in which solid substances are heated in an inert or reducing atmosphere and any volatile fractions, containing high-boiling liquids and products of pyrolysis, are collected. The destructive distillation of wood to give methanol is the root of its common name - *wood alcohol*.
- Freeze distillation is an analogous method of purification using freezing instead of evaporation. It is not truly distillation, but a recrystallization where the product is

the mother liquor, and does not produce products equivalent to distillation. This process is used in the production of ice beer and ice wine to increase ethanol and sugar content, respectively. It is also used to produce applejack. Unlike distillation, freeze distillation concentrates poisonous congeners rather than removing them.

Azeotropic distillation

Interactions between the components of the solution create properties unique to the solution, as most processes entail nonideal mixtures, where Raoult's law does not hold. Such interactions can result in a constant-boiling **azeotrope** which behaves as if it were a pure compound (i.e., boils at a single temperature instead of a range). At an azeotrope, the solution contains the given component in the same proportion as the vapor, so that evaporation does not change the purity, and distillation does not effect separation. For example, ethyl alcohol and water form an azeotrope of 95.6% at 78.1 °C.

If the azeotrope is not considered sufficiently pure for use, there exist some techniques to break the azeotrope to give a pure distillate. This set of techniques are known as **azeotropic distillation**. Some techniques achieve this by "jumping" over the azeotropic composition (by adding an additional component to create a new azeotrope, or by varying the pressure). Others work by chemically or physically removing or sequestering the impurity. For example, to purify ethanol beyond 95%, a drying agent or a (desiccant such as potassium carbonate) can be added to convert the soluble water into insoluble water of crystallization. Molecular sieves are often used for this purpose as well.

Immiscible liquids, such as water and toluene, easily form azeotropes. Commonly, these azeotropes are referred to as a low boiling azeotrope because the boiling point of the azeotrope is lower than the boiling point of either pure component. The temperature and composition of the azeotrope is easily predicted from the vapor pressure of the pure components, without use of Raoult's law. The azeotrope is easily broken in a distillation set-up by using a liquid-liquid separator (a decanter) to separate the two liquid layers that are condensed overhead. Only one of the two liquid layers is refluxed to the distillation set-up.

High boiling azeotropes, such as a 20 weight percent mixture of hydrochloric acid in water, also exist. As implied by the name, the boiling point of the azeotrope is greater than the boiling point of either pure component.

To break azeotropic distillations and cross distillation boundaries, such as in the DeRosier Problem, it is necessary to increase the composition of the light key in the distillate.

Breaking an azeotrope with unidirectional pressure manipulation

The boiling points of components in an azeotrope overlap to form a band. By exposing an azeotrope to a vacuum or positive pressure, it's possible to bias the boiling point of one component away from the other by exploiting the differing vapour pressure curves of each; the curves may overlap at the azeotropic point, but are unlikely to be remain identical further along the pressure axis either side of the azeotropic point. When the bias is great enough, the two boiling points no longer overlap and so the azeotropic band disappears.

This method can remove the need to add other chemicals to a distillation, but it has two potential drawbacks.

Under negative pressure, power for a vacuum source is needed and the reduced boiling points of the distillates requires that the condenser be run cooler to prevent distillate vapours being lost to the vacuum source. Increased cooling demands will often require additional energy and possibly new equipment or a change of coolant.

Alternatively, if positive pressures are required, standard glassware can not be used, energy must be used for pressurization and there is a higher chance of side reactions occurring in the distillation, such as decomposition, due to the higher temperatures required to effect boiling.

A unidirectional distillation will rely on a pressure change in one direction, either positive or negative.

Pressure-swing distillation

Pressure-swing distillation is essentially the same as the unidirectional distillation used to break azeotropic mixtures, but here both positive and negative pressures may be employed.

This has an important impact on the selectivity of the distillation and allows a chemist to optimize a process such that fewer extremes of pressure and temperature are required and less energy is consumed. This is particularly important in commercial applications.

Pressure-swing distillation is employed during the industrial purification of ethyl acetate after its catalytic synthesis from ethanol.

Industrial distillation

Typical industrial distillation towers

Large scale **industrial distillation** applications include both batch and continuous fractional, vacuum, azeotropic, extractive, and steam distillation. The most widely used industrial applications of continuous, steady-state fractional distillation are in petroleum refineries, petrochemical and chemical plants and natural gas processing plants.

Industrial distillation is typically performed in large, vertical cylindrical columns known as **distillation towers** or **distillation columns** with diameters ranging from about 65 centimeters to 16 meters and heights ranging from about 6 meters to 90 meters or more. When the process feed has a diverse composition, as in distilling crude oil, liquid outlets at intervals up the column allow for the withdrawal of different *fractions* or products having different boiling points or boiling ranges. The "lightest" products (those with the lowest boiling point) exit from the top of the columns and the "heaviest" products (those with the highest boiling point) exit from the bottom of the column and are often called the **bottoms**.

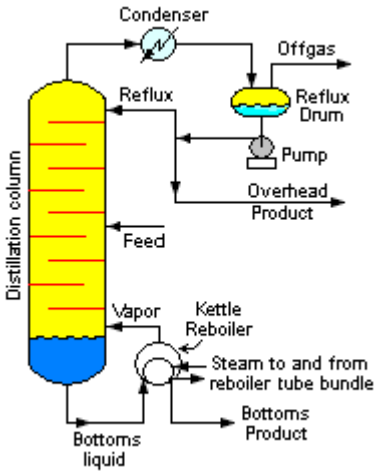
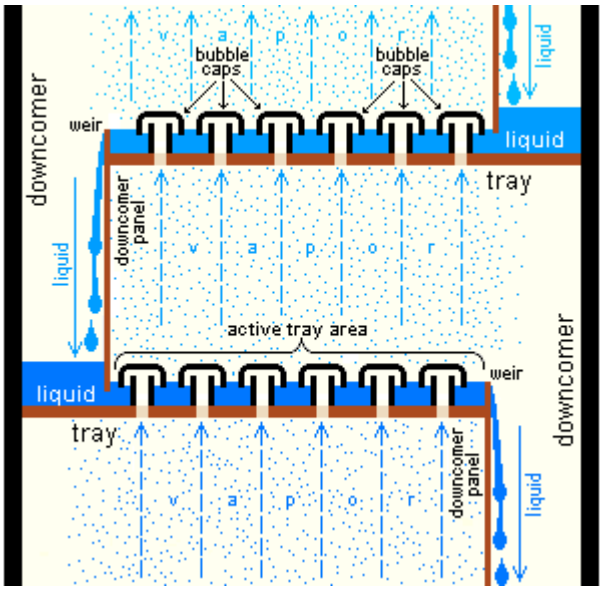


Diagram of a typical industrial distillation tower

Industrial towers use reflux to achieve a more complete separation of products. Reflux refers to the portion of the condensed overhead liquid product from a distillation or fractionation tower that is returned to the upper part of the tower as shown in the schematic diagram of a typical, large-scale industrial distillation tower. Inside the tower, the downflowing reflux liquid provides cooling and condensation of the upflowing vapors thereby increasing the efficiency of the distillation tower. The more reflux that is provided for a given number of theoretical plates, the better the tower's separation of lower boiling materials from higher boiling materials. Alternatively, the more reflux that is provided for a given desired separation, the fewer the number of theoretical plates required.

Such industrial fractionating towers are also used in air separation, producing liquid oxygen, liquid nitrogen, and high purity argon. Distillation of chlorosilanes also enables the production of high-purity silicon for use as a semiconductor.



Section of an industrial distillation tower showing detail of trays with bubble caps

Design and operation of a distillation tower depends on the feed and desired products. Given a simple, binary component feed, analytical methods such as the McCabe-Thiele method or the Fenske equation can be used. For a multi-component feed, simulation

models are used both for design and operation. Moreover, the efficiencies of the vapor-liquid contact devices (referred to as "plates" or "trays") used in distillation towers are typically lower than that of a theoretical 100% efficient equilibrium stage. Hence, a distillation tower needs more trays than the number of theoretical vapor-liquid equilibrium stages.

In modern industrial uses, generally a packing material is used in the column instead of trays, especially when low pressure drops across the column are required, as when operating under vacuum.

Large-scale, industrial vacuum distillation column

This packing material can either be random dumped packing (1-3" wide) such as Raschig rings or structured sheet metal. Liquids tend to wet the surface of the packing and the vapors pass across this wetted surface, where mass transfer takes place. Unlike conventional tray distillation in which every tray represents a separate point of vapor-liquid equilibrium, the vapor-liquid equilibrium curve in a packed column is continuous. However, when modeling packed columns, it is useful to compute a number of "theoretical stages" to denote the separation efficiency of the packed column with respect to more traditional trays. Differently shaped packings have different surface areas and void space between packings. Both of these factors affect packing performance.

Another factor in addition to the packing shape and surface area that affects the performance of random or structured packing is the liquid and vapor distribution entering the packed bed. The number of theoretical stages required to make a given separation is calculated using a specific vapor to liquid ratio. If the liquid and vapor are not evenly distributed across the superficial tower area as it enters the packed bed, the liquid to vapor ratio will not be correct in the packed bed and the required separation will not be achieved. The packing will appear to not be working properly. The height equivalent of a theoretical plate (HETP) will be greater than expected. The problem is not the packing itself but the mal-distribution of the fluids entering the packed bed. Liquid mal-distribution is more frequently the problem than vapor. The design of the liquid distributors used to introduce the feed and reflux to a packed bed is critical to making the

packing perform to it maximum efficiency. Methods of evaluating the effectiveness of a liquid distributor to evenly distribute the liquid entering a packed bed can be found in references. Considerable work as been done on this topic by Fractionation Research, Inc. (commonly known as FRI).

Multi-effect distillation

The goal of multi-effect distillation is to increase the energy efficiency of the process, for use in desalination, or in some cases one stage in the production of ultrapure water. The number of effects is proportional to the $kW \cdot h/m^3$ of water recovered figure, and refers to the volume of water recovered per unit of energy compared with single-effect distillation. One effect is roughly 636 $kW \cdot h/m^3$.

- Multi-stage flash distillation Can achieve more than 20 effects with thermal energy input, as mentioned here.
- Vapor compression evaporation Commercial large-scale units can achieve around 72 effects with electrical energy input, according to manufacturers.

There are many other types of multi-effect distillation processes, including one referred to as simply multi-effect distillation (MED), in which multiple chambers, with intervening heat exchangers, are employed.

Distillation in food processing

Distilled beverages

Carbohydrate-containing plant materials are allowed to ferment, producing a dilute solution of ethanol in the process. Spirits such as whiskey and rum are prepared by distilling these dilute solutions of ethanol. Components other than ethanol, including water, esters, and other alcohols, are collected in the condensate, which account for the flavor of the beverage.

Gallery

Chemistry on its beginnings used retorts as laboratory equipment exclusively for distillation processes.

A simple set-up to distill dry and oxygen-free toluene.

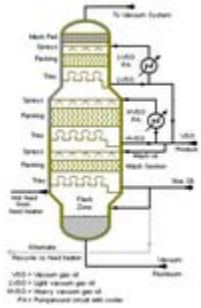


Diagram of an industrial-scale vacuum distillation column as commonly used in oil refineries

A rotary evaporator is able to distill solvents more quickly at lower temperatures through the use of a vacuum.

Distillation using semi-microscale apparatus. The jointless design eliminates the need to fit pieces together. The pear-shaped flask allows the last drop of residue to be removed, compared with a similarly-sized round-bottom flask The small holdup volume prevents losses. A pig is used to channel the various distillates into three receiving flasks. If necessary the distillation can be carried out under vacuum using the vacuum adapter at the pig.