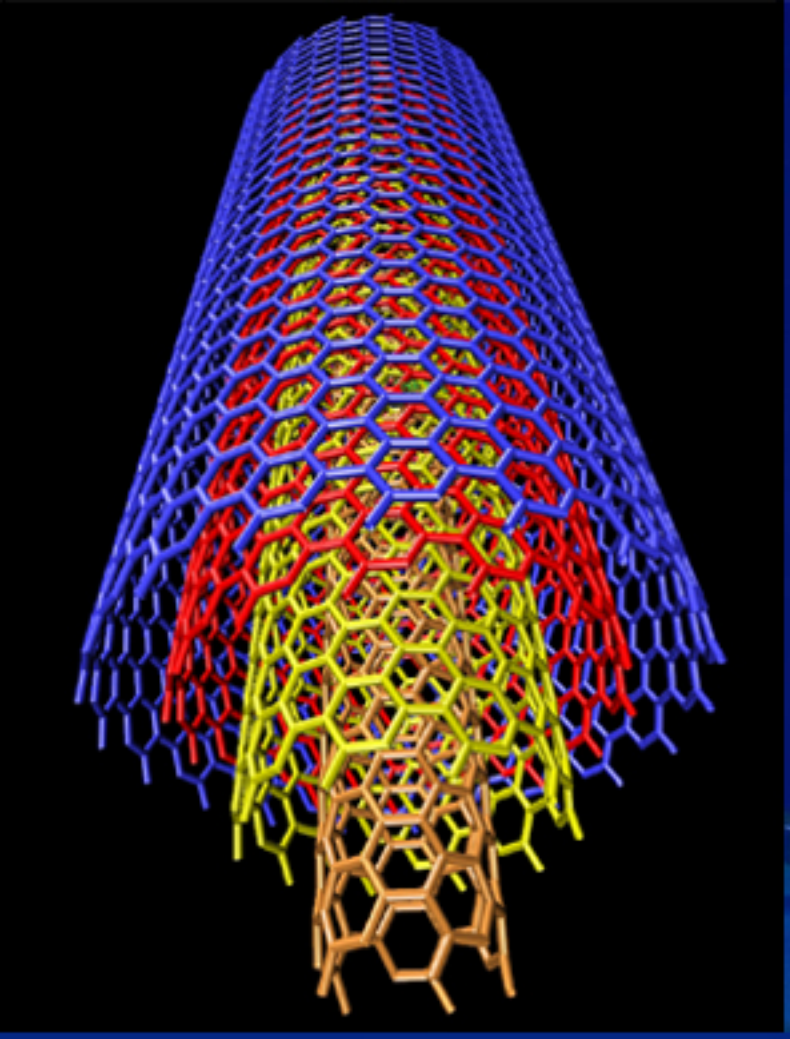


Handbook of Molecular Electronics and Microelectromechanical Systems

Hillary Narvaez

Rylee Phelps

First Edition, 2012

ISBN 978-81-323-0924-6

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

Table of Contents

Chapter- 1

Introduction to Molecular Electronics

Molecular electronics (sometimes called *moletronics*) involves the study and application of molecular building blocks for the fabrication of electronic components. This includes both passive and active electronic components. Molecular electronics is a branch of nanotechology.

An interdisciplinary pursuit, molecular electronics spans physics, chemistry, and materials science. The unifying feature is the use of molecular building blocks for the fabrication of electronic components. This includes both passive (e.g. resistive wires) and active components such as transistors and molecular-scale switches. Due to the prospect of size reduction in electronics offered by molecular-level control of properties, molecular electronics has aroused much excitement both in science fiction and among scientists. Molecular electronics provides means to extend Moore's Law beyond the foreseen limits of small-scale conventional silicon integrated circuits.

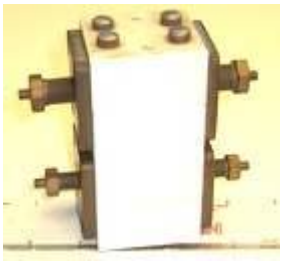
Molecular electronics is split into two related but separate subdisciplines: *molecular materials for electronics* utilizes the properties of the molecules to affect the bulk properties of a material, while *molecular scale electronics* focuses on single-molecule applications.

Concept genesis and theory

In their 1940's discussion of so-called "donor-acceptor" complexes, Robert Mulliken and Albert Szent-Gyorgi advanced the concept of charge transfer in molecules. They subsequently further refined the study of both charge transfer and energy transfer in molecules. Likewise, a 1974 paper from Mark Ratner and Ari Aviram [1] illustrated a theoretical molecular rectifier. In 1988, Aviram described in detail a theoretical single-molecule field-effect transistor. Further concepts were proposed by Forrest Carter of the Naval Research Laboratory, including single-molecule logic gates.

These were all theoretical constructs and not concrete devices. The *direct* measurement of the electronic characteristics of individual molecules awaited the development of methods for making molecular-scale electrical contacts. This was no easy task. Thus, the

first experiment directly-measuring the conductance of a single molecule was only reported in 1997 by Mark Reed and co-workers. Since then, this branch of the field has progressed rapidly. Likewise, as it has become possible to measure such properties directly, the theoretical predictions of the early workers have been substantially confirmed.



Voltage-controlled switch, a molecular electronic device from 1974. From Smithsonian Chip collection

However, while mostly operating in the quantum realm of less than 100 nanometers, "molecular" electronic processes can collectively manifest on a macro scale. Examples include quantum tunneling, negative resistance, phonon-assisted hopping, polarons, and the like. Thus, macro-scale active organic electronic devices were described decades before molecular-scale ones. E.g., in 1974, John McGinness and his coworkers described the putative "first experimental demonstration of an operating molecular electronic device". This was a voltage-controlled switch. As its active element, this device used DOPA melanin, an oxidized mixed polymer of polyacetylene, polypyrrole, and polyaniline. The "ON" state of this switch exhibited almost metallic conductivity.

Since the 1970s, scientists have developed an entire panoply of new materials and devices. These findings have opened the door to plastic electronics and optoelectronics, which are beginning to find commercial application.

Charge transfer complexes

The first highly-conductive organic compounds were the charge transfer complexes. In 1954, researchers at Bell Labs and elsewhere reported organic charge transfer complexes with resistivities as low as 8 ohms-cm. In the early 1970s, salts of tetrathiafulvalene were shown to exhibit almost metallic conductivity, while superconductivity was demonstrated in 1980. Broad research on charge transfer salts continues today.

Conducting polymers

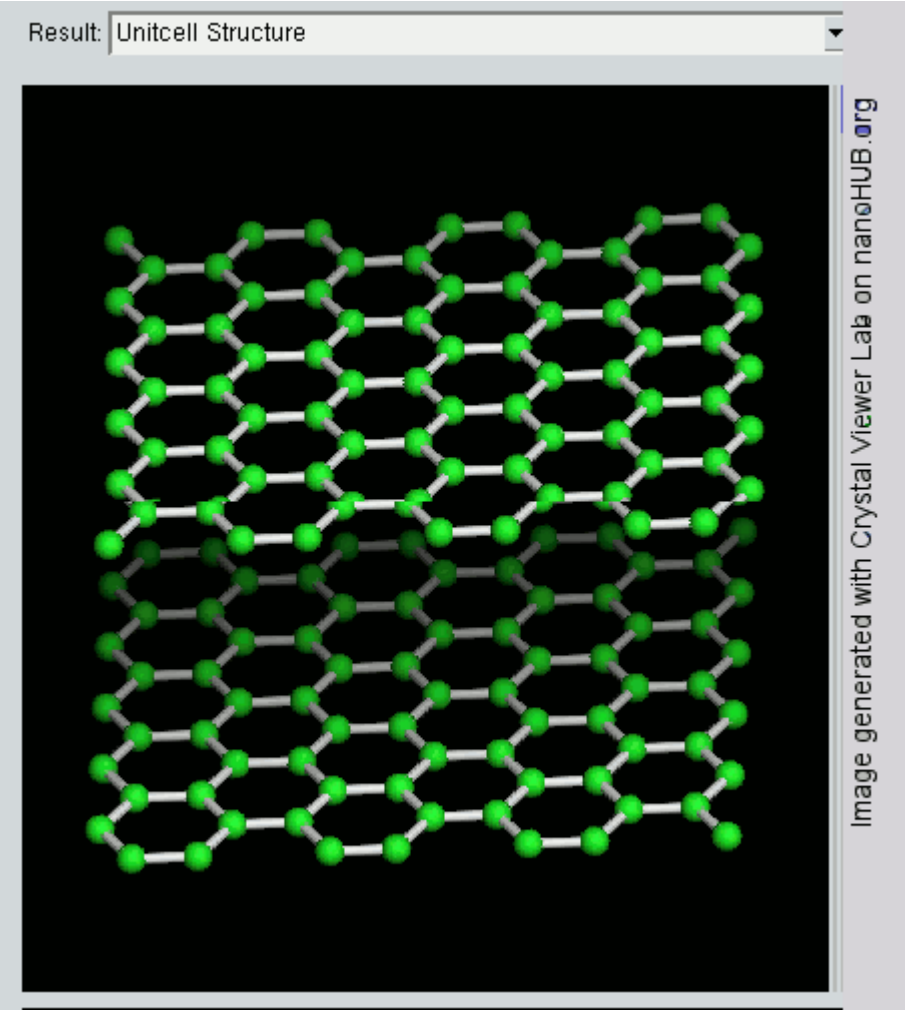
The linear-backbone "polymer blacks" (polyacetylene, polypyrrole, and polyaniline) and their copolymers are the main class of conductive polymers. Historically, these are known as melanins. In 1963 Australians DE Weiss and coworkers reported iodine-doped oxidized polypyrrole blacks with resistivities as low as 1 ohm/cm. Subsequent papers reported resistances as low as 0.03 Ohm/cm. With the notable exception of Charge

transfer complexes (some of which are even superconductors), organic molecules were previously considered insulators or at best weakly conducting semiconductors.

Over a decade later in 1977, Shirakawa, Heeger, and MacDiarmid reported equivalent high conductivity in rather similarly oxidized and iodine-doped polyacetylene. They later received the 2000 Nobel prize in chemistry "for the discovery and development of conductive polymers". The Nobel citation made no reference to Weiss *et al.'s* similar earlier work.

C_{60} and carbon nanotubes

From graphite to C_{60}



Rotating view of a graphite crystal (2 graphene layers)

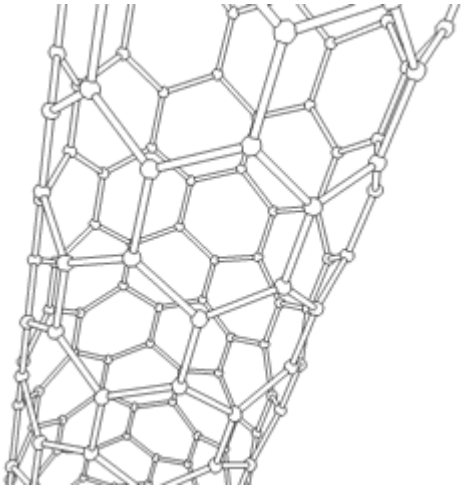
In polymers, classical organic molecules are composed of both carbon and hydrogen (and

sometimes additional compounds such as nitrogen, chlorine or sulphur). They are obtained from petrol and can often be synthesized in large amounts. Most of these molecules are insulating when their length exceeds a few nanometers. However, naturally occurring carbon is conducting. In particular, graphite (recovered from coal or encountered naturally) is conducting. From a theoretical point of view, graphite is a semi-metal, a category in between metals and semi-conductors. It has a layered structure, each sheet being one atom thick. Between each sheet, the interactions are weak enough to allow an easy manual cleavage.

Tailoring the graphite sheet to obtain well defined nanometer-sized objects remains a challenge. However, by the close of the twentieth century, chemists were exploring methods to fabricate extremely small graphitic objects that could be considered single molecules. After studying the interstellar conditions under which carbon is known to form clusters, Richard Smalley's group (Rice University, Texas) set up an experiment in which graphite was vaporized using laser irradiation. Mass spectrometry revealed that clusters containing specific "magic numbers" of atoms were stable, in particular those clusters of 60 atoms. Harry Kroto, an English chemist who assisted in the experiment, suggested a possible geometry for these clusters - atoms covalently bound with the exact symmetry of a soccer ball. Coined buckminsterfullerenes, buckyballs or C_{60}, the clusters retained some properties of graphite, such as conductivity. These objects were rapidly envisioned as possible building blocks for molecular electronics.

Carbon nanotubes

Carbon nanotube



Carbon nanotubes (**CNTs**; also known as **buckytubes**), not to be confused with Carbon Fiber, are allotropes of carbon with a cylindrical nanostructure. Nanotubes have been constructed with length-to-diameter ratio of up to 132,000,000:1, significantly larger than any other material. These cylindrical carbon molecules have novel properties, making

them potentially useful in many applications in nanotechnology, electronics, optics, and other fields of materials science, as well as potential uses in architectural fields. They may also have applications in the construction of body armor. They exhibit extraordinary strength and unique electrical properties, and are efficient thermal conductors.

Nanotubes are members of the fullerene structural family, which also includes the spherical buckyballs. The ends of a nanotube may be capped with a hemisphere of the buckyball structure. Their name is derived from their size, since the diameter of a nanotube is on the order of a few nanometers (approximately 1/50,000th of the width of a human hair), while they can be up to 18 centimeters in length (as of 2010). Nanotubes are categorized as single-walled nanotubes (SWNTs) and multi-walled nanotubes (MWNTs).

Applied quantum chemistry, specifically, orbital hybridization best describes chemical bonding in nanotubes. The chemical bonding of nanotubes is composed entirely of sp^2 bonds, similar to those of graphite. These bonds, which are stronger than the sp^3 bonds found in alkanes, provide nanotubules with their unique strength. Moreover, nanotubes naturally align themselves into "ropes" held together by van der Waals forces.

Most single-walled nanotubes (SWNT) have a diameter of close to 1 nanometer, with a tube length that can be many millions of times longer. The structure of a SWNT can be conceptualized by wrapping a one-atom-thick layer of graphite called graphene into a seamless cylinder. The way the graphene sheet is wrapped is represented by a pair of indices (n,m) called the chiral vector. The integers n and m denote the number of unit vectors along two directions in the honeycomb crystal lattice of graphene. If $m = 0$, the nanotubes are called "zigzag". If $n = m$, the nanotubes are called "armchair". Otherwise, they are called "chiral". The diameter of a nanotube can be calculated from its (n,m) indices as follows

$$d = \frac{a}{\pi}\sqrt{(n^2 + nm + m^2)}.$$

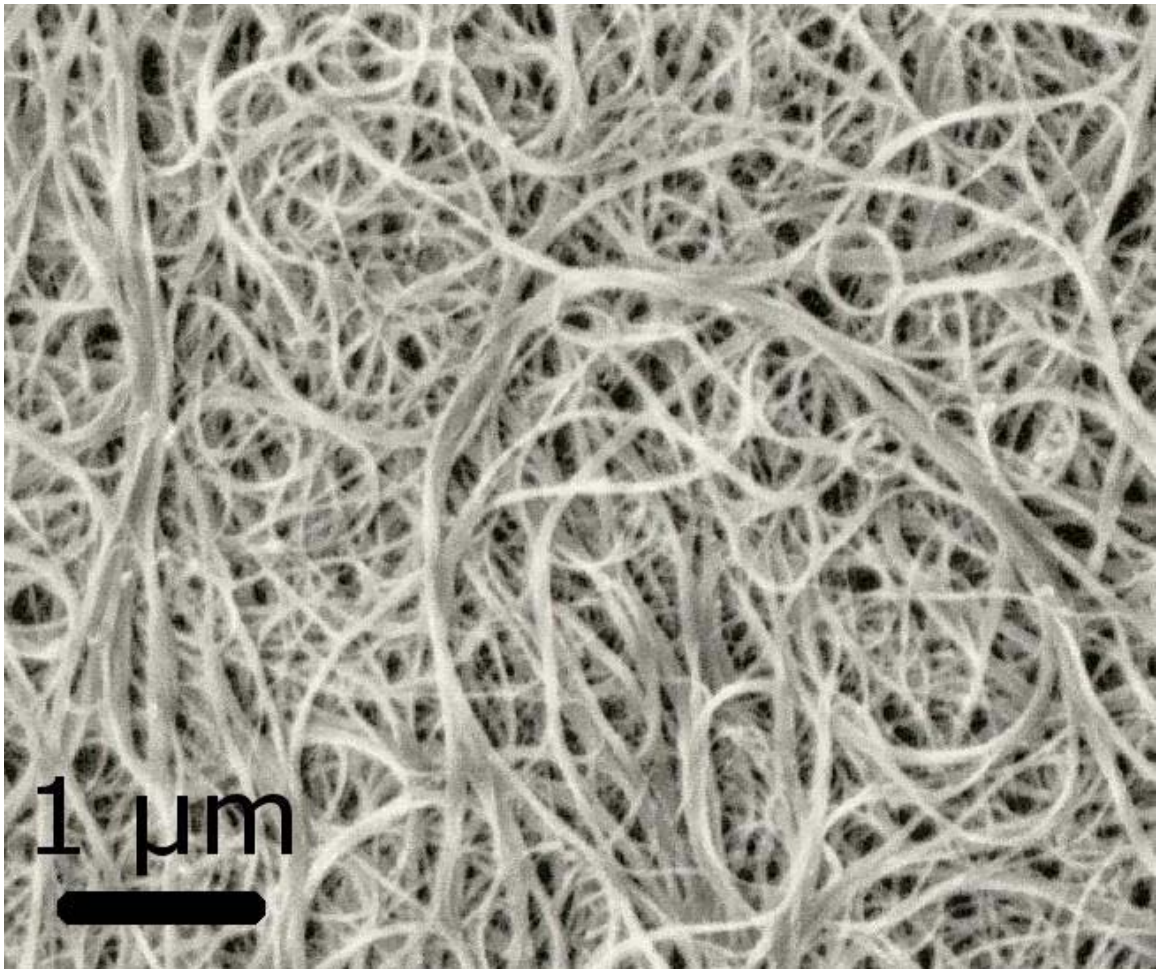
where $a = 0.246$ nm.

Single-walled nanotubes are an important variety of carbon nanotube because they exhibit electric properties that are not shared by the multi-walled carbon nanotube (MWNT) variants. In particular, their band gap can vary from zero to about 2 eV and their electrical conductivity can show metallic or semiconducting behavior, whereas MWNTs are zero-gap metals. Single-walled nanotubes are the most likely candidate for miniaturizing electronics beyond the micro electromechanical scale currently used in electronics. The most basic building block of these systems is the electric wire, and SWNTs can be excellent conductors. One useful application of SWNTs is in the development of the first intramolecular field effect transistors (FET). Production of the first intramolecular logic gate using SWNT FETs has recently become possible as well. To create a logic gate you must have both a p-FET and an n-FET. Because SWNTs are p-FETs when exposed to oxygen and n-FETs otherwise, it is possible to protect half of an SWNT from oxygen exposure, while exposing the other half to oxygen. This results in a

single SWNT that acts as a NOT logic gate with both p and n-type FETs within the same molecule.

Single-walled nanotubes are dropping precipitously in price, from around $1500 per gram as of 2000 to retail prices of around $50 per gram of as-produced 40–60% by weight SWNTs as of March 2010.

Multi-walled



SEM image of carbon nanotubes bundles.

Multi-walled nanotubes (MWNT) consist of multiple rolled layers (concentric tubes) of graphite. There are two models which can be used to describe the structures of multi-walled nanotubes. In the *Russian Doll* model, sheets of graphite are arranged in concentric cylinders, e.g. a (0,8) single-walled nanotube (SWNT) within a larger (0,17) single-walled nanotube. In the *Parchment* model, a single sheet of graphite is rolled in around itself, resembling a scroll of parchment or a rolled newspaper. The interlayer distance in multi-walled nanotubes is close to the distance between graphene layers in graphite, approximately 3.4 Å.

The special place of double-walled carbon nanotubes (DWNT) must be emphasized here because their morphology and properties are similar to SWNT but their resistance to chemicals is significantly improved. This is especially important when functionalization is required (this means grafting of chemical functions at the surface of the nanotubes) to add new properties to the CNT. In the case of SWNT, covalent functionalization will break some C=C double bonds, leaving "holes" in the structure on the nanotube and thus modifying both its mechanical and electrical properties. In the case of DWNT, only the outer wall is modified. DWNT synthesis on the gram-scale was first proposed in 2003 by the CCVD technique, from the selective reduction of oxide solutions in methane and hydrogen.

Torus

A stable nanobud structure

In theory, a nanotorus is a carbon nanotube bent into a torus (doughnut shape). Nanotori are predicted to have many unique properties, such as magnetic moments 1000 times larger than previously expected for certain specific radii. Properties such as magnetic moment, thermal stability, etc. vary widely depending on radius of the torus and radius of the tube.

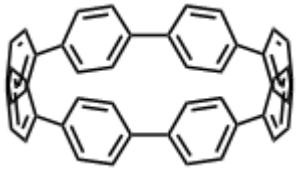
Nanobud

Carbon nanobuds are a newly created material combining two previously discovered allotropes of carbon: carbon nanotubes and fullerenes. In this new material, fullerene-like "buds" are covalently bonded to the outer sidewalls of the underlying carbon nanotube. This hybrid material has useful properties of both fullerenes and carbon nanotubes. In particular, they have been found to be exceptionally good field emitters. In composite materials, the attached fullerene molecules may function as molecular anchors preventing slipping of the nanotubes, thus improving the composite's mechanical properties.

Cup stacked carbon nanotubes

Cup-stacked carbon nanotubes (CSCNTs) differ from other quasi-1D carbon structures, which normally behave as quasi-metallic conductors of electrons. CSCNTs exhibit semiconducting behaviors due to the stacking microstructure of graphene layers.

Extreme carbon nanotubes



Cycloparaphenylene

The observation of the *longest* carbon nanotubes (18.5 cm long) was reported in 2009. These nanotubles were grown on Si substrates using an improved chemical vapor deposition (CVD) method and represent electrically uniform arrays of single-walled carbon nanotubes.

The *shortest* carbon nanotube is the organic compound cycloparaphenylene which was synthesized in early 2009.

The *thinnest* carbon nanotube is armchair (2,2) CNT with a diameter of 3 Å. This nanotube was grown inside a multi-walled carbon nanotube. Assigning of carbon nanotube type was done by combination of high-resolution transmission electron microscopy (HRTEM), Raman spectroscopy and density functional theory (DFT) calculations.

The *thinnest freestanding* single-walled carbon nanotube is about 4.3 Å in diameter. Researchers suggested that it can be either (5,1) or (4,2) SWCNT, but exact type of carbon nanotube remains questionable. (3,3), (4,3) and (5,1) carbon nanotubes (all about 4 Å in diameter) were unambiguously identified using more precise aberration-corrected high-resolution transmission electron microscopy. However, they were found inside of double-walled carbon nanotubes.

Properties

Strength

Carbon nanotubes are the strongest and stiffest materials yet discovered in terms of tensile strength and elastic modulus respectively. This strength results from the covalent sp^2 bonds formed between the individual carbon atoms. In 2000, a multi-walled carbon nanotube was tested to have a tensile strength of 63 gigapascals (GPa). (This, for illustration, translates into the ability to endure tension of a weight equivalent to 6422 kg on a cable with cross-section of 1 mm^2.) Since carbon nanotubes have a low density for a solid of 1.3 to 1.4 $g \cdot cm^{-3}$, its specific strength of up to 48,000 $kN \cdot m \cdot kg^{-1}$ is the best of known materials, compared to high-carbon steel's 154 $kN \cdot m \cdot kg^{-1}$.

Under excessive tensile strain, the tubes will undergo plastic deformation, which means the deformation is permanent. This deformation begins at strains of approximately 5%

and can increase the maximum strain the tubes undergo before fracture by releasing strain energy.

CNTs are not nearly as strong under compression. Because of their hollow structure and high aspect ratio, they tend to undergo buckling when placed under compressive, torsional or bending stress.

Comparison of mechanical properties

Material	Young's modulus (TPa)	Tensile strength (GPa)	Elongation at break (%)
SWNT	~1 (from 1 to 5)	13–53[E]	16
Armchair SWNT	0.94[T]	126.2[T]	23.1
Zigzag SWNT	0.94[T]	94.5[T]	15.6–17.5
Chiral SWNT	0.92		
MWNT	0.27[E]–0.8[E]–0.95[E]	11[E]–63[E]–150[E]	
Stainless steel	0.186[E]–0.214[E]	0.38[E]–1.55[E]	15–50
Kevlar–29&149	0.06[E]–0.18[E]	3.6[E]–3.8[E]	~2

[E]Experimental observation; [T]Theoretical prediction

The above discussion referred to axial properties of the nanotube, whereas simple geometrical considerations suggest that carbon nanotubes should be much softer in the radial direction than along the tube axis. Indeed, TEM observation of radial elasticity suggested that even the van der Waals forces can deform two adjacent nanotubes. Nanoindentation experiments, performed by several groups on multiwalled carbon nanotubes, indicated Young's modulus of the order of several GPa confirming that CNTs are indeed rather soft in the radial direction.

Hardness

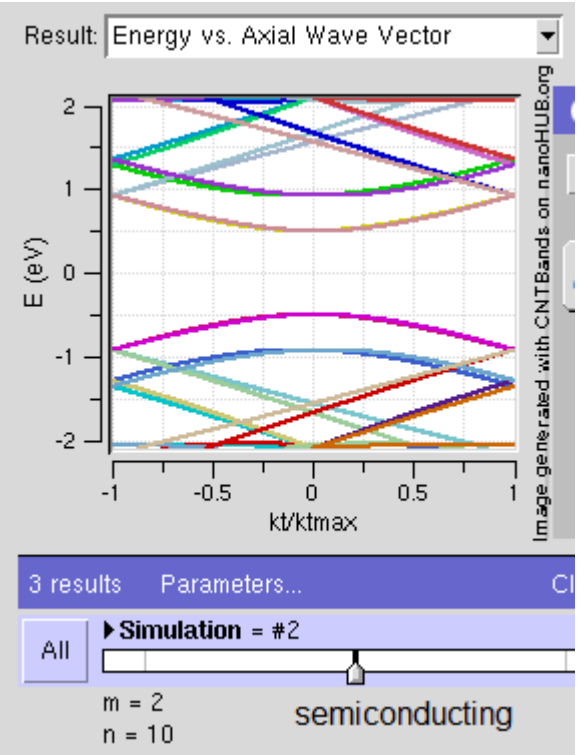
Diamond is considered to be the hardest material. Under conditions of high temperature and high pressure, graphite transforms into diamond. One study succeeded in the synthesis of a super-hard material by compressing SWNTs to above 24 GPa at *room temperature*. The hardness of this material was measured with a nanoindenter as 62–152 GPa. The hardness of reference diamond and boron nitride samples was 150 and 62 GPa, respectively. The bulk modulus of compressed SWNTs was 462–546 GPa, surpassing the value of 420 GPa for diamond.

Kinetic

Multi-walled nanotubes are multiple concentric nanotubes precisely nested within one another. These exhibit a striking telescoping property whereby an inner nanotube core

may slide, almost without friction, within its outer nanotube shell, thus creating an atomically perfect linear or rotational bearing. This is one of the first true examples of molecular nanotechnology, the precise positioning of atoms to create useful machines. Already, this property has been utilized to create the world's smallest rotational motor. Future applications such as a gigahertz mechanical oscillator are also envisaged.

Electrical



Band structures computed using tight binding approximation for (6,0) CNT (zigzag, metallic) (10,2) CNT (semiconducting) and (10,10) CNT (armchair, metallic).

Because of the symmetry and unique electronic structure of graphene, the structure of a nanotube strongly affects its electrical properties. For a given (n,m) nanotube, if $n = m$, the nanotube is metallic; if $n - m$ is a multiple of 3, then the nanotube is semiconducting with a very small band gap, otherwise the nanotube is a moderate semiconductor. Thus all armchair ($n = m$) nanotubes are metallic, and nanotubes (6,4), (9,1), etc. are semiconducting.

However, this rule has exceptions, because curvature effects in small diameter carbon nanotubes can influence strongly electrical properties. Thus, a (5,0) SWCNT that should be semiconducting in fact is metallic according to the calculations. Likewise, *vice versa*--zigzag and chiral SWCNTs with small diameters that should be metallic have finite gap (armchair nanotubes remain metallic). In theory, metallic nanotubes can carry an electric current density of 4×10^9 A/cm^2 which is more than 1,000 times greater than metals such as copper.

Multiwalled carbon nanotubes with interconnected inner shells show superconductivity with a relatively high transition temperature $T_c = 12$ K. In contrast, the T_c value is an order of magnitude lower for ropes of single-walled carbon nanotubes or for MWNTs with usual, non-interconnected shells.

Optical

Thermal

All nanotubes are expected to be very good thermal conductors along the tube, exhibiting a property known as "ballistic conduction", but good insulators laterally to the tube axis. Measurements show that a SWNT has a room-temperature thermal conductivity along its axis of about 3500 $W \cdot m^{-1} \cdot K^{-1}$; compare this to copper, a metal well-known for its good thermal conductivity, which transmits 385 $W \cdot m^{-1} \cdot K^{-1}$. A SWNT has a room-temperature thermal conductivity across its axis (in the radial direction) of about 1.52 $W \cdot m^{-1} \cdot K^{-1}$, which is about as thermally conductive as soil. The temperature stability of carbon nanotubes is estimated to be up to 2800 °C in vacuum and about 750 °C in air.

Defects

As with any material, the existence of a crystallographic defect affects the material properties. Defects can occur in the form of atomic vacancies. High levels of such defects can lower the tensile strength by up to 85%. Another form of carbon nanotube defect is the Stone Wales defect, which creates a pentagon and heptagon pair by rearrangement of the bonds. Because of the very small structure of CNTs, the tensile strength of the tube is dependent on its weakest segment in a similar manner to a chain, where the strength of the weakest link becomes the maximum strength of the chain.

Crystallographic defects also affect the tube's electrical properties. A common result is lowered conductivity through the defective region of the tube. A defect in armchair-type tubes (which can conduct electricity) can cause the surrounding region to become semiconducting, and single monoatomic vacancies induce magnetic properties.

Crystallographic defects strongly affect the tube's thermal properties. Such defects lead to phonon scattering, which in turn increases the relaxation rate of the phonons. This reduces the mean free path and reduces the thermal conductivity of nanotube structures. Phonon transport simulations indicate that substitutional defects such as nitrogen or boron will primarily lead to scattering of high-frequency optical phonons. However, larger-scale defects such as Stone Wales defects cause phonon scattering over a wide range of frequencies, leading to a greater reduction in thermal conductivity.

One-dimensional transport

Because of the nanoscale dimensions, electrons propagate only along the tube's axis and electron transport involves many quantum effects. Because of this, carbon nanotubes are frequently referred to as “one-dimensional”.

Toxicity

Determining the toxicity of carbon nanotubes has been one of the most pressing questions in nanotechnology. Unfortunately, such research has only just begun. Thus, the data are still fragmentary and subject to criticism. Preliminary results highlight the difficulties in evaluating the toxicity of this heterogeneous material. Parameters such as structure, size distribution, surface area, surface chemistry, surface charge, and agglomeration state as well as purity of the samples, have considerable impact on the reactivity of carbon nanotubes. However, available data clearly show that, under some conditions, nanotubes can cross membrane barriers, which suggests that if raw materials reach the organs they can induce harmful effects such as inflammatory and fibrotic reactions.

A study led by Alexandra Porter from the University of Cambridge shows that CNTs can enter human cells and accumulate in the cytoplasm, causing cell death.

Results of rodent studies collectively show that regardless of the process by which CNTs were synthesized and the types and amounts of metals they contained, CNTs were capable of producing inflammation, epithelioid granulomas (microscopic nodules), fibrosis, and biochemical/toxicological changes in the lungs. Comparative toxicity studies in which mice were given equal weights of test materials showed that SWCNTs were more toxic than quartz, which is considered a serious occupational health hazard when chronically inhaled. As a control, ultrafine carbon black was shown to produce minimal lung responses.

The needle-like fiber shape of CNTs, similar to asbestos fibers, raises fears that widespread use of carbon nanotubes may lead to mesothelioma, cancer of the lining of the lungs often caused by exposure to asbestos. A recently-published pilot study supports this prediction. Scientists exposed the mesothelial lining of the body cavity of mice, as a surrogate for the mesothelial lining of the chest cavity, to long multiwalled carbon nanotubes and observed asbestos-like, length-dependent, pathogenic behavior which included inflammation and formation of lesions known as granulomas. Authors of the study conclude:

> "This is of considerable importance, because research and business communities continue to invest heavily in carbon nanotubes for a wide range of products under the assumption that they are no more hazardous than graphite. Our results suggest the need for further research and great caution before introducing such products into the market if long-term harm is to be avoided."

According to co-author Dr. Andrew Maynard:

> "This study is exactly the kind of strategic, highly focused research needed to ensure the safe and responsible development of nanotechnology. It looks at a specific nanoscale material expected to have widespread commercial applications and asks specific questions about a specific health hazard. Even though scientists have been raising concerns about the safety of long, thin carbon nanotubes for

over a decade, none of the research needs in the current U.S. federal nanotechnology environment, health and safety risk research strategy address this question."

Although further research is required, results presented today clearly demonstrate that, under certain conditions, especially those involving chronic exposure, carbon nanotubes can pose a serious risk to human health.

Synthesis

Powder of carbon nanotubes

Techniques have been developed to produce nanotubes in sizeable quantities, including arc discharge, laser ablation, high pressure carbon monoxide (HiPco), and chemical vapor deposition (CVD). Most of these processes take place in vacuum or with process gases. CVD growth of CNTs can occur in vacuum or at atmospheric pressure. Large quantities of nanotubes can be synthesized by these methods; advances in catalysis and continuous growth processes are making CNTs more commercially viable.

Arc discharge

Nanotubes were observed in 1991 in the carbon soot of graphite electrodes during an arc discharge, by using a current of 100 amps, that was intended to produce fullerenes. However the first macroscopic production of carbon nanotubes was made in 1992 by two researchers at NEC's Fundamental Research Laboratory. The method used was the same as in 1991. During this process, the carbon contained in the negative electrode sublimates because of the high discharge temperatures. Because nanotubes were initially discovered using this technique, it has been the most widely-used method of nanotube synthesis.

The yield for this method is up to 30 percent by weight and it produces both single- and multi-walled nanotubes with lengths of up to 50 micrometers with few structural defects.

Laser ablation

In the laser ablation process, a pulsed laser vaporizes a graphite target in a high-temperature reactor while an inert gas is bled into the chamber. Nanotubes develop on the cooler surfaces of the reactor as the vaporized carbon condenses. A water-cooled surface may be included in the system to collect the nanotubes.

This process was developed by Dr. Richard Smalley and co-workers at Rice University, who at the time of the discovery of carbon nanotubes, were blasting metals with a laser to produce various metal molecules. When they heard of the existence of nanotubes they replaced the metals with graphite to create multi-walled carbon nanotubes. Later that year the team used a composite of graphite and metal catalyst particles (the best yield was from a cobalt and nickel mixture) to synthesize single-walled carbon nanotubes.

The laser ablation method yields around 70% and produces primarily single-walled carbon nanotubes with a controllable diameter determined by the reaction temperature. However, it is more expensive than either arc discharge or chemical vapor deposition.

Chemical vapor deposition (CVD)

Nanotubes being grown by plasma enhanced chemical vapor deposition

The catalytic vapor phase deposition of carbon was first reported in 1959, but it was not until 1993 that carbon nanotubes were formed by this process. In 2007, researchers at the University of Cincinnati (UC) developed a process to grow aligned carbon nanotube arrays of 18 mm length on a FirstNano ET3000 carbon nanotube growth system.

During CVD, a substrate is prepared with a layer of metal catalyst particles, most commonly nickel, cobalt, iron, or a combination. The metal nanoparticles can also be produced by other ways, including reduction of oxides or oxides solid solutions. The diameters of the nanotubes that are to be grown are related to the size of the metal

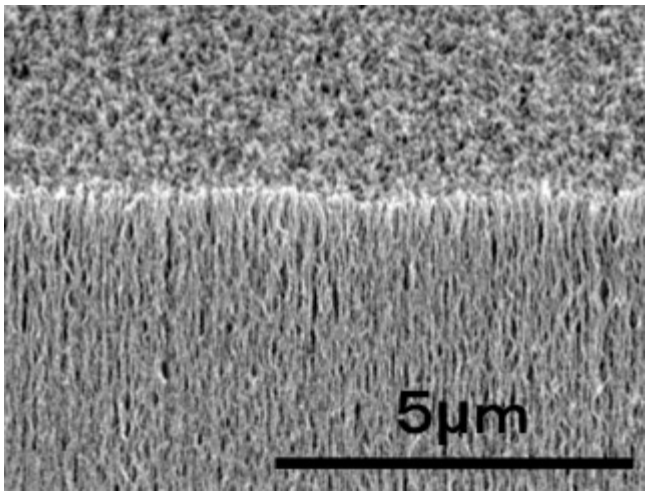
particles. This can be controlled by patterned (or masked) deposition of the metal, annealing, or by plasma etching of a metal layer. The substrate is heated to approximately 700°C. To initiate the growth of nanotubes, two gases are bled into the reactor: a process gas (such as ammonia, nitrogen or hydrogen) and a carbon-containing gas (such as acetylene, ethylene, ethanol or methane). Nanotubes grow at the sites of the metal catalyst; the carbon-containing gas is broken apart at the surface of the catalyst particle, and the carbon is transported to the edges of the particle, where it forms the nanotubes. This mechanism is still being studied. The catalyst particles can stay at the tips of the growing nanotube during the growth process, or remain at the nanotube base, depending on the adhesion between the catalyst particle and the substrate. Thermal catalytic decomposition of hydrocarbon has become an active area of research and can be a promising route for the bulk production of CNTs. Fluidised bed reactor is the most widely used reactor for CNT preparation. Scale-up of the reactor is the major challenge.

CVD is a common method for the commercial production of carbon nanotubes. For this purpose, the metal nanoparticles are mixed with a catalyst support such as MgO or Al_2O_3 to increase the surface area for higher yield of the catalytic reaction of the carbon feedstock with the metal particles. One issue in this synthesis route is the removal of the catalyst support via an acid treatment, which sometimes could destroy the original structure of the carbon nanotubes. However, alternative catalyst supports that are soluble in water have proven effective for nanotube growth.

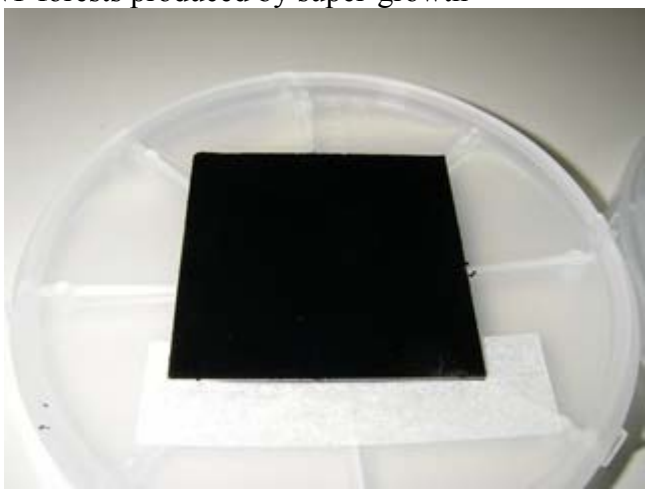
If a plasma is generated by the application of a strong electric field during the growth process (plasma enhanced chemical vapor deposition*), then the nanotube growth will follow the direction of the electric field. By adjusting the geometry of the reactor it is possible to synthesize vertically aligned carbon nanotubes (i.e., perpendicular to the substrate), a morphology that has been of interest to researchers interested in the electron emission from nanotubes. Without the plasma, the resulting nanotubes are often randomly oriented. Under certain reaction conditions, even in the absence of a plasma, closely spaced nanotubes will maintain a vertical growth direction resulting in a dense array of tubes resembling a carpet or forest.

Of the various means for nanotube synthesis, CVD shows the most promise for industrial-scale deposition, because of its price/unit ratio, and because CVD is capable of growing nanotubes directly on a desired substrate, whereas the nanotubes must be collected in the other growth techniques. The growth sites are controllable by careful deposition of the catalyst. In 2007, a team from Meijo University demonstrated a high-efficiency CVD technique for growing carbon nanotubes from camphor. Researchers at Rice University, until recently led by the late Richard Smalley, have concentrated upon finding methods to produce large, pure amounts of particular types of nanotubes. Their approach grows long fibers from many small seeds cut from a single nanotube; all of the resulting fibers were found to be of the same diameter as the original nanotube and are expected to be of the same type as the original nanotube.

Super-growth CVD



SEM photo of SWNT forests produced by super-growth



A small SWNT sample produced by super-growth

Super-growth CVD (water-assisted chemical vapour deposition) process was developed by Kenji Hata, Sumio Iijima and co-workers at AIST, Japan. In this process, the activity and lifetime of the catalyst are enhanced by addition of water into the CVD reactor. Dense millimeter-tall nanotube "forests", aligned normal to the substrate, were produced. The forests growth rate could be expressed, as

$$H(t) = \beta\tau_o(1 - e^{-t/\tau_o}).$$

In this equation, β is the initial growth rate and τ_o is the characteristic catalyst lifetime.

Their specific surface exceeds 1,000 m^2/g (capped) or 2,200 m^2/g (uncapped), surpassing the value of 400–1,000 m^2/g for HiPco samples. The synthesis efficiency is about 100 times higher than for the laser ablation method. The time required to make SWNT forests

of the height of 2.5 mm by this method was 10 minutes in 2004. Those SWNT forests can be easily separated from the catalyst, yielding clean SWNT material (purity >99.98%) without further purification. For comparison, the as-grown HiPco CNTs contain about 5-35% of metal impurities; it is therefore purified through dispersion and centrifugation that damages the nanotubes. The super-growth process avoids this problem. Patterned highly organized single-walled nanotube structures were successfully fabricated using the super-growth technique.

The mass density of super-growth CNTs is about 0.037 g/cm^3. It is much lower than that of conventional CNT powders (~1.34 g/cm^3), probably because the latter contain metals and amorphous carbon.

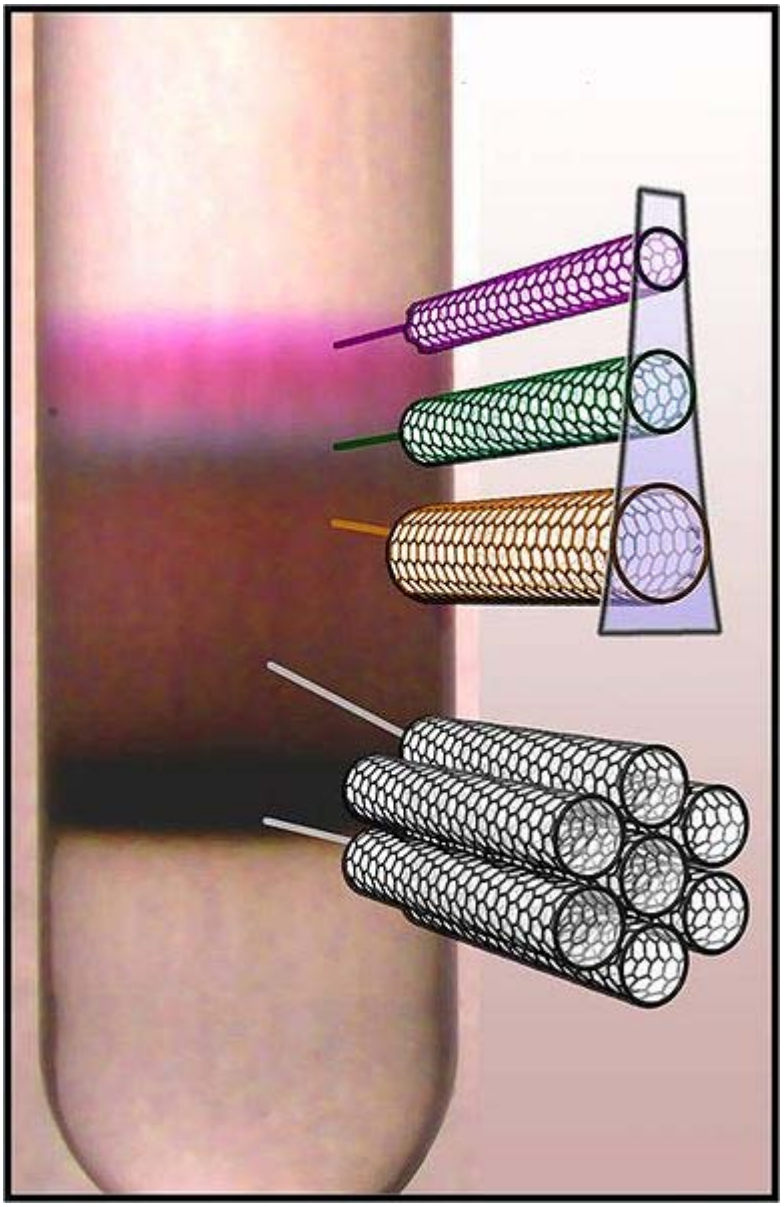
The super-growth method is basically a variation of CVD. Therefore, it is possible to grow material containing SWNT, DWNTs and MWNTs, and to alter their ratios by tuning the growth conditions. Their ratios change by the thinness of the catalyst. Many MWNTs are included so that the diameter of the tube is wide.

The vertically aligned nanotube forests originate from a "zipping effect" when they are immersed in a solvent and dried. The zipping effect is caused by the surface tension of the solvent and the van der Waals forces between the carbon nanotubes. It aligns the nanotubes into a dense material, which can be formed in various shapes, such as sheets and bars, by applying weak compression during the process. Densification increases the Vickers hardness by about 70 times and density is 0.55 g/cm^3. The packed carbon nanotubes are more than 1 mm long and have a carbon purity of 99.9% or higher; they also retain the desirable alignment properties of the nanotubes forest.

Natural, incidental, and controlled flame environments

Fullerenes and carbon nanotubes are not necessarily products of high-tech laboratories; they are commonly formed in such mundane places as ordinary flames, produced by burning methane, ethylene, and benzene, and they have been found in soot from both indoor and outdoor air. However, these naturally occurring varieties can be highly irregular in size and quality because the environment in which they are produced is often highly uncontrolled. Thus, although they can be used in some applications, they can lack in the high degree of uniformity necessary to satisfy the many needs of both research and industry. Recent efforts have focused on producing more uniform carbon nanotubes in controlled flame environments. Such methods have promise for large-scale, low-cost nanotube synthesis, though they must compete with rapidly developing large scale CVD production.

Application related issues



Centrifuge tube with a solution of carbon nanotubes, which were sorted by diameter using density-gradient ultracentrifugation.

Many electronic applications of carbon nanotubes crucially rely on techniques of selectively producing either semiconducting or metallic CNTs, preferably of a certain chirality. Several methods of separating semiconducting and metallic CNTs are known, but most of them are not yet suitable for large-scale technological processes. The most efficient method relies on density-gradient ultracentrifugation which separates surfactant-wrapped nanotubes by the minute difference in their density. This density difference often translates into difference in the nanotube diameter and (semi)conducting properties. Another method of separation uses a sequence of freezing, thawing, and compression of

SWNTs embedded in agarose gel. This process results in a solution containing 70% metallic SWNTs and leaves a gel containing 95% semiconducting SWNTs. The diluted solutions separated by this method show various colors. Moreover, SWNTs can be separated by the column chromatography method. Yield is 95% in semiconductor type SWNT and 90% in metallic type SWNT.

In addition to separation of semiconducting and metallic SWNTs, it is possible to sort SWNTs by length, diameter, and chirality. The highest resolution length sorting, with length variation of <10%, has thus far been achieved by size exclusion chromatography (SEC) of DNA-dispersed carbon nanotubes (DNA-SWNT). SWNT diameter separation has been achieved by density-gradient ultracentrifugation (DGU) using surfactant-dispersed SWNTs and by ion-exchange chromatography (IEC) for DNA-SWNT. Purification of individual chiralities has also been demonstrated with IEC of DNA-SWNT: specific short DNA oligomers can be used to isolate individual SWNT chiralities. Thus far, 12 chiralities have been isolated at purities ranging from 70% for (8,3) and (9,5) SWNTs to 90% for (6,5), (7,5) and (10,5)SWNTs. There have been successful efforts to integrate these purified nanotubes into devices, e. g. FETs.

An alternative to separation is development of a selective growth of semiconducting or metallic CNTs. Recently, a new CVD recipe was announced which involves a combination of ethanol and methanol gases and quartz substrates resulting in horizontally aligned arrays of 95–98% semiconducting nanotubes.

Nanotubes are usually grown on nanoparticles of magnetic metal (Fe, Co), which facilitates production of electronic (spintronic) devices. In particular control of current through a field-effect transistor by magnetic field has been demonstrated in such a single-tube nanostructure.

Current applications

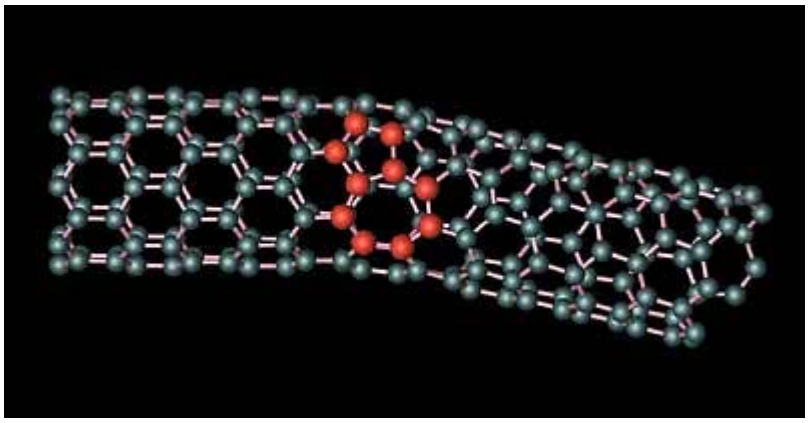
Current use and application of nanotubes has mostly been limited to the use of bulk nanotubes, which is a mass of rather unorganized fragments of nanotubes. Bulk nanotube materials may never achieve a tensile strength similar to that of individual tubes, but such composites may nevertheless yield strengths sufficient for many applications. Bulk carbon nanotubes have already been used as composite fibers in polymers to improve the mechanical, thermal and electrical properties of the bulk product.

Easton-Bell Sports, Inc. have been in partnership with Zyvex Performance Materials, using CNT technology in a number of their bicycle components—including flat and riser handlebars, cranks, forks, seatposts, stems and aero bars.

Zyvex Performance Materials has also built a 54' maritime vessel, the Piranha Unmanned Surface Vessel, as a technology demonstrator for what is possible using CNT technology. CNTs help improve the structural performance of the vessel, resulting in a lightweight 8,000lb boat.

Amroy Europe Oy manufactures Hybtonite carbon nanoepoxy resins where carbon nanotubes have been chemically bond to epoxy, resulting composite material that is 20% to 30% stronger than other composite materials. It has been used for wind turbines, marine paints and variety of sports gear such as skis, ice hockey sticks, baseball bats, hunting arrows and surfboards.

Potential applications



The joining of two carbon nanotubes with different electrical properties to form a diode has been proposed . L Chico et al. Phys Rev Lett 76, 971 (1996)

The strength and flexibility of carbon nanotubes makes them of potential use in controlling other nanoscale structures, which suggests they will have an important role in nanotechnology engineering. The highest tensile strength of an individual multi-walled carbon nanotube has been tested to be is 63 GPa. Carbon nanotubes were found in Damascus steel from the 17th century, possibly helping to account for the legendary strength of the swords made of it.

Structural

Because of the carbon nanotube's superior mechanical properties, many structures have been proposed ranging from everyday items like clothes and sports gear to combat jackets and space elevators. However, the space elevator will require further efforts in refining carbon nanotube technology, as the practical tensile strength of carbon nanotubes can still be greatly improved.

For perspective, outstanding breakthroughs have already been made. Pioneering work led by Ray H. Baughman at the NanoTech Institute has shown that single and multi-walled nanotubes can produce materials with toughness unmatched in the man-made and natural worlds.

Because of the high mechanical strength of carbon nanotubes, research is being made into weaving them into clothes to create stab-proof and bulletproof clothing. The nanotubes

would effectively stop the bullet from penetrating the body, although the bullet's kinetic energy would likely cause broken bones and internal bleeding.

In electrical circuits

Nanotube-based transistors, also known as carbon nanotube field-effect transistors (CNFETs), have been made that operate at room temperature and that are capable of digital switching using a single electron. However, one major obstacle to realization of nanotubes has been the lack of technology for mass production. In 2001 IBM researchers demonstrated how metallic nanotubes can be destroyed, leaving semiconducting ones behind for use as transistors. Their process is called "constructive destruction" which includes the automatic destruction of defective nanotubes on the wafer. This process, however, only gives control over the electrical properties on a statistical scale.

The potential of carbon nanotubes was demonstrated in 2003 when room-temperature ballistic transistors with ohmic metal contacts and high-k gate dielectric were reported, showing 20–30x higher ON current than state-of-the-art Si MOSFETs. This presented an important advance in the field as CNT was shown to potentially outperform Si. At the time, a major challenge was ohmic metal contact formation. In this regard, palladium, which is a high work function metal was shown to exhibit Schottky barrier-free contacts to semiconducting nanotubes with diameters >1.7 nm.

The first nanotube integrated memory circuit was made in 2004. One of the main challenges has been regulating the conductivity of nanotubes. Depending on subtle surface features a nanotube may act as a plain conductor or as a semiconductor. A fully automated method has however been developed to remove non-semiconductor tubes.

Another way to make carbon nanotube transistors has been to use random networks of them. By doing so one averages all of their electrical differences and one can produce devices in large scale at the wafer level. This approach was first patented by Nanomix Inc.(date of original application June 2002). It was first published in the academic literature by the United States Naval Research Laboratory in 2003 through independent research work. This approach also enabled Nanomix to make the first transistor on a flexible and transparent substrate.

Large structures of carbon nanotubes can be used for thermal management of electronic circuits. An approximately 1 mm–thick carbon nanotube layer was used as a special material to fabricate coolers, this materials has very low density, ~20 times lower weight than a similar copper structure, while the cooling properties are similar for the two materials.

Overall, incorporating carbon nanotubes as transistors into logic-gate circuits with densities comparable to modern CMOS technology has not yet been demonstrated.

As paper batteries

A paper battery is a battery engineered to use a paper-thin sheet of cellulose (which is the major constituent of regular paper, among other things) infused with aligned carbon nanotubes. The nanotubes act as electrodes; allowing the storage devices to conduct electricity. The battery, which functions as both a lithium-ion battery and a supercapacitor, can provide a long, steady power output comparable to a conventional battery, as well as a supercapacitor's quick burst of high energy—and while a conventional battery contains a number of separate components, the paper battery integrates all of the battery components in a single structure, making it more energy efficient.

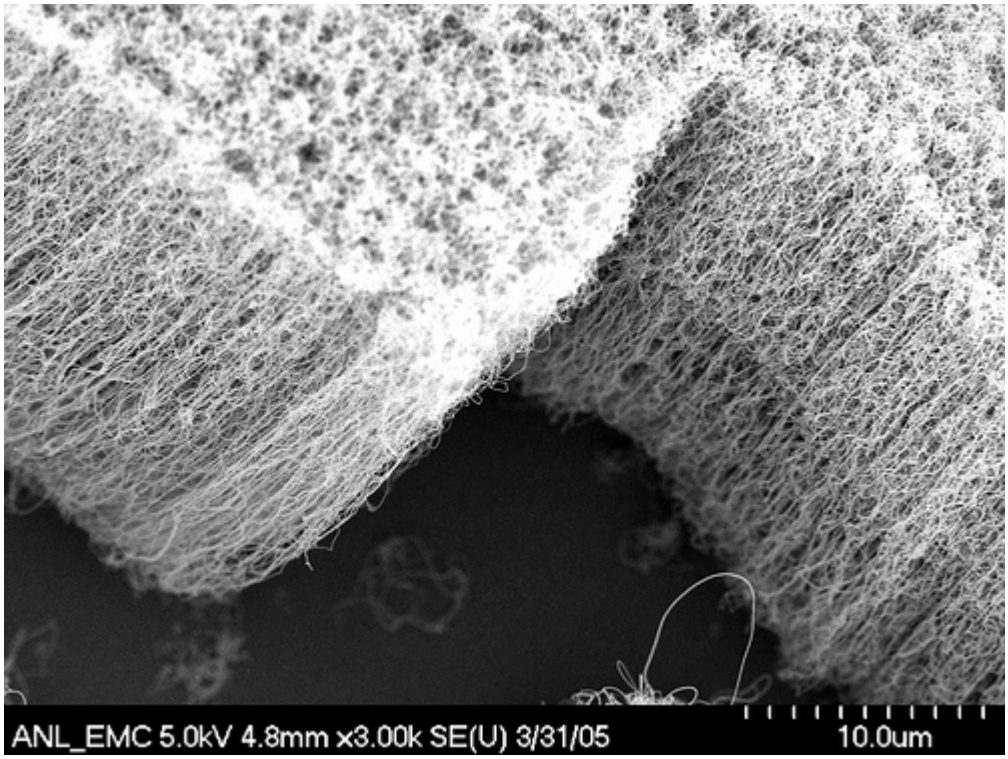
Solar cells

Solar cells developed at the New Jersey Institute of Technology use a carbon nanotube complex, formed by a mixture of carbon nanotubes and carbon buckyballs (known as fullerenes) to form snake-like structures. Buckyballs trap electrons, although they can't make electrons flow. Add sunlight to excite the polymers, and the buckyballs will grab the electrons. Nanotubes, behaving like copper wires, will then be able to make the electrons or current flow.

Ultracapacitors

MIT Laboratory for Electromagnetic and Electronic Systems uses nanotubes to improve ultracapacitors. The activated charcoal used in conventional ultracapacitors has many small hollow spaces of various size, which create together a large surface to store electric charge. But as charge is quantized into elementary charges, i.e. electrons, and each such elementary charge needs a minimum space, a significant fraction of the electrode surface is not available for storage because the hollow spaces are not compatible with the charge's requirements. With a nanotube electrode the spaces may be tailored to size—few too large or too small—and consequently the capacity should be increased considerably.

Other applications



Aligned nanotubes are preferred for many applications.

Carbon nanotubes have been implemented in nanoelectromechanical systems, including mechanical memory elements (NRAM being developed by Nantero Inc.) and nanoscale electric motors.

In May 2005, Nanomix Inc placed on the market a hydrogen sensor which integrated carbon nanotubes on a silicon platform. Since then Nanomix has been patenting many such sensor applications such as in the field of carbon dioxide, nitrous oxide, glucose, DNA detection, etc.

Research at University of California, Riverside has shown that carbon nanotubes are suitable scaffold materials for osteoblast proliferation and bone formation.

Eikos Inc of Franklin, Massachusetts and Unidym Inc. of Silicon Valley, California are developing transparent, electrically conductive films of carbon nanotubes to replace indium tin oxide (ITO). Carbon nanotube films are substantially more mechanically robust than ITO films, making them ideal for high-reliability touchscreens and flexible displays. Printable water-based inks of carbon nanotubes are desired to enable the production of these films to replace ITO. Nanotube films show promise for use in displays for computers, cell phones, PDAs, and ATMs.

A nanoradio, a radio receiver consisting of a single nanotube, was demonstrated in 2007. In 2008 it was shown that a sheet of nanotubes can operate as a loudspeaker if an alternating current is applied. The sound is not produced through vibration but thermoacoustically.

A flywheel made of carbon nanotubes could be spun at extremely high velocity on a floating magnetic axis in a vacuum, and potentially store energy at a density approaching that of conventional fossil fuels. Since energy can be added to and removed from flywheels very efficiently in the form of electricity, this might offer a way of storing electricity, making the electrical grid more efficient and variable power suppliers (like wind turbines) more useful in meeting energy needs. The practicality of this depends heavily upon the cost of making massive, unbroken nanotube structures, and their failure rate under stress.

Ultra-short SWNTs (US-tubes) have been used as nanoscaled capsules for delivering MRI contrast agents in vivo.

Nitrogen-doped carbon nanotubes may replace platinum catalysts used to reduce oxygen in fuel cells. A forest of vertically-aligned nanotubes can reduce oxygen in alkaline solution more effectively than platinum, which has been used in such applications since the 1960s. The nanotubes have the added benefit of not being subject to carbon monoxide poisoning.

Discovery

A 2006 editorial written by Marc Monthioux and Vladimir Kuznetsov in the journal *Carbon* described the interesting and often misstated origin of the carbon nanotube. A large percentage of academic and popular literature attributes the discovery of hollow, nanometer-size tubes composed of graphitic carbon to Sumio Iijima of NEC in 1991.

In 1952 L. V. Radushkevich and V. M. Lukyanovich published clear images of 50 nanometer diameter tubes made of carbon in the Soviet *Journal of Physical Chemistry*. This discovery was largely unnoticed, as the article was published in the Russian language, and Western scientists' access to Soviet press was limited during the Cold War. It is likely that carbon nanotubes were produced before this date, but the invention of the transmission electron microscope (TEM) allowed direct visualization of these structures.

Carbon nanotubes have been produced and observed under a variety of conditions prior to 1991. A paper by Oberlin, Endo, and Koyama published in 1976 clearly showed hollow carbon fibers with nanometer-scale diameters using a vapor-growth technique. Additionally, the authors show a TEM image of a nanotube consisting of a single wall of graphene. Later, Endo has referred to this image as a single-walled nanotube.

In 1979 John Abrahamson presented evidence of carbon nanotubes at the 14th Biennial Conference of Carbon at Pennsylvania State University. The conference paper described carbon nanotubes as carbon fibers which were produced on carbon anodes during arc

discharge. A characterization of these fibers was given as well as hypotheses for their growth in a nitrogen atmosphere at low pressures.

In 1981 a group of Soviet scientists published the results of chemical and structural characterization of carbon nanoparticles produced by a thermocatalytical disproportionation of carbon monoxide. Using TEM images and XRD patterns, the authors suggested that their “carbon multi-layer tubular crystals” were formed by rolling graphene layers into cylinders. They speculated that by rolling graphene layers into a cylinder, many different arrangements of graphene hexagonal nets are possible. They suggested two possibilities of such arrangements: circular arrangement (armchair nanotube) and a spiral, helical arrangement (chiral tube).

In 1987, Howard G. Tennett of Hyperion Catalysis was issued a U.S. patent for the production of "cylindrical discrete carbon fibrils" with a "constant diameter between about 3.5 and about 70 nanometers..., length 10^2 times the diameter, and an outer region of multiple essentially continuous layers of ordered carbon atoms and a distinct inner core...."

Iijima's discovery of multi-walled carbon nanotubes in the insoluble material of arc-burned graphite rods in 1991 and Mintmire, Dunlap, and White's independent prediction that if single-walled carbon nanotubes could be made, then they would exhibit remarkable conducting properties helped create the initial buzz that is now associated with carbon nanotubes. Nanotube research accelerated greatly following the independent discoveries by Bethune at IBM and Iijima at NEC of *single-walled* carbon nanotubes and methods to specifically produce them by adding transition-metal catalysts to the carbon in an arc discharge. The arc discharge technique was well-known to produce the famed Buckminster fullerene on a preparative scale, and these results appeared to extend the run of accidental discoveries relating to fullerenes. The original observation of fullerenes in mass spectrometry was not anticipated, and the first mass-production technique by Krätschmer and Huffman was used for several years before realizing that it produced fullerenes.

The discovery of nanotubes remains a contentious issue. Many believe that Iijima's report in 1991 is of particular importance because it brought carbon nanotubes into the awareness of the scientific community as a whole.

Theory of molecular electronics

Molecular electronics operates in the quantum realm of distances less than 100 nanometers. The theory of single molecule devices is particularly interesting since the system under consideration is an open quantum system in nonequilibrium (driven by voltage).

In the low bias voltage regime, the nonequilibrium nature of the molecular junction can be ignored, and the current-voltage characteristics of the device can be calculated using the equilibrium electronic structure of the system. However, in stronger bias regimes a

more sophisticated treatment is required, as there is no longer a variational principle. In the elastic tunneling case (where the passing electron does not exchange energy with the system), the formalism of Rolf Landauer can be used to calculate the transmission through the system as a function of bias voltage, and hence the current.

In inelastic tunneling, an elegant formalism based on the non-equilibrium Green's functions of Leo Kadanoff and Gordon Baym, and independently by Leonid Keldysh was put forth by Ned Wingreen and Yigal Meir. This Meir-Wingreen formulation has been used to great success in the molecular electronics community to examine the more difficult and interesting cases where the transient electron exchanges energy with the molecular system (for example through electron-phonon coupling or electronic excitations).

Recent progress

Recent progress in nanotechnology and nanoscience has facilitated both experimental and theoretical study of molecular electronics. In particular, the development of the scanning tunneling microscope (STM) and later the atomic force microscope (AFM) have facilitated manipulation of single-molecule electronics. In addition, theoretical advances in molecular electronics have facilitated further understanding of non-adibatic charge transfer events at electrode-electrolyte interfaces.

The first measurement of the conductance of a single molecule was realised in 1994 by C. Joachim and J. K. Gimzewski and published in 1995. This was the conclusion of 10 years of research started at IBM TJ Watson, using the scanning tunnelling microscope tip apex to switch a single molecule as already explored by A. Aviram, C. Joachim and M. Pomerantz at the end of the 80's. The trick was to use an UHV Scanning Tunneling microscope to allow the tip apex to gently touch the top of a single C_{60} molecule adsorbed on a Au(110) surface. A resistance of 55 MOhms was recorded together with a low voltage linear I-V. The contact was certified by recording the I-z current distance characteristic, which allows the measurement of the deformation of the C_{60} cage under contact. This first experiment was followed by the reported result using a mechanical break junction approach to connect two gold electrodes to a sulfur-terminated molecular wire by Mark Reed and James Tour.

A single-molecule amplifier was implemented by C. Joachim and J.K. Gimzewski in IBM Zurich. This experiment involving a single C_{60} molecule demonstrated that a single C_{60} molecule can provide gain in a circuit just by playing with through C_{60} intramolecular quantum interference effects.

A collaboration of researchers at HP and UCLA, led by James Heath, Fraser Stoddart, R. Stanley Williams, and Philip Kuekes, has developed molecular electronics based on rotaxanes and catenanes.

Work is also being done on the use of single-wall carbon nanotubes as field-effect transistors. Most of this work is being done by IBM.

Until recently entirely theoretical, the Aviram-Ratner model for a unimolecular rectifier has been unambiguously-confirmed in experiments by a group led by Geoffrey J. Ashwell at Bangor University, UK. Many rectifying molecules have so far been identified, and the number and efficiency of these systems is expanding rapidly.

Supramolecular electronics is a new field that tackles electronics at a supramolecular level.

An important issue in molecular electronics is the determination of the resistance of a single molecule (both theoretical and experimental). For example, Bumm, et al. used STM to analyze a single molecular switch in a self-assembled monolayer to determine how conductive such a molecule can be. Another problem faced by this field is the difficulty of performing direct characterization since imaging at the molecular scale is often difficult in many experimental devices.

Chapter- 2

Single Molecular Electronics

Single molecule electronics is a branch of molecular electronics that uses single molecules as electronic components. Because single molecules constitute the smallest stable structures imaginable this miniaturization is the ultimate goal for shrinking electrical circuits.

Concepts

In **single molecule electronics**, the bulk material is replaced by single molecules. That is, instead of creating structures by removing or applying material after a pattern scaffold, the atoms are put together in a chemistry lab. This way billions of billions of copies are made simultaneously (typically more than 10^{20} molecules are made at once) while the composition of molecules are controlled down to the last atom. The molecules utilized have properties that resemble traditional electronic components such as a wire, transistor or rectifier.

The miniaturization down to single molecules brings the scale down to a regime where quantum effects are important. As opposed to the case in conventional electronic components, where electrons can be filled in or drawn out more or less like a continuous flow of charge, the transfer of a single electron alters the system significantly. This means that when an electron has been transferred from the source electrode to the molecule, the molecule gets charged up and makes it much harder for the next one to transfer. The significant amount of energy due to charging has to be taken into account when making calculations about the electronic properties of the setup and is highly sensitive to distances to conducting surfaces nearby.

Because of the small size of the molecules, quantum mechanics put severe restraints on the states (or orbitals) the electrons can be in on the molecule. These states determine the energy and spatial distribution that an electron can have and hence the electronic properties of the setup. Unfortunately, even though the molecules seem small and simple

when drawn schematically, the possible electronic states can only be deduced approximately and this limits the predictability of the molecular electronic properties.

Further, connecting single molecules reliably to a larger scale circuit has proven to be a great challenge and constitute a significant hindrance to commercialization.

History

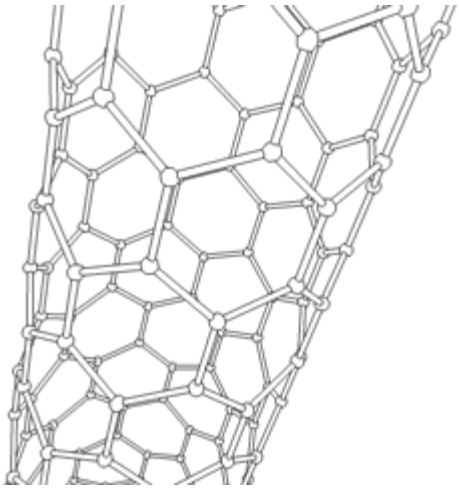
Conventionally, the electronics are made up of bulk material. Ever since its invention in 1958 the performance and complexity of integrated circuits has been growing exponentially (a trend also known as Moore's law) and has forced the feature sizes of the embedded components to shrink accordingly. As the structures become smaller the sensitivity for deviations increases and in a few generations, when the minimum feature sizes reaches 13 nm, the composition of the devices will have to be controlled to a precision of a few atoms in order for the devices to work. With the bulk approach having inherent limitations in addition to becoming increasingly demanding and expensive, the idea was born that the components could instead be built up atom for atom in a chemistry lab (bottom up) as opposed to carving them out of bulk material (top down). This idea is the reasoning behind molecular electronics with the ultimate miniaturization being components contained in single molecules.

The concept of molecular electronics was first published in 1974 when Aviram and Ratner suggested an organic molecule that could work as a rectifier . Having both huge commercial and fundamental interest much effort was put into proving its feasibility and 16 years later in 1990 the first demonstration of an intrinsic molecular rectifier was realized by Ashwell and coworkers for a thin film of molecules. It wasn't until 1997 that the first measurements on the conductance of a single molecule were published and then found to be fraudulent .

Examples

Common for molecules utilized in molecular electronics is that the structures contain a lot of alternating double and single bonds. The reason for this is that such a pattern delocalizes the molecular orbitals making it possible for electrons to move freely over the conjugated area.

Wires



This image of a rotating carbon nanotube shows its 3D structure.

The sole purpose of molecular wires is to electrically connect different parts of a molecular electrical circuit. As the assembly of these and their connection to a macroscopic circuit is still to be mastered, the focus of research in single molecule electronics is primarily on the functionalized molecules: molecular wires are characterized by containing no functional groups and are hence composed of plain repetitions of a conjugated building block. Among these are the carbon nanotubes that are quite large compared to the other suggestions but have shown very promising electrical properties.

The main problem with the molecular wires is to obtain good electrical contact with the electrodes so that the electrons can move freely in and out of the wire.

Transistors

Single molecule transistors are fundamentally different than the ones known from bulk electronics. The gate in a conventional (field-emission) transistor determines the conductance between the source and drain electrode by controlling the density of charge carriers between them, whereas the gate in a single molecule transistor controls the feasibility of a single electron to jump on and off the molecule by modifying the energy of the molecular orbitals. One of the effects of this difference is that the single molecule transistor is almost binary: it is either ON or OFF. This opposes its bulk counterparts which have quadratic responses to gate voltage.

It is the quantization of charge into electrons that is responsible for the markedly different behavior compared to bulk electronics. Because of the size of a single molecule, the charging due to a single electron is significant and provides a mean to turn the transistor ON or OFF. For this to work, the electronic orbitals on the transistor molecule cannot be too well integrated with the orbitals on the electrodes. If they are, an electron cannot be

said to be located on the molecule or the electrodes and the molecule will function as a wire.

A popular group of molecules, that can work as the semiconducting channel material in a molecular transistor, is the oligopolyphenylenevinylenes (OPVs) that works by the Coulomb blockade mechanism when placed between the source and drain electrode in an appropriate way . Fullerenes work by the same mechanism and have also been commonly utilized.

Semiconducting carbon nanotubes have also been demonstrated to work as channel material but although molecular, these molecules are sufficiently large to behave almost as bulk semiconductors.

The size of the molecules and the low temperature the measurements are being conducted at makes the quantum mechanical states well defined. It is therefore being researched if the quantum mechanical properties can be used for more advanced purposes than simple transistors (e.g. spintronics).

Physicists at the University of Arizona, in collaboration with chemists from the University of Madrid, have designed a single molecule transistor using a ring-shaped molecule similar to benzene. Physicists at Canada's National Institute for Nanotechnology have designed a single-molecule transistor using styrene. Both groups expect (their designs have yet to be experimentally verified) their respective devices to function at room temperature, and to be controlled by a single electron.

Rectifiers (diodes)

Molecular rectifiers are mimics of their bulk counterparts and have an asymmetric construction so that the molecule can accept electrons in one end but not the other. The molecules have an electron donor (D) in one end and an electron acceptor (A) in the other. This way, the unstable state $D^+ - A^-$ will be more readily made than $D^- - A^+$. The result is that an electric current can be drawn through the molecule if the electrons are added through the acceptor end, but not so easily if the reverse is attempted. An example of a molecular rectifier was made by Geoffrey J. Ashwell's Ph.D. students.

Techniques

One of the biggest problems with measuring on single molecules is to establish reproducible electrical contact with only one molecule and doing so without shortcutting the electrodes. Because the current photolithographic technology is unable to produce electrode gaps small enough to contact both ends of the molecules tested (in the order of nanometers) alternative strategies is put into use.

Molecular gaps

One way to produce electrodes with a molecular sized gap between them is break junctions, in which a thin electrode is stretched until it breaks. Another is electromigration. Here a current is lead through a thin wire until it melts and the atoms migrate to produce the gap. Further, the reach of conventional photolithography can be enhanced by chemically etching or depositing metal on the electrodes.

Probably the easiest way to conduct measurements on several molecules is to use the tip of a scanning tunneling microscope (STM) to contact molecules adhered at the other end to a metal substrate. The connection obtained this way is, however, not stable enough to conduct the advanced measurements possible with the above techniques.

Anchoring

A popular way to anchor molecules to the electrodes is to make use of sulfurs' high affinity to gold. In these setups, the molecules are synthesized so that sulfur atoms are placed strategically to function as crocodile clips connecting the molecules to the gold electrodes. Though useful, the anchoring is non-specific and thus anchors the molecules randomly to all gold surfaces. Further, the contact resistance is highly dependent on the precise atomic geometry around the site of anchoring and thereby inherently compromises the reproducibility of the connection.

To circumvent the latter issue, experiments has shown that fullerenes could be a good candidate for use instead of sulfur because of the large conjugated π-system that can electrically contact many more atoms at once than a single atom of sulfur.

Problems

Artifacts

When trying to measure electronic characteristics of molecules, artificial phenomena can occur that can be hard to distinguish from truly molecular behavior . Before they were discovered these artifacts have mistakenly been published as being features pertaining to the molecules in question.

Applying a voltage drop in the order of volts across a nanometer sized junction results in a very strong electrical field. The field can cause metal atoms to migrate and eventually close the gap by a thin filament which can be broken again when carrying a current. The two levels of conductance imitate molecular switching between a conductive and an isolating state of a molecule.

Another encountered artifact is when the electrodes undergo chemical reactions due to the high field strength in the gap. When the bias is reversed the reaction will cause hysteresis in the measurements that can be interpreted as being of molecular origin.

A metallic grain between the electrodes can act as a single electron transistor by the mechanism described above thus resembling the characteristics of a molecular transistor. This artifact is especially common with nanogaps produced by the electromigration technique.

Commercialization

One of the biggest hindrances for single molecule electronics to be commercially exploited is the lack of techniques to connect a molecular sized circuit to bulk electrodes in a way that gives reproducible results. At the current state, the difficulty of connecting single molecules vastly outweighs any possible performance increase that could be gained from such shrinkage. The picture becomes even worse if the molecules are to have a certain spatial orientation and/or have multiple poles to connect.

Also problematic is the fact that some measurements on single molecules are carried out in cryogenic temperatures (close to absolute zero) which is very energy consuming. This is done in order to reduce the signal noise to a degree where the faint currents of single molecules can be measured.

Perspectives

Single molecule electronics is an emerging field, and entire electronic circuits consisting exclusively of molecular sized compounds are still very far from being realized. However, the continuous demand for more computing power together with the inherent limitations of the present day lithographic methods make the transition seem unavoidable.

Currently, the focus is on discovering molecules with interesting properties and on finding ways to obtaining reliable and reproducible contacts between the molecular components and the bulk material of the electrodes.

Chapter- 3

Charge-Transfer Complex

A **charge-transfer complex** (**CT complex**) or **electron-donor-acceptor complex** is an association of two or more molecules, or of different parts of one very large molecule, in which a fraction of electronic charge is transferred between the molecular entities. The resulting electrostatic attraction provides a stabilizing force for the molecular complex. The source molecule from which the charge is transferred is called the electron donor and the receiving species is called the electron acceptor.

The nature of the attraction in a charge-transfer complex is not a stable chemical bond, and is much weaker than covalent forces. The attraction is created by an electronic transition into an excited electronic state, and is best characterized as a weak electron resonance. The excitation energy of this resonance occurs very frequently in the visible region of the electro-magnetic spectrum, which produces the usually intense color characteristic for these complexes. These optical absorption bands are often referred to as *charge-transfer bands* (CT bands). Optical spectroscopy is a powerful technique to characterize charge-transfer bands.

Charge-transfer complexes exist in many types of molecules, inorganic as well as organic, and in all phases of matter, i.e. in solids, liquids, and even gases. A well-known example is the blue charge-transfer band exhibited by iodine when combined with starch.

In inorganic chemistry, most charge-transfer complexes involve electron transfer between metal atoms and ligands. The charge-transfer bands in transition metal complexes result from movement of electrons between molecular orbitals (MO) that are predominantly metal in character and those that are predominantly ligand in character. If the electron moves from the MO with ligand like character to the metal like one, the complex is called ligand-to-metal charge-transfer (LMCT) complex. If the electron moves from the MO with metal like character to the ligand-like one, the complex is called a metal-to-ligand charge-transfer (MLCT) complex. Thus, a MLCT results in oxidation of the metal center whereas a LMCT results in the reduction of the metal center. Resonance Raman

Spectroscopy is also a powerful technique to assign and characterize charge-transfer bands in these complexes.

Donor-acceptor association equilibrium

Charge-transfer complexes are formed by weak association of molecules or molecular subgroups, one acting as an electron donor and another as an electron acceptor. The association does not constitute a strong covalent bond and is subject to significant temperature, concentration, and host (e.g., solvent) dependencies.

The charge-transfer association occurs in a chemical equilibrium with the independent donor (D) and acceptor (A) molecules:

$$D + A \rightleftharpoons DA$$

Quantum mechanically, this is described as a resonance between the non-bonded state |D, A> and the dative state $|D^+...A^->$. The formation of the dative state is an electronic transition giving rise to the colorful absorption bands.

The intensity of charge-transfer bands in the absorbance spectrum is strongly dependent upon the degree (equilibrium constant) of this association reaction. Methods have been developed to determine the equilibrium constant for these complexes in solution by measuring the intensity of absorption bands as a function of the concentration of donor and acceptor components in solution. The methods were first described for the association of iodine dissolved in aromatic hydrocarbons. The procedure is called the Benesi-Hildebrand method, named after the authors of the study.

Charge-transfer transition energy

The color of charge-transfer bands, i.e., the charge-transfer transition energy, is characteristic of the specific type of donor and acceptor entities.

The electron donating power of a donor molecule is measured by its ionization potential which is the energy required to remove an electron from the highest occupied molecular orbital. The electron accepting power of the electron acceptor is determined by its electron affinity which is the energy released when filling the lowest unoccupied molecular orbital.

The overall energy balance (ΔE) is the energy gained in a spontaneous charge transfer. It is determined by the difference between the acceptor's electron affinity (E_A) and the donor's ionization potential (E_I), adjusted by the resulting electrostatic attraction (J) between donor and acceptor:

$$\Delta E = E_A - E_I + J$$

The positioning of the characteristic CT bands in the electromagnetic spectrum is directly related to this energy difference and the balance of resonance contributions of non-bonded and dative states in the resonance equilibrium.

Identification of CT bands

Charge-transfer complexes are identified by

- *Color*: The color of CT complexes is reflective of the relative energy balance resulting from the transfer of electronic charge from donor to acceptor.
- *Solvatochromism*: In solution, the transition energy and therefore the complex color varies with variation in solvent permittivity, indicating variations in shifts of electron density as a result of the transition. This distinguishes it from the $\pi^* \leftarrow \pi$ transitions on the ligand.
- *Intensity*: CT absorptions bands are intense and often lie in the ultraviolet or visible portion of the spectrum. For inorganic complexes, the typical molar absorptivities, ε, are about 50000 L mol^{-1} cm^{-1}, that are three orders of magnitude higher than typical ε of 20 L mol^{-1} cm^{-1} or lower, for d-d transitions (transition from t_2g to e_g). This is because the CT transitions are spin-allowed and Laporte-allowed. However, d-d transitions are only spin-allowed; they are Laporte-forbidden.

Inorganic charge-transfer complexes

Charge-transfer occurs often in inorganic ligand chemistry involving metals. Depending on the direction of charge transfer they are either classified as ligand-to-metal (LMCT) or metal-to-ligand (MLCT) charge transfer.

Ligand-to-metal charge transfer

LMCT complexes arise from transfer of electrons from MO with ligand like character to those with metal like character. This type of transfer is predominant if complexes have ligands with relatively high energy lone pairs (example S or Se) or if the metal has low lying empty orbitals. Many such complexes have metals in high oxidation states (even d^0). These conditions imply that the acceptor level is available and low in energy.

Consider a d^6 octahedral complex (example $IrBr_6^{3-}$). The t_2g levels are filled as shown in Figure 1. Consequently an intense absorption is observed around 250 nm corresponding to a transition from ligand σ MO to the empty e_g MO. However, in $IrBr_6^{2-}$ that is a d^5 complex two absorptions, one near 600 nm and another near 270 nm, are observed. This is because two transitions are possible, one to t_{2g} (that can now accommodate one more electron) and another to e_g. The 600 nm band corresponds to transition to the t_{2g} MO and the 270 nm band to the e_g MO.

Figure 1. MO diagram showing ligand to metal charge transfer for a d^6 octahedral complex Another thing to note is that CT bands might also arise from transfer of electrons from nonbonding orbitals of the ligand to the e_g MO.

Trend of LMCT energies

Oxidation Number

+7 MnO_4^- < TcO_4^- < ReO_4^-
+6 CrO_4^{2-} < MoO_4^{2-} < WO_4^{2-}
+5 VO_4^{3-} < NbO_4^{3-} < TaO_4^{3-}

The energies of transitions correlate with the order of the electrochemical series. The metal ions that are most easily reduced correspond to the lowest energy transitions. The above trend is consistent with transfer of electrons from the ligand to the metal, thus resulting in a reduction of metal ions by the ligand.

Examples include:

1. MnO_4^- : The permanganate ion having tetrahedral geometry is intensely purple due to strong absorption involving charge transfer from MO derived primarily from filled oxygen p orbitals to empty MO derived from manganese(VII).
2. CdS: The color of artist's pigment cadmium yellow is due to transition from Cd^{2+} (5s) ← $S^{2-}(\pi)$.
3. HgS: it is red due to Hg^{2+} (6s) ← $S^{2-}(\pi)$ transition.
4. Fe Oxides: they are red and yellow due to transition from Fe (3d) ← $O^{2-}(\pi)$.

Metal-to-ligand charge transfer

Metal-to-ligand charge-transfer (MLCT) complexes arise from transfer of electrons from MO with metal like character to those with ligand like character. This is most commonly observed in complexes with ligands having low-lying π^* orbitals especially aromatic ligands. The transition will occur at low energy if the metal ion has a low oxidation number for its d orbitals will relatively be high in energy.

Examples of such ligands taking part in MLCT include 2,2'-bipyridine (bipy), 1,10-phenanthroline (phen), CO, CN^- and SCN^-. Examples of these complexes include:

1. Tris(2,2'-bipyridyl)ruthenium(II) : This orange colored complex is being studied as the excited state resulting from this charge transfer has a lifetime of microseconds and the complex is a versatile photochemical redox reagent.
2. $W(CO)_4$(phen)
3. $Fe(CO)_3$(bipy)

Photoreactivity of MLCT excited states

The photoreactivity of MLCT complexes result from the nature of the oxidized metal and the reduced ligand. Though the states of traditional MLCT complexes like $Ru(bipy)_3^{2+}$ and $Re(bipy)(CO)_3Cl$ were intrinsically not reactive, several MLCT complexes have been synthesized that are characterized by reactive MLCT states.

Vogler and Kunkely considered the MLCT complex to be an isomer of the ground state which contains an oxidized metal and reduced ligand. Thus various reactions like electrophilic attack and radical reactions on the reduced ligand, oxidative addition at the metal center due to the reduced ligand, and outer sphere charge-transfer reactions can be attributed to states arising from MLCT transitions. MLCT states' reactivity often depends on the oxidation of the metal. Subsequent processes include associative ligand substitution, exciplex formation and cleavage of metal---metal bonds.

Color of charge-transfer complexes

Many metal complexes are colored due to d-d electronic transitions. Visible light of the correct wavelength is absorbed, promoting a lower d-electron to a higher excited state. This absorption of light causes color. These colors are usually quite faint, though. This is because of two selection rules:

The spin rule: $\Delta S = 0$

On promotion, the electron should not experience a change in spin. Electronic transitions which experience a change in spin are said to be *spin forbidden*.

Laporte's rule: $\Delta l = \pm 1$

d-d transitions for complexes which have a center of symmetry are forbidden - *symmetry forbidden* or *Laporte forbidden*.

Charge-transfer complexes do not experience d-d transitions. Thus, these rules do not apply and the absorptions are generally very intense.

For example, the classic example of a charge-transfer complex is that between iodine and starch to form an intense purple color. This has widespread use as a rough screen for counterfeit currency. Unlike most paper, the paper used in US currency is not sized with starch. Thus, formation of this purple color on application of an iodine solution indicates a counterfeit.

Other examples

Hexaphenylbenzenes like **H** (fig. 3) lend themselves extremely well to forming charge-transfer complexes. Cyclic voltammetry for **H** displays 4 well separated maxima corresponding to H^+ right up to H^{4+} with the first ionization at $E_{1/2}$ of only 0.51 eV.

oxidation of these arenes by for instance dodecamethylcarboranyl (**B**) to the blue crystal solid H^+B^- complex is therefore easy.

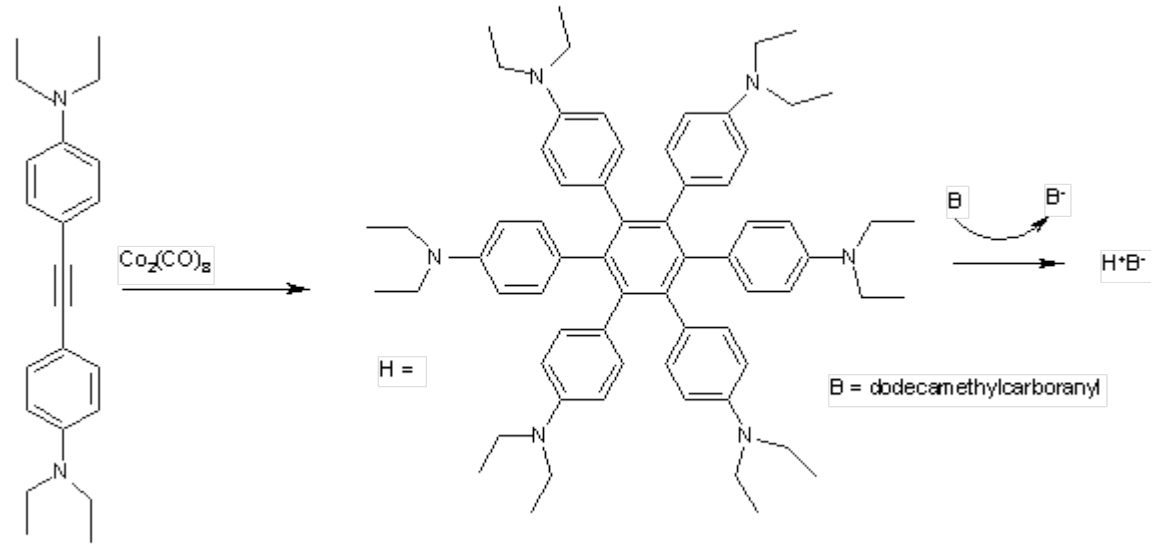


Fig. 3 Synthesis of H^+B^- complex: Alkyne trimerisation of bisubstituted alkyne with dicobalt octacarbonyl, delocalization is favored with activating groups such as a di(ethylamino) group

The phenyl groups are all positioned in an angle of around 45° with respect to the central aromatic ring and the positive charge in the radical cation is therefore through space delocalised through the 6 benzene rings in the shape of a toroid. The complex has 5 absorption bands in the near infrared region which can be assigned to specific electronic transitions with the aid of deconvolution and the Mulliken-Hush theory.

Electrical conductivity

In 1954 researchers at Bell Laboratories and elsewhere reported charge-transfer complexes with resistivities as low as 8 ohms·cm in combinations of perylene with iodine or bromine. In 1962, the well-known acceptor tetracyanoquinodimethane (TCNQ) was reported. Tetrathiafulvalene (TTF) was synthesized in 1970 and found to be a strong electron donor. In 1973 it was discovered that a combination of these components formed a strong charge-transfer complex, henceforth referred to as TTF-TCNQ. The complex is formed in solution and may be crystallized into a well-formed crystalline solid. The solid shows almost metallic electrical conductance and was the first discovered purely organic conductor. In a TTF-TCNQ crystal, TTF and TCNQ molecules are arranged independently in separate parallel-aligned stacks and an electron transfer occurs from donor (TTF) to acceptor (TCNQ) stacks. Hence, electrons and electron holes are separated and concentrated in the stacks and can traverse in a one-dimensional direction along the TCNQ and TTF columns, respectively, when an electric potential is applied to the ends of a crystal in the stack direction.

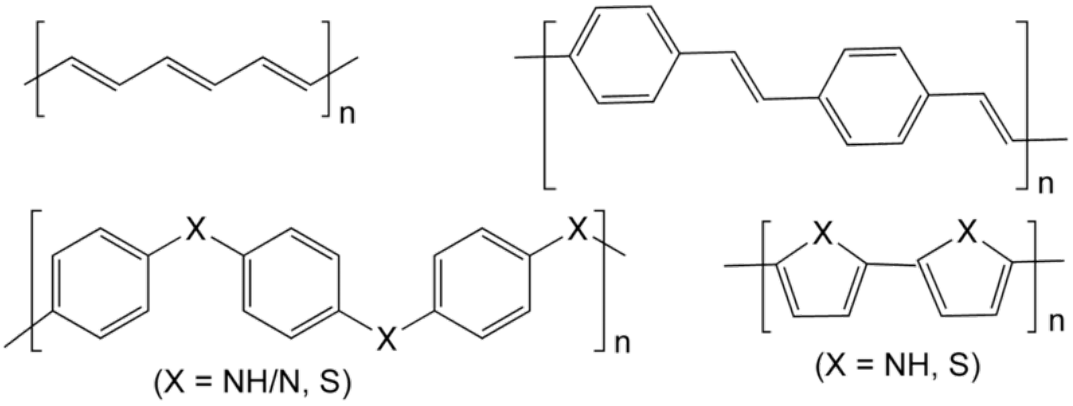
The first organic molecule that forms a superconductor was discovered in 1980. Tetramethyl-tetraselenafulvalene-phosphorus hexafloride ($TMTSF_2PF_6$), a semi-conductor at ambient conditions, shows superconductivity at low temperature (critical temperature) and high pressure: 0.9 K and 12 kbar. Since 1980, many organic superconductors have been synthesized, and the critical temperature has been raised to over 100 K as of 2001 . Unfortunately, critical current densities in these complexes are very small.

CT complexation involved in diseases

In humans, elevated systemic levels of transition-series metals, electron-donors, etc. are associated with specific disease symptoms. These include psychosis, movement disorders, pigmentary abnormalities, and deafness. This may involve charge-transfer complexes with the Melanin in the midbrain, skin, and the stria vascularis of the inner ear.

Chapter- 4

Conductive Polymer



Conductive polymers or more precisely **intrinsically conducting polymers** (ICPs) are organic polymers that conduct electricity. Such compounds may have metallic conductivity or be semiconductors. The biggest advantage of conductive polymers is their processability. Conductive polymers are also plastics, which are organic polymers. Therefore, they can combine the mechanical properties (flexibility, toughness, malleability, elasticity, etc.) of plastics with high electrical conductivity. These properties can be fine-tuned using the methods of organic synthesis.

Correlation of chemical structure and electrical conductivity

In traditional polymers such as polyethylenes, the valence electrons are bound in sp^3 hybridized covalent bonds. Such "sigma-bonding electrons" have low mobility and do not contribute to the electrical conductivity of the material. The situation is completely different in conjugated materials. Conducting polymers have backbones of contiguous sp^2 hybridized carbon centers. One valence electron on each center resides in a p_z orbital, which is orthogonal to the other three sigma-bonds. The electrons in these delocalized orbitals have high mobility when the material is "doped" by oxidation, which removes some of these delocalized electrons. Thus the conjugated p-orbitals form a one-

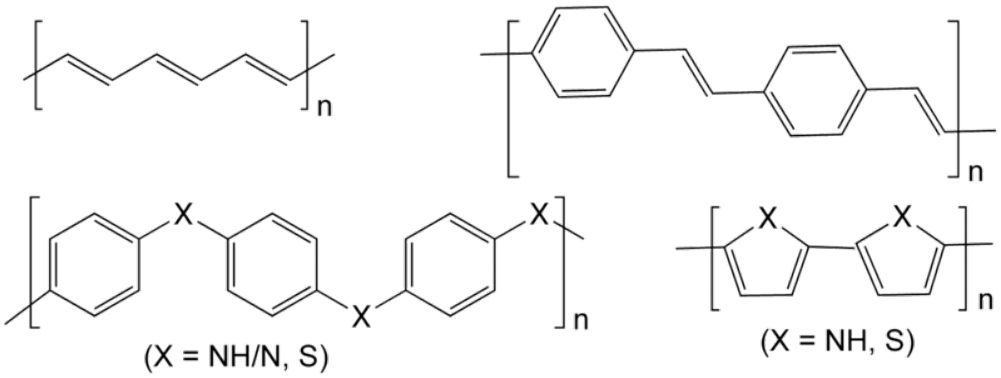
dimensional electronic band, and the electrons within this band become mobile when it is partially emptied. The band structures of conductive polymers can easily be calculated with a tight binding model. In principle, these same materials can be doped by reduction, which adds electrons to an otherwise unfilled band. In practice, most organic conductors are doped oxidatively to give p-type materials. The redox doping of organic conductors is analogous to the doping of silicon semiconductors, whereby a small fraction silicon atoms are replaced by electron-rich (e.g., phosphorus) or electron-poor (e.g. boron) atoms to create n-type and p-type semiconductors, respectively.

Although typically "doping" conductive polymers involves oxidizing or reducing the material, conductive organic polymers associated with a protic solvent may also be "self-doped."

The most notable difference between conductive polymers and inorganic semiconductors is the mobility, which until very recently was dramatically lower in conductive polymers than their inorganic counterparts. This difference is diminishing with the invention of new polymers and the development of new processing techniques. Low charge carrier mobility is related to structural disorder. In fact, as with inorganic amorphous semiconductors, conduction in such relatively disordered materials is mostly a function of "mobility gaps" with phonon-assisted hopping, polaron-assisted tunneling, etc., between localized states. Recently, it has been reported that Quantum Decoherence on localized electron states might be the fundamental mechanism behind electron transport in conductive polymers.

The conjugated polymers in their undoped, pristine state are semiconductors or insulators. As such, the energy gap can be > 2 eV, which is too great for thermally activated conduction. Therefore, undoped conjugated polymers, such as polythiophenes, polyacetylenes only have a low electrical conductivity of around 10^{-10} to 10^{-8} S/cm. Even at a very low level of doping (< 1%), electrical conductivity increases several orders of magnitude up to values of around 0.1 S/cm. Subsequent doping of the conducting polymers will result in a saturation of the conductivity at values around 0.1–10 kS/cm for different polymers. Highest values reported up to now are for the conductivity of stretch oriented polyacetylene with confirmed values of about 80 kS/cm. Although the pi-electrons in polyactetylene are delocalized along the chain, pristine polyacetylene is not a metal. Polyacetylene has alternating single and double bonds which have lengths of 1.44 and 1.36 Å, respectively. Upon doping, the bond alteration is diminished in conductivity increases. Non-doping increases in conductivity can also be accomplished in a field effect transistor (organic FET or OFET) and by irradiation. Some materials also exhibit negative differential resistance and voltage-controlled "switching" analogous to that seen in inorganic amorphous semiconductors.

Classes of materials



Structures of various conductive organic polymers. Clockwise; polyacetylene, polyphenylenevinylene, polypyrrole (X = NH), and polythiophene (X = S), polyaniline (X = N, NH) and polyphenylene sulfide (X = S).

The following table presents some organic conductive polymers according to their composition. **The well-studied classes are written in bold** and *the less well studied ones are in italic*.

The main chain contains	**No heteroatom**	**NH or NR**	**S**
Aromatic cycles	*Poly(fluorene)s, polypyrenes, polyazulenes, polynaphthalenes*	The N is in the aromatic cycle: **poly(pyrrole)s (PPY)**, *polycarbazoles, polyindoles, polyazepines*... The N is outside the aromatic cycle: **polyanilines (PANI)**	The S is in the aromatic cycle: **poly(thiophene)s (PT)**... The S is outside the aromatic cycle: **poly(p-phenylene sulfide) (PPS)**
Double bonds	**Poly(acetylene)s (PAC)**		
Aromatic cycles and double bonds	**Poly(p-phenylene vinylene) (PPV)**		

PPV and its soluble derivatives have emerged as the prototypical electroluminescent semiconducting polymers. Today, poly(3-alkylthiophenes) are the archetypical materials for solar cells and transistors.

Synthesis of conductive polymers

Many methods for the synthesis of conductive polymers have been developed. Most conductive polymers are prepared by oxidative coupling of monocyclic precursors. Such reactions entail dehydrogenation:

$$n\ H\text{–}[X]\text{–}H \rightarrow H\text{–}[X]_n\text{–}H + 2(n\text{–}1)\ H^+ + 2(n\text{–}1)\ e^-$$

One challenge is usually the low solubility of the polymer. This has been addressed by some researchers through the formation of nanostructures and surfactant stabilized conducting polymer dispersions in water. These include polyaniline nanofibers and PEDOT:PSS. These materials have lower molecular weights than that of some materials previously explored in the literature. However, in some cases, the molecular weight need not be high to achieve the desired properties.

Properties and applications

Conductive polymers enjoy few large-scale applications due to their poor processability. They have been known to have promise in antistatic materials and they have been incorporated into commercial displays and batteries, but there have had limitations due to the manufacturing costs, material inconsistencies, toxicity, poor solubility in solvents, and inability to directly melt process. Literature suggests they're also promising in organic solar cells, printing electronic circuits, organic light-emitting diodes, actuators, electrochromism, supercapacitors, biosensors, flexible transparent displays, electromagnetic shielding and possibly replacement for the popular transparent conductor indium tin oxide. Conducting polymers are rapidly gaining attraction in new applications with increasingly processable materials with better electrical and physical properties and lower costs. The new nanostructured forms of conducting polymers particularly, provide fresh air to this field with their higher surface area and better dispersability.

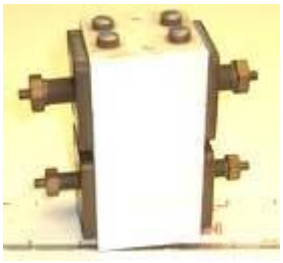
Electroluminescence

Electroluminescence is light emission stimulated by electrical current. In organic compounds, electroluminescence has been known since the early 1950s, when Bernanose and coworkers first produced electroluminescence in crystalline thin films of acridine orange and quinacrine. In 1960, researchers at Dow Chemical developed AC-driven electroluminescent cells using doping. In some cases, similar light emission is observed when a voltage is applied to a thin layer of a conductive organic polymer film. While electroluminescence was originally mostly of academic interest, the increased conductivity of modern conductive polymers means enough power can be put through the device at low voltages to generate practical amounts of light. This property has led to the development of flat panel displays using Organic LEDs, solar panels, and optical amplifiers.

Barriers to applications

Since most conductive polymers require oxidative doping, the properties of the resulting state are crucial. Such materials are often salt-like, which diminishes their solubility in organic solvents and hence their processability. Furthermore, the charged organic backbone is often unstable towards atmospheric moisture. Compared to metals, organic conductors can be expensive requiring multi-step synthesis. The poor processability for many polymers requires the introduction of solubilizing substituents, which can further complicate the synthesis.

History



voltage-controlled switch, an organic polymer electronic device from 1974. Now in the Smithsonian Chip collection.

There are multiple reviews of the history of the field. The first report on polyaniline goes back to the discovery of aniline. In the mid-19th century, Letheby reported the electrochemical and chemical oxidation products of aniline in acidic media, noting that reduced form was colourless but the oxidized forms were deep blue. In the early 20th century, German chemists named several compounds "aniline black" and "pyrrole black" and used them industrially. Classically, such polymer "blacks", their parent compound polyacetylene, and their co-polymers were called "Melanins".

The first highly-conductive polymers were mixtures of plastics with embedded conducting particles (metal powders, carbon black...) or coated plastics. For example, page 52 of the December 26, 1949, issue of LIFE magazine shows a light bulb in a circuit without wires. The conductor was Markite, a clear pyrolyzed inorganic silica. The conductivity of Markite ranges from as low as distilled water to as high as mercury.

In the 1950s, researchers reported that polycyclic aromatic compounds formed semi-conducting charge-transfer complex salts with halogens. This indicated that organic compounds could carry current. While organic conductors were previously intermittently discussed, the field was particularly energized by the prediction of superconductivity following the discovery of BCS theory.

As for "pure" conductive organic polymers—in 1963, Bolto and co-workers reported conductivity in iodine-doped polypyrroles. This Australian group eventually claimed to reach resistivities as low as 0.03 ohm·cm with other conductive organic polymers. This resistivity is roughly equivalent to present-day efforts.

Subsequently, DeSurville and coworkers reported high conductivity in a polyaniline. Similarly, in 1980, Diaz and Logan reported films of polyaniline that could serve as electrodes.

Similarly, much early work on the physics and chemistry of conductive polymers was done under the melanin rubrick. This was because of the medical relevance of this material. For example, in the 1960s Blois *et al.* showed semiconduction in melanins, as well as further defining their physical structures and properties Nicolaus *et al.* further defined the conductive polymer structures. Classically, all polyacetylenes, polypyrroles and polyanilines are melanins, "The most simple melanin can be considered the acetylene-black from which it is possible to derive all the others.. Substitution does not qualitatively influence the physical properties like conductivity, colour, EPR, which remain unaltered."

In 1974, McGinness and coworkers described an "active" organic-polymer electronic device, a voltage-controlled bistable switch. This device used DOPA-melanin, a well-characterized self-doping copolymer of polyaniline, polypyrrole, and polyacetylene. The "ON" state of this device exhibited low conductivity with switching, with as much as five orders of magnitude shifts in current. Their material also exhibited classic negative differential resistance.

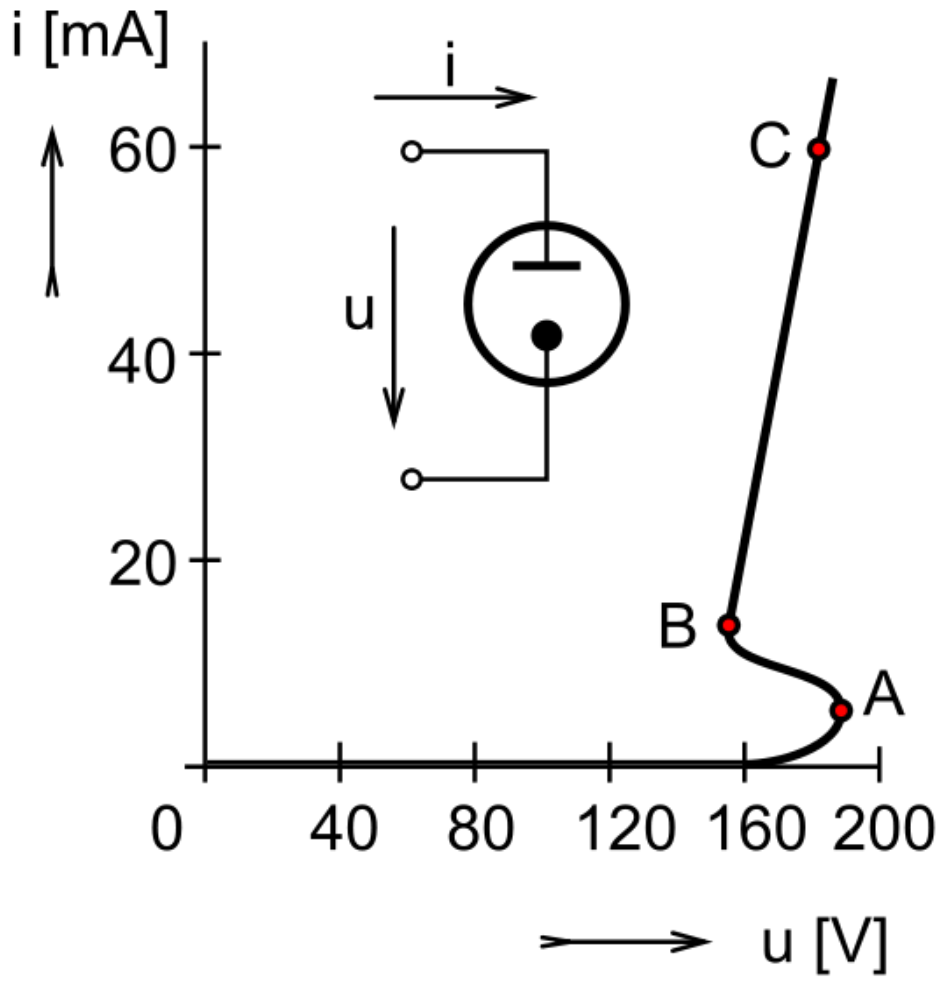
In 1977, Alan J. Heeger, Alan MacDiarmid and Hideki Shirakawa reported similar high conductivity in oxidized iodine-doped polyacetylene. This research earned them the 2000 Nobel prize in Chemistry *"For the discovery and development of conductive polymers."* Some reviewers have questioned the Nobel citation's *discovery* assignment. Thus, Inzelt notes that, while the Nobelists deserve credit for publicising and popularizing the field, conductive polymers were " *..produced, studied and even applied* " well before their work.

Trends

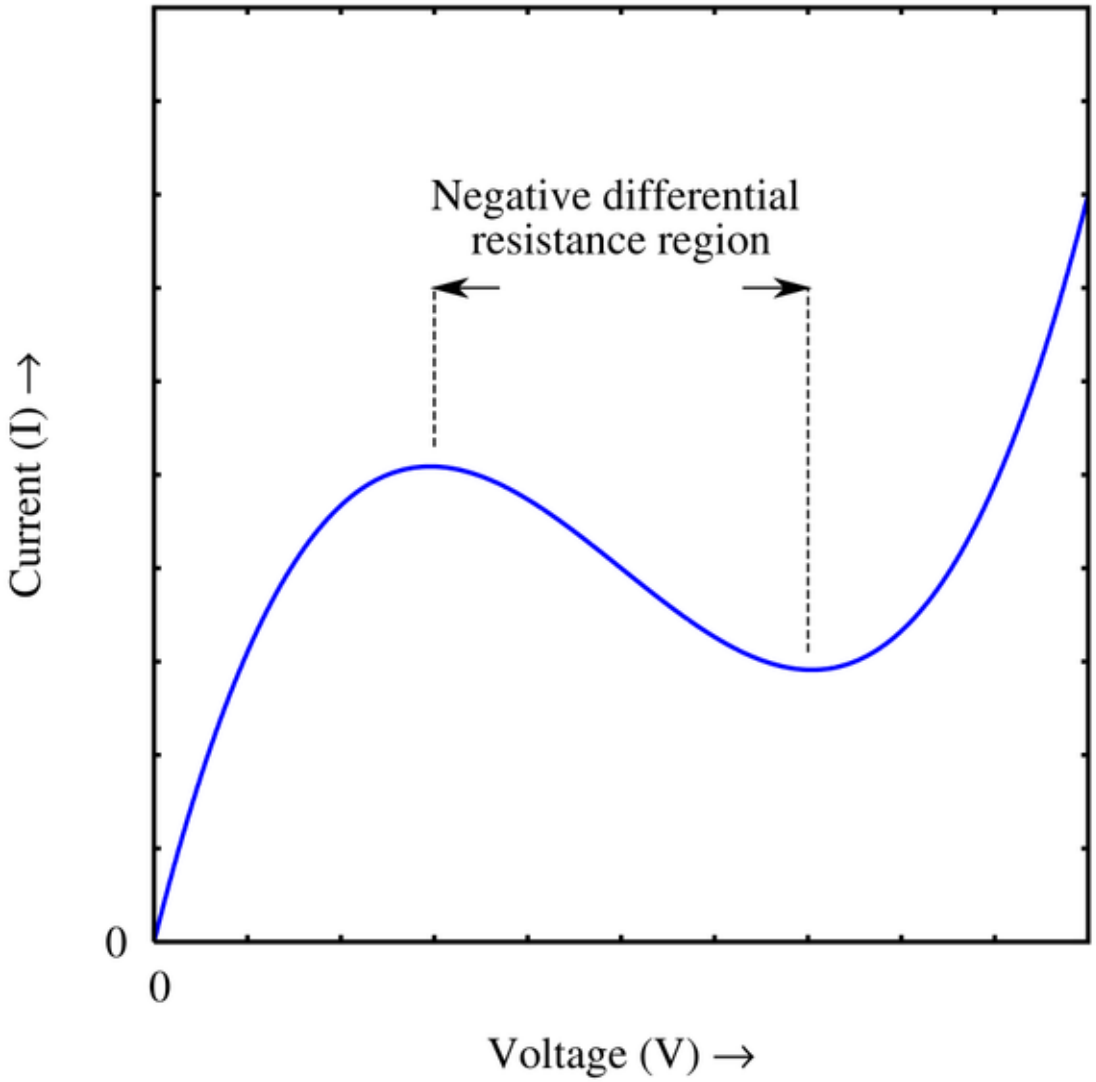
Most recent emphasis is on organic light emitting diodes and organic polymer solar cells. The Organic Electronics Association is an international platform to promote applications of organic semiconductors. Conductive polymer products with embedded and improved electromagnetic interference (EMI) and electrostatic discharge (ESD) protection have led to both prototypes and products. For example, Polymer Electronics Research Center at University of Auckland is developing a range of novel DNA sensor technologies based on conducting polymers, photoluminescent polymers and inorganic nanocrystals (quantum dots) for simple, rapid and sensitive gene detection. Typical conductive polymers must be "doped" to produce high conductivity. To date, there remains to be discovered an organic polymer that is *intrinsically* electrically conducting.

Chapter- 5

Negative Resistance



A neon lamp exhibits S-shaped negative differential resistance in A–B section of its IV curve where an increase in the current results in a decreased voltage; total resistance is still positive.



A tunnel diode exhibits N-shaped negative differential resistance in the middle region of its IV curve where an increase in the voltage results in a decreased current; total resistance is still positive.

Negative resistance is a property of some electric circuits where current through and voltage across the same port change in opposite directions. This is in contrast to a simple ohmic resistor, where current and voltage change in the same direction under the same conditions. Negative resistors are theoretical and do not exist as a discrete component. However, some types of diodes (e.g., tunnel diodes) can be built that exhibit negative resistance in some part of their operating range. Such a differential negative resistance can be illustrated with a resonant tunneling diode. Similarly, some chalcogenide glasses, organic semiconductors, and conductive polymers exhibit a similar region of negative resistance as a bulk property of the material.

Properties

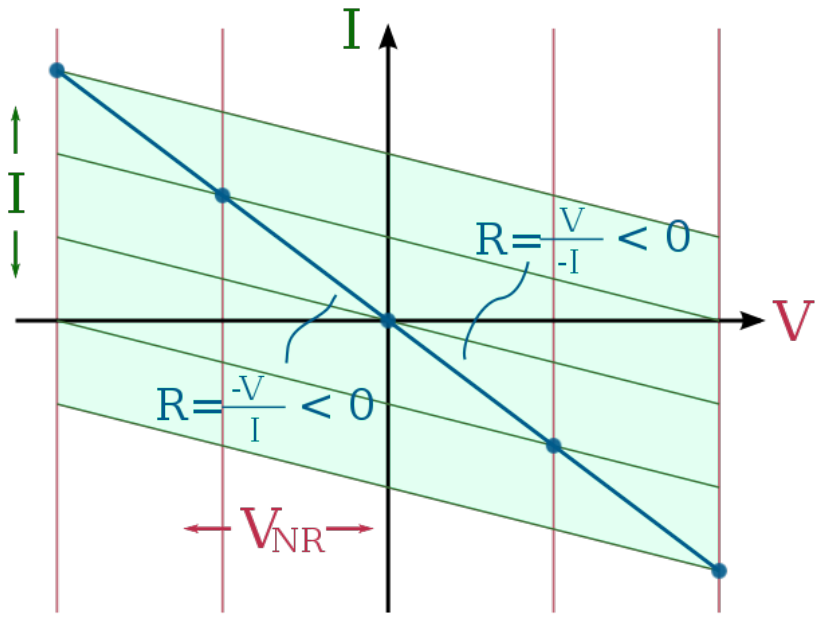


Figure 2: The IV curve of a theoretical negative resistor

Fig. 2 shows a graph of a negative resistor, showing the negative slope. In contrast to this, a resistor will have a positive slope. Tunnel diodes and Gunn diodes exhibit a negative resistance region in their IV (current - voltage) curve. They have two terminals like a resistor; but are not linear devices. Unijunction transistors also have negative resistance properties when a circuit is built using other components.

For negative resistance to be present there must be active components in the circuit providing a source of energy. This is because current through a negative resistance implies a source of energy just as current through positive resistance implies that energy is being dissipated. A resistor produces voltage that is proportional to the current through it according to Ohm's law. The IV curve of a true negative resistor has a negative slope and passes through the origin of the coordinate system (the curve can only enter the 2nd and 4th quadrants if energy is being supplied). This is to be compared with devices such as the tunnel diode where the negative slope portion of the curve does not pass through the origin. Clearly, there is no source of energy in a two terminal diode.

History

In early research it was noticed that arc discharge devices and some vacuum tube devices such as the dynatron exhibit negative differential resistance effects. Practical and economic devices only became available with solid state technology. The typical true negative impedance circuit—the negative impedance converter -- is due to John G. Linvill (1953) and the popular element with negative differential resistance—the tunnel diode -- is due to Leo Esaki (1958).

Implementations

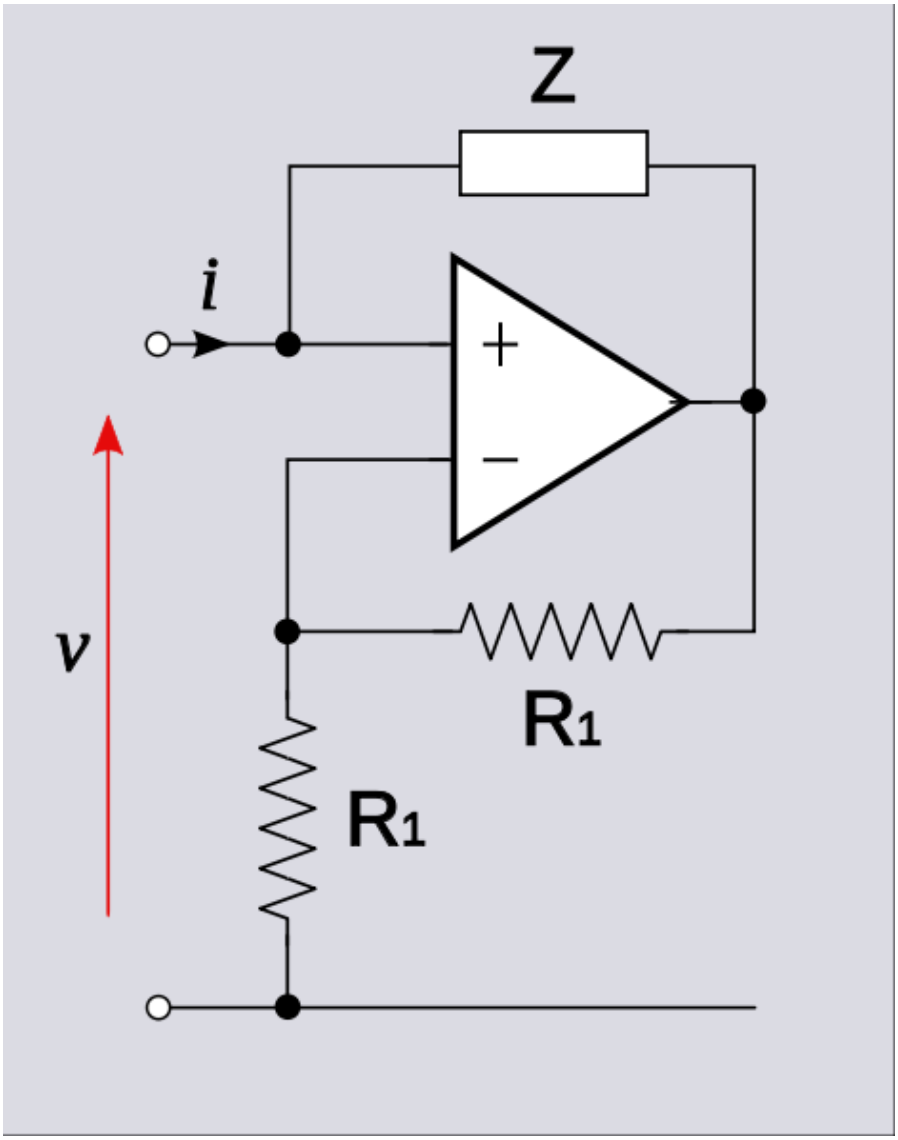


Figure 3: Negative impedance circuit with $Z_{\text{in}} \triangleq \frac{v}{i} = -Z$

Diodes

Tunnel diodes are heavily doped semiconductor junctions that have an "N" shaped transfer curve. A vacuum tube can also be made to exhibit negative resistance.. Other negative resistance diodes have been built that have an "S" shaped transfer curve. When biased so that the operating point is in the negative resistance region, these devices can be used as an Amplifier. These devices can also be biased so that they will switch between two states very quickly, as the applied voltage changes.

Operational Amplifiers

The negative resistance circuit shown in Figure 3 is an opamp implementation of the negative impedance converter (see below). The two resistors R1 and the op amp constitute a negative feedback non-inverting amplifier with gain A = 2. In the case $Z = R$, the input resistance (for an ideal opamp) is given by;

$$R_{\text{in}} = -Z = -R$$

The input port of the circuit can be connected into another network as if it were a negative resistance component.

In the general case Z can be selected to produce negative capacitances or negative inductances.

Applications

Oscillators

All feedback oscillators imply the presence of negative resistance. There are many such topologies, including the Dynatron oscillator, Colpitts oscillator, Hartley oscillator, Wien bridge oscillator, and some types of relaxation oscillators. If the feedback loop is broken and the input impedance examined it will be found to include negative resistance. Negative resistance characteristics of Gunn diodes are often used in microwave frequencies as well.

Amplifiers

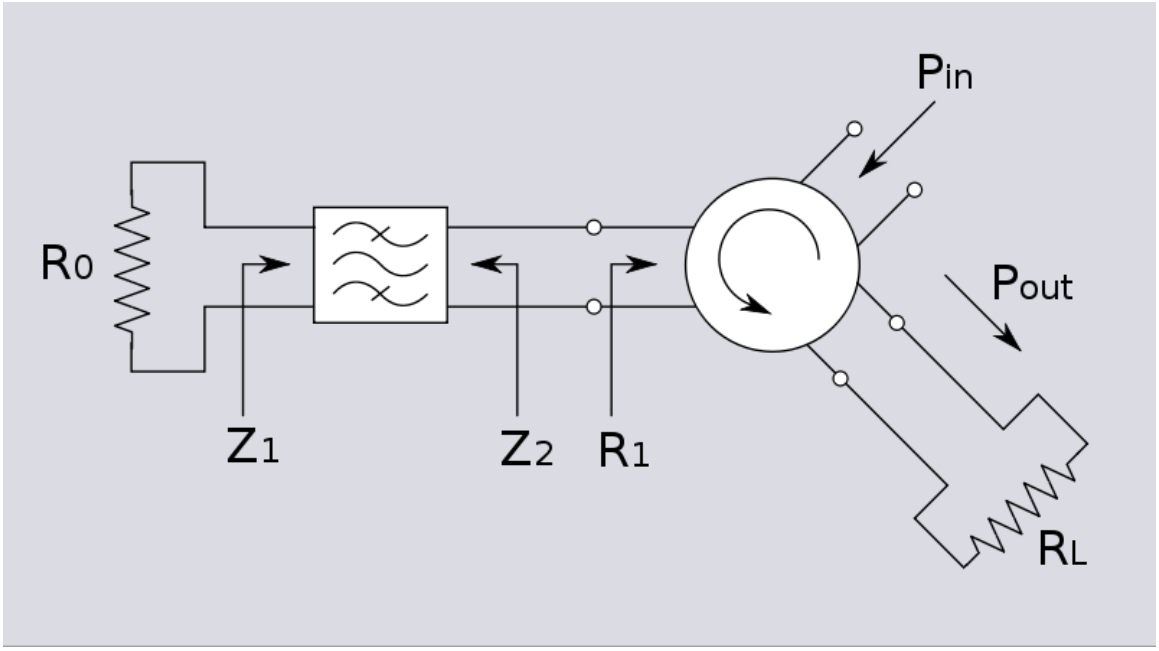


Figure 4: Negative resistance microwave amplifier using a circulator

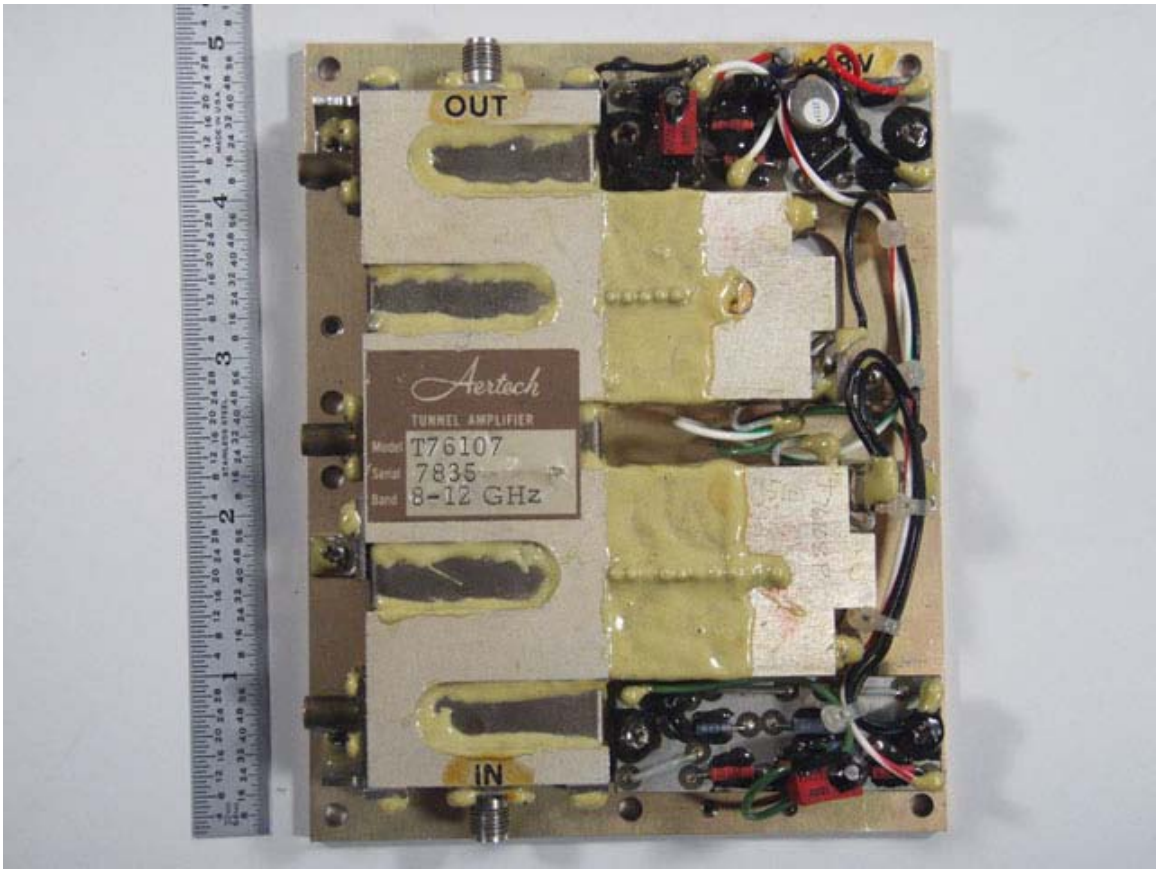


Figure 5: 8 - 12 GHz tunnel diode amplifier, circa 1970

A device exhibiting negative resistance can be used to amplify a signal and this is an especially useful technique at microwave frequencies. Such devices do not present as pure negative resistance at these frequencies (in the case of the tunnel diode a large parallel capacitance is also present) and a matching filter is usually required. The reactive components of the device's equivalent circuit can be absorbed into the filter design so the circuit can be represented as a pure resistance followed by a bandpass filter. The output of this arrangement is fed into one port of a three-port circulator. The other two ports constitute the input and output of the amplifier with the direction of circulation as shown in the diagram. Treating R_0 as being positive, the reflection coefficients at the two ends of the filter are given by;

$$\Gamma_1 = \frac{Z_1 - R_0}{Z_1 + R_0} \text{and,} \quad \Gamma_2 = \frac{Z_2 - R_1}{Z_2 + R_1}$$

Since the filter has no resistive elements, there is no dissipation and the magnitudes of the two reflection coefficients must be equal,

$$|\Gamma_1| = |\Gamma_2|$$

The input power entering the cirulator is directed at the matching filter, is reflected at both the input and output of the filter and a portion finally arrives at the load. This portion is given by;

$$\frac{P_{\text{out}}}{P_{\text{in}}} = |\Gamma_1|^2$$

For a well matched filter, the reflection coefficients will be very small in the passband and very little power will reach the load. On the other hand if R_0 is replaced by a negative resistance such that,

$$R_0' = -R_0 \text{then,}$$
$$\Gamma_1' = \frac{Z_1 + R_0}{Z_1 - R_0} \text{and,}$$
$$|\Gamma_1'| = |\Gamma_2'| = \frac{1}{|\Gamma_1|}$$

Now the reflection coefficients are very large and more power is reaching the load than was injected in the input port. The net result of terminating one port in a negative resistance is amplification between the remaining two ports.

Mixers and frequency converters

The highly non-linear characteristics of tunnel diodes makes them useful as frequency mixers. The conversion gain of a tunnel diode mixer can be as high as 20 dB if it is biased to operate in the negative resistance region.

Antenna design

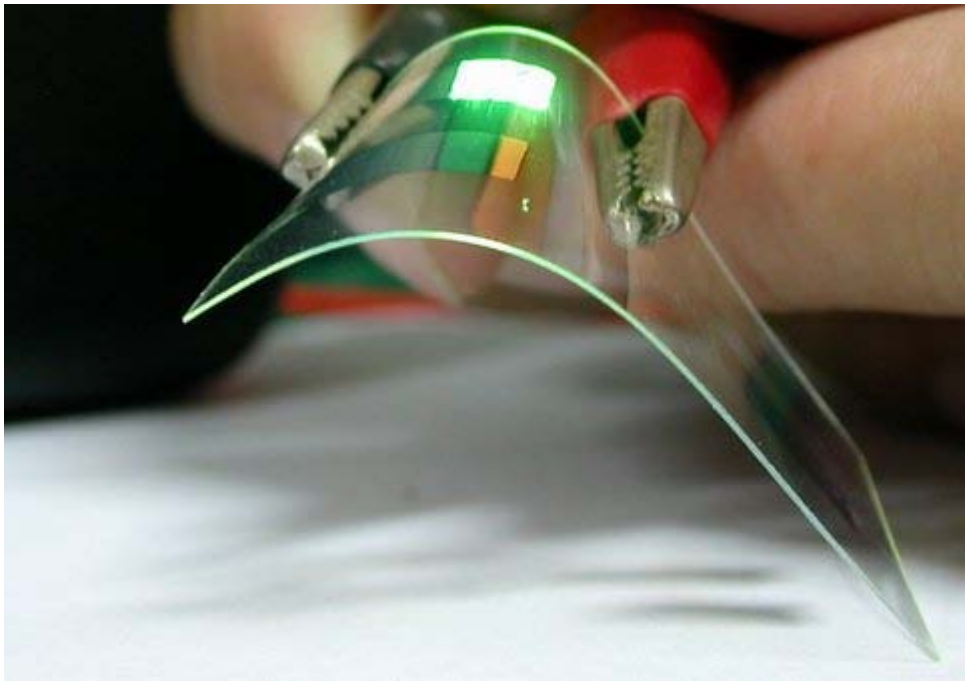
Another concept of negative resistance exists in the domain of radio frequency antenna design. This is also known as negative impedance. It is not uncommon for an antenna containing multiple driven elements to exhibit apparent negative impedance in one or more of the driven elements.

Impedance cancellation

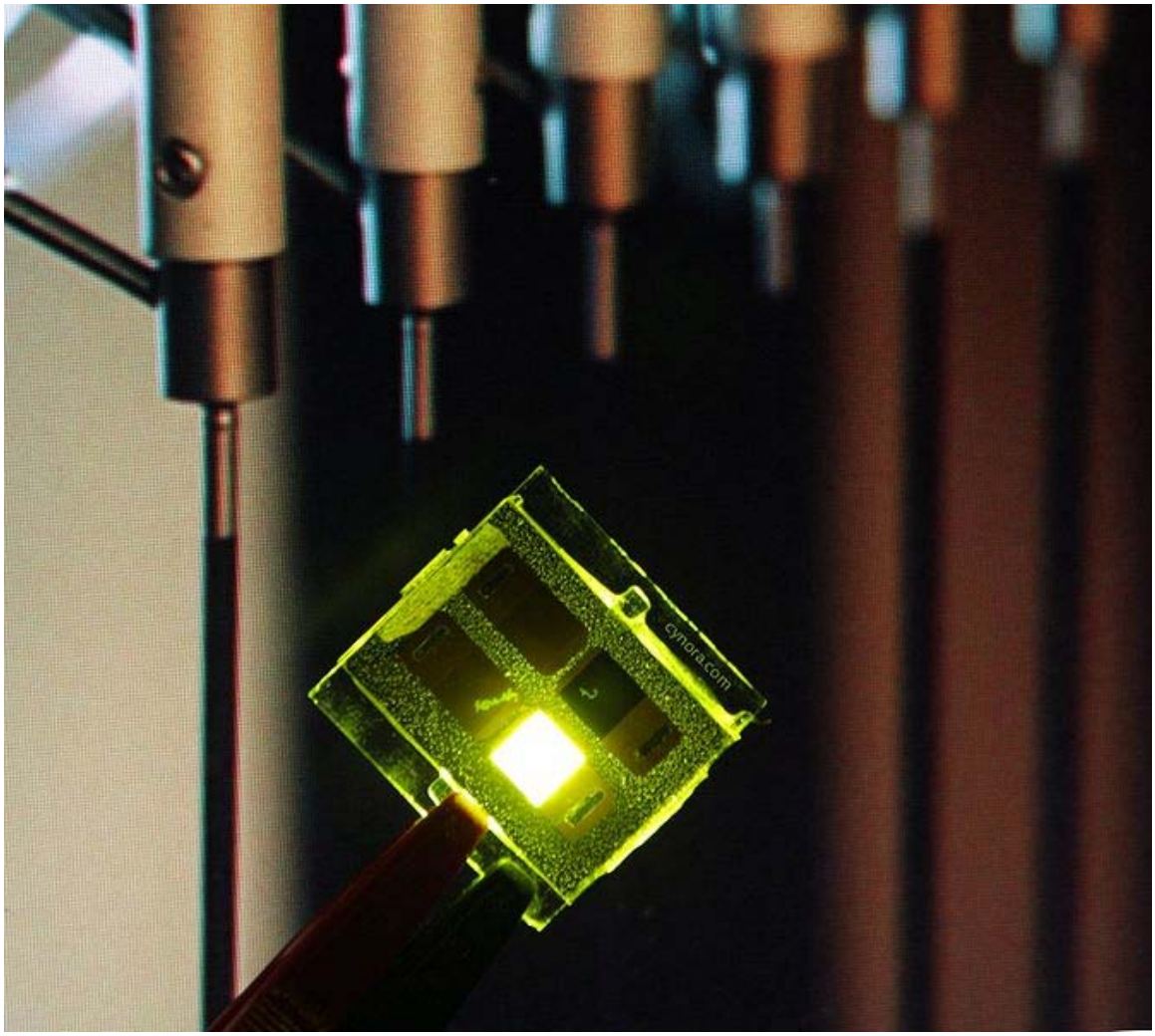
Negative impedances can be used to cancel the effects of positive impedances, for example, by eliminating the internal resistance of a voltage source or making the internal resistance of a current source infinite. This property is used in telephony line repeaters and in circuits such as the Howlan

Chapter- 6

Organic Light-Emitting Diode



Demonstration of a flexible OLED device



A green emitting OLED device

Sony XEL-1, the world's first OLED TV.

An **organic light emitting diode** (**OLED**) is a light-emitting diode (LED) in which the emissive electroluminescent layer is a film of organic compounds which emit light in response to an electric current. This layer of organic semiconductor material is situated between two electrodes. Generally, at least one of these electrodes is transparent.

OLEDs are used in television screens, computer monitors, small, portable system screens such as mobile phones and PDAs, watches, advertising, information and indication. OLEDs are also used in light sources for general space illumination and in large-area light-emitting elements. Due to their comparatively early stage of development, they typically emit less light per unit area than inorganic solid-state based LED point-light sources.

An OLED display functions without a backlight. Thus, it can display deep black levels and can also be thinner and lighter than established liquid crystal displays. Similarly, in low ambient light conditions such as dark rooms, an OLED screen can achieve a higher contrast ratio than an LCD screen using either cold cathode fluorescent lamps or the more recently developed LED backlight.

There are two main families of OLEDs: those based upon small molecules and those employing polymers. Adding mobile ions to an OLED creates a Light-emitting Electrochemical Cell or LEC, which has a slightly different mode of operation.

OLED displays can use either passive-matrix (PMOLED) or active-matrix addressing schemes. Active-matrix OLEDs (AMOLED) require a thin-film transistor backplane to switch each individual pixel on or off, and can make higher resolution and larger size displays possible.

History

The first observations of electroluminescence in organic materials were in the early 1950s by A. Bernanose and co-workers at the Nancy-Université, France. They applied high-voltage alternating current (AC) fields in air to materials such as acridine orange, either deposited on or dissolved in cellulose or cellophane thin films. The proposed mechanism was either direct excitation of the dye molecules or excitation of electrons.

In 1960, Martin Pope and co-workers at New York University developed ohmic dark-injecting electrode contacts to organic crystals. They further described the necessary energetic requirements (work functions) for hole and electron injecting electrode contacts. These contacts are the basis of charge injection in all modern OLED devices. Pope's group also first observed direct current (DC) electroluminescence under vacuum on a pure single crystal of anthracene and on anthracene crystals doped with tetracene in 1963 using a small area silver electrode at 400V. The proposed mechanism was field-accelerated electron excitation of molecular fluorescence.

Pope's group reported in 1965 that in the absence of an external electric field, the electroluminescence in anthracene crystals is caused by the recombination of a thermalized electron and hole, and that the conducting level of anthracene is higher in energy than the exciton energy level. Also in 1965, W. Helfrich and W. G. Schneider of the National Research Council in Canada produced double injection recombination electroluminescence for the first time in an anthracene single crystal using hole and electron injecting electrodes, the forerunner of modern double injection devices. In the same year, Dow Chemical researchers patented a method of preparing electroluminescent cells using high voltage (500–1500 V) AC-driven (100–3000 Hz) electrically-insulated one millimetre thin layers of a melted phosphor consisting of ground anthracene powder, tetracene, and graphite powder. Their proposed mechanism involved electronic excitation at the contacts between the graphite particles and the anthracene molecules.

Device performance was limited by the poor electrical conductivity of contemporary organic materials. However this was overcome by the discovery and development of highly conductive polymers.

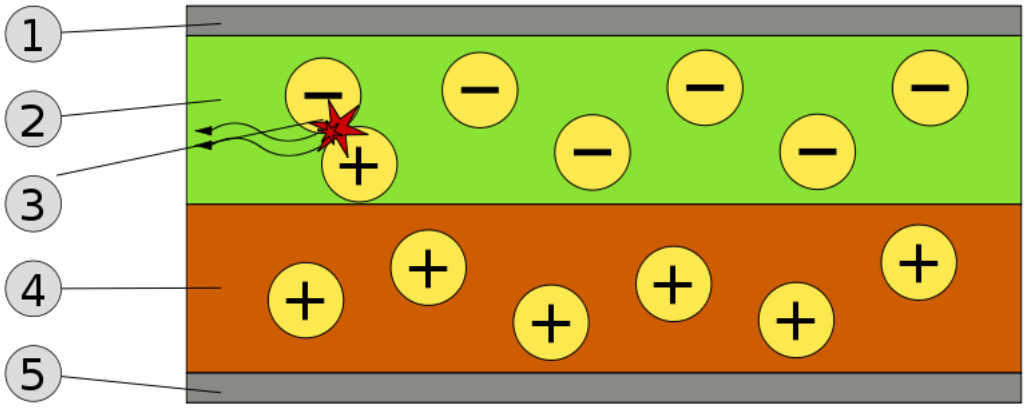
Electroluminescence from polymer films was first observed by Roger Partridge at the National Physical Laboratory in the United Kingdom. The device consisted of a film of

poly(n-vinylcarbazole) up to 2.2 micrometres thick located between two charge injecting electrodes. The results of the project were patented in 1975 and published in 1983.

The first diode device was reported at Eastman Kodak by Ching W. Tang and Steven Van Slyke in 1987. This device used a novel two-layer structure with separate hole transporting and electron transporting layers such that recombination and light emission occurred in the middle of the organic layer. This resulted in a reduction in operating voltage and improvements in efficiency and led to the current era of OLED research and device production.

Research into polymer electroluminescence culminated in 1990 with J. H. Burroughes *et al.* at the Cavendish Laboratory in Cambridge reporting a high efficiency green light-emitting polymer based device using 100 nm thick films of poly(p-phenylene vinylene).

Working principle



Schematic of a bilayer OLED: 1. Cathode (−), 2. Emissive Layer, 3. Emission of radiation, 4. Conductive Layer, 5. Anode (+)

A typical OLED is composed of a layer of organic materials situated between two electrodes, the anode and cathode, all deposited on a substrate. The organic molecules are electrically conductive as a result of delocalization of pi electrons caused by conjugation over all or part of the molecule. These materials have conductivity levels ranging from insulators to conductors, and therefore are considered organic semiconductors. The highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) of organic semiconductors are analogous to the valence and conduction bands of inorganic semiconductors.

Originally, the most basic polymer OLEDs consisted of a single organic layer. One example was the first light-emitting device synthesised by J. H. Burroughes *et al.*, which involved a single layer of poly(p-phenylene vinylene). However multilayer OLEDs can be fabricated with two or more layers in order to improve device efficiency. As well as conductive properties, different materials may be chosen to aid charge injection at

electrodes by providing a more gradual electronic profile, or block a charge from reaching the opposite electrode and being wasted. Many modern OLEDs incorporate a simple bilayer structure, consisting of a conductive layer and an emissive layer.

During operation, a voltage is applied across the OLED such that the anode is positive with respect to the cathode. A current of electrons flows through the device from cathode to anode, as electrons are injected into the LUMO of the organic layer at the cathode and withdrawn from the HOMO at the anode. This latter process may also be described as the injection of electron holes into the HOMO. Electrostatic forces bring the electrons and the holes towards each other and they recombine forming an exciton, a bound state of the electron and hole. This happens closer to the emissive layer, because in organic semiconductors holes are generally more mobile than electrons. The decay of this excited state results in a relaxation of the energy levels of the electron, accompanied by emission of radiation whose frequency is in the visible region. The frequency of this radiation depends on the band gap of the material, in this case the difference in energy between the HOMO and LUMO.

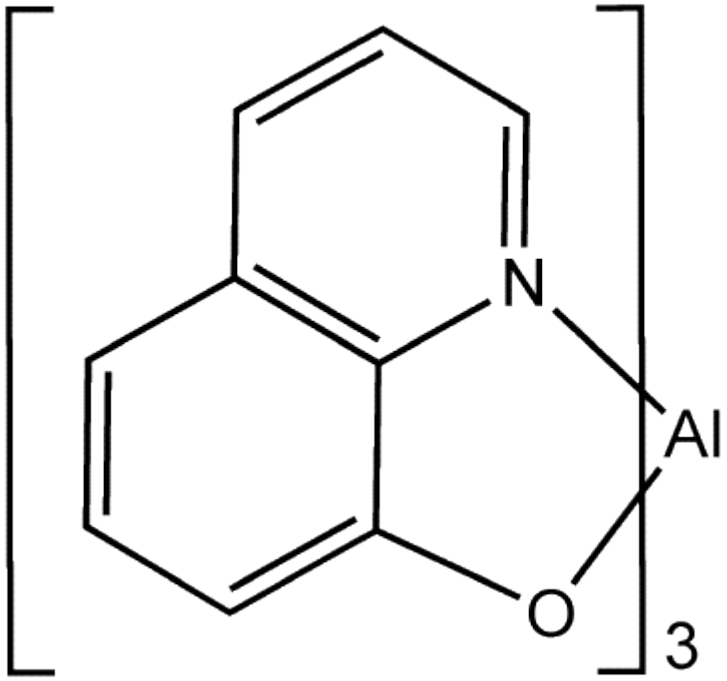
As electrons and holes are fermions with half integer spin, an exciton may either be in a singlet state or a triplet state depending on how the spins of the electron and hole have been combined. Statistically three triplet excitons will be formed for each singlet exciton. Decay from triplet states (phosphorescence) is spin forbidden, increasing the timescale of the transition and limiting the internal efficiency of fluorescent devices. Phosphorescent organic light-emitting diodes make use of spin–orbit interactions to facilitate intersystem crossing between singlet and triplet states, thus obtaining emission from both singlet and triplet states and improving the internal efficiency.

Indium tin oxide (ITO) is commonly used as the anode material. It is transparent to visible light and has a high work function which promotes injection of holes into the HOMO level of the organic layer. A typical conductive layer may consist of PEDOT:PSS as the HOMO level of this material generally lies between the workfunction of ITO and the HOMO of other commonly used polymers, reducing the energy barriers for hole injection. Metals such as barium and calcium are often used for the cathode as they have low work functions which promote injection of electrons into the LUMO of the organic layer. Such metals are reactive, so require a capping layer of aluminium to avoid degradation.

Single carrier devices are typically used to study the kinetics and charge transport mechanisms of an organic material and can be useful when trying to study energy transfer processes. As current through the device is composed of only one type of charge carrier, either electrons or holes, recombination does not occur and no light is emitted. For example, electron only devices can be obtained by replacing ITO with a lower work function metal which increases the energy barrier of hole injection. Similarly, hole only devices can be made by using a cathode comprised solely of aluminium, resulting in an energy barrier too large for efficient electron injection.

Material technologies

Small molecules



Alq_3, commonly used in small molecule OLEDs.

Efficient OLEDs using small molecules were first developed by Dr. Ching W. Tang *et al.* at Eastman Kodak. The term OLED traditionally refers specifically to this type of device, though the term SM-OLED is also in use.

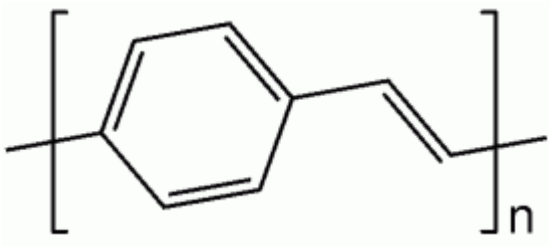
Molecules commonly used in OLEDs include organometallic chelates (for example Alq_3, used in the organic light-emitting device reported by Tang *et al.*), fluorescent and phosphorescent dyes and conjugated dendrimers. A number of materials are used for their charge transport properties, for example triphenylamine and derivatives are commonly used as materials for hole transport layers. Fluorescent dyes can be chosen to obtain light emission at different wavelengths, and compounds such as perylene, rubrene and quinacridone derivatives are often used. Alq_3 has been used as a green emitter, electron transport material and as a host for yellow and red emitting dyes.

The production of small molecule devices and displays usually involves thermal evaporation in a vacuum. This makes the production process more expensive and of limited use for large-area devices than other processing techniques. However, contrary to polymer-based devices, the vacuum deposition process enables the formation of well controlled, homogeneous films, and the construction of very complex multi-layer structures. This high flexibility in layer design, enabling distinct charge transport and

charge blocking layers to be formed, is the main reason for the high efficiencies of the small molecule OLEDs.

Coherent emission from a laser dye-doped tandem SM-OLED device, excited in the pulsed regime, has been demonstrated. The emission is nearly diffraction limited with a spectral width similar to that of broadband dye lasers.

Polymer light-emitting diodes



poly(*p*-phenylene vinylene), used in the first PLED.

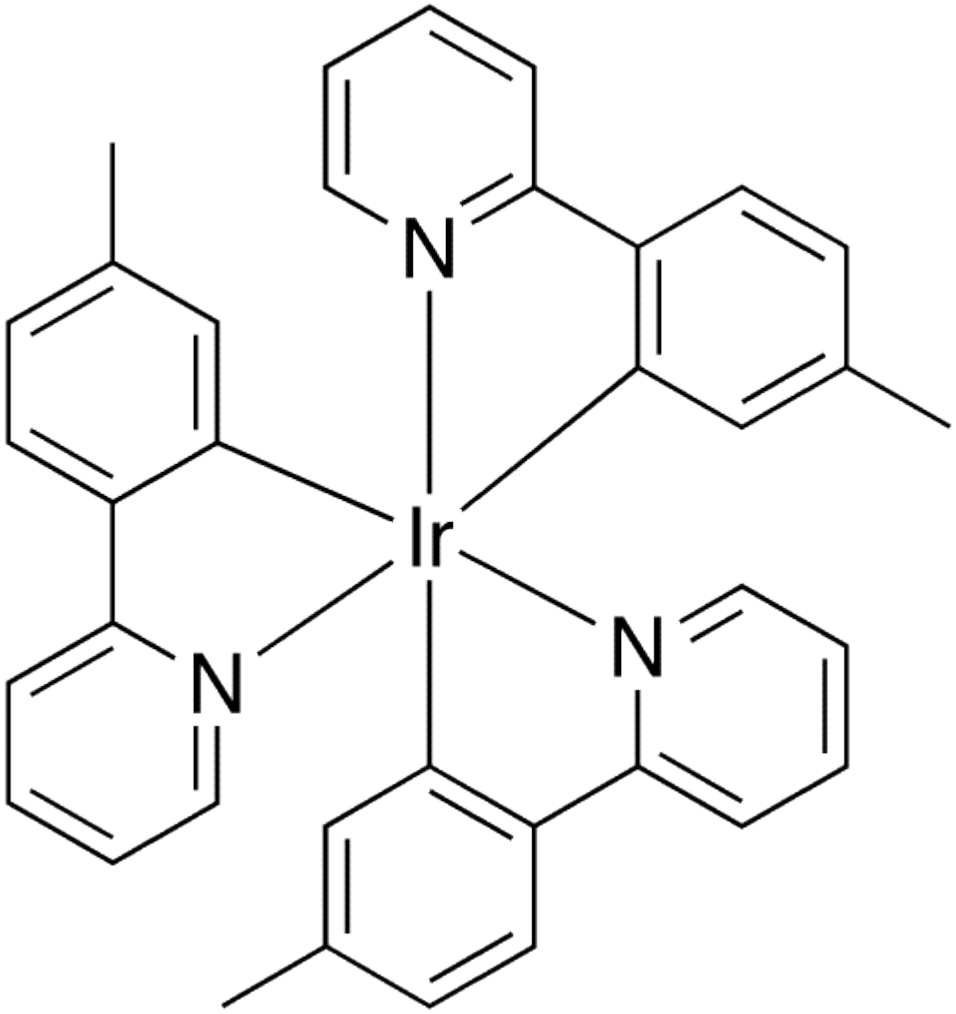
Polymer light-emitting diodes (PLED), also light-emitting polymers (LEP), involve an electroluminescent conductive polymer that emits light when connected to an external voltage. They are used as a thin film for full-spectrum colour displays. Polymer OLEDs are quite efficient and require a relatively small amount of power for the amount of light produced.

Vacuum deposition is not a suitable method for forming thin films of polymers. However, polymers can be processed in solution, and spin coating is a common method of depositing thin polymer films. This method is more suited to forming large-area films than thermal evaporation. No vacuum is required, and the emissive materials can also be applied on the substrate by a technique derived from commercial inkjet printing. However, as the application of subsequent layers tends to dissolve those already present, formation of multilayer structures is difficult with these methods. The metal cathode may still need to be deposited by thermal evaporation in vacuum.

Typical polymers used in PLED displays include derivatives of poly(*p*-phenylene vinylene) and polyfluorene. Substitution of side chains onto the polymer backbone may determine the colour of emitted light or the stability and solubility of the polymer for performance and ease of processing.

While unsubstituted poly(p-phenylene vinylene) (PPV) is typically insoluble, a number of PPVs and related poly(naphthalene vinylene)s (PNVs) that are soluble in organic solvents or water have been prepared via ring opening metathesis polymerization.

Phosphorescent materials



$Ir(mppy)_3$, a phosphorescent dopant which emits green light.

Phosphorescent organic light emitting diodes use the principle of electrophosphorescence to convert electrical energy in an OLED into light in a highly efficient manner, with the internal quantum efficiencies of such devices approaching 100%.

Typically, a polymer such as poly(n-vinylcarbazole) is used as a host material to which an organometallic complex is added as a dopant. Iridium complexes such as $Ir(mppy)_3$ are currently the focus of research, although complexes based on other heavy metals such as platinum have also been used.

The heavy metal atom at the centre of these complexes exhibits strong spin-orbit coupling, facilitating intersystem crossing between singlet and triplet states. By using these phosphorescent materials, both singlet and triplet excitons will be able to decay

radiatively, hence improving the internal quantum efficiency of the device compared to a standard PLED where only the singlet states will contribute to emission of light.

Applications of OLEDs in solid state lighting require the achievement of high brightness with good CIE coordinates (for white emission). The use of macromolecular species like polyhedral oligomeric silsesquioxanes (POSS) in conjunction with the use of phosphorescent species such as Ir for printed OLEDs have exhibited brightnesses as high as 10,000 cd/m^2.

Device Architectures

Structure

- **Bottom or top emission:** Bottom emission devices use a transparent or semi-transparent bottom electrode to get the light through a transparent substrate. Top emission devices use a transparent or semi-transparent top electrode emitting light directly. Top-emitting OLEDs are better suited for active-matrix applications as they can be more easily integrated with a non-transparent transistor backplane.

- **Transparent OLEDs** use transparent or semi-transparent contacts on both sides of the device to create displays that can be made to be both top and bottom emitting (transparent). TOLEDs can greatly improve contrast, making it much easier to view displays in bright sunlight. This technology can be used in Head-up displays, smart windows or augmented reality applications. Novaled's OLED panel presented in Finetech Japan 2010, boasts a transparency of 60-70%.

- **Stacked OLEDs** use a pixel architecture that stacks the red, green, and blue subpixels on top of one another instead of next to one another, leading to substantial increase in gamut and color depth, and greatly reducing pixel gap. Currently, other display technologies have the RGB (and RGBW) pixels mapped next to each other decreasing potential resolution.

- **Inverted OLED:** In contrast to a conventional OLED, in which the anode is placed on the substrate, an Inverted OLED uses a bottom cathode that can be connected to the drain end of an n-channel TFT especially for the low cost amorphous silicon TFT backplane useful in the manufacturing of AMOLED displays.

Patterning technologies

Patternable organic light-emitting devices use a light or heat activated electroactive layer. A latent material (PEDOT-TMA) is included in this layer that, upon activation, becomes highly efficient as a hole injection layer. Using this process, light-emitting devices with arbitrary patterns can be prepared.

Colour patterning can be accomplished by means of laser, such as radiation-induced sublimation transfer (RIST).

Organic vapour jet printing (OVJP) uses an inert carrier gas, such as argon or nitrogen, to transport evaporated organic molecules (as in Organic Vapor Phase Deposition). The gas is expelled through a micron sized nozzle or nozzle array close to the substrate as it is being translated. This allows printing arbitrary multilayer patterns without the use of solvents.

Conventional OLED displays are formed by vapor thermal evaporation (VTE) and are patterned by shadow-mask. A mechanical mask has openings allowing the vapor to pass only on the desired location.

Backplane technologies

For a high resolution display like a TV, a TFT backplane is necessary to drive the pixels correctly. Currently, Low Temperature Polycrystalline silicon LTPS-TFT is used for commercial AMOLED displays. LTPS-TFT has variation of the performance in a display, so various compensation circuits have been reported. Due to the size limitation of the excimer laser used for LTPS, the AMOLED size was limited. To cope with the hurdle related to the panel size, amorphous-silicon/microcrystalline-silicon backplanes have been reported with large display prototype demonstrations.

Advantages

Demonstration of a 4.1" prototype flexible display from Sony

The different manufacturing process of OLEDs lends itself to several advantages over flat-panel displays made with LCD technology.

- **Lower cost in the future:** OLEDs can be printed onto any suitable substrate by an inkjet printer or even by screen printing,theoretically making them cheaper to produce than LCD or plasma displays. However, fabrication of the OLED substrate is more costly than that of a TFT LCD, until mass production methods lower cost through scaleability.

- **Light weight & flexible plastic substrates:** OLED displays can be fabricated on flexible plastic substrates leading to the possibility of Organic light-emitting diode roll-up display being fabricated or other new applications such as roll-up displays embedded in fabrics or clothing. As the substrate used can be flexible such as PET., the displays may be produced inexpensively.

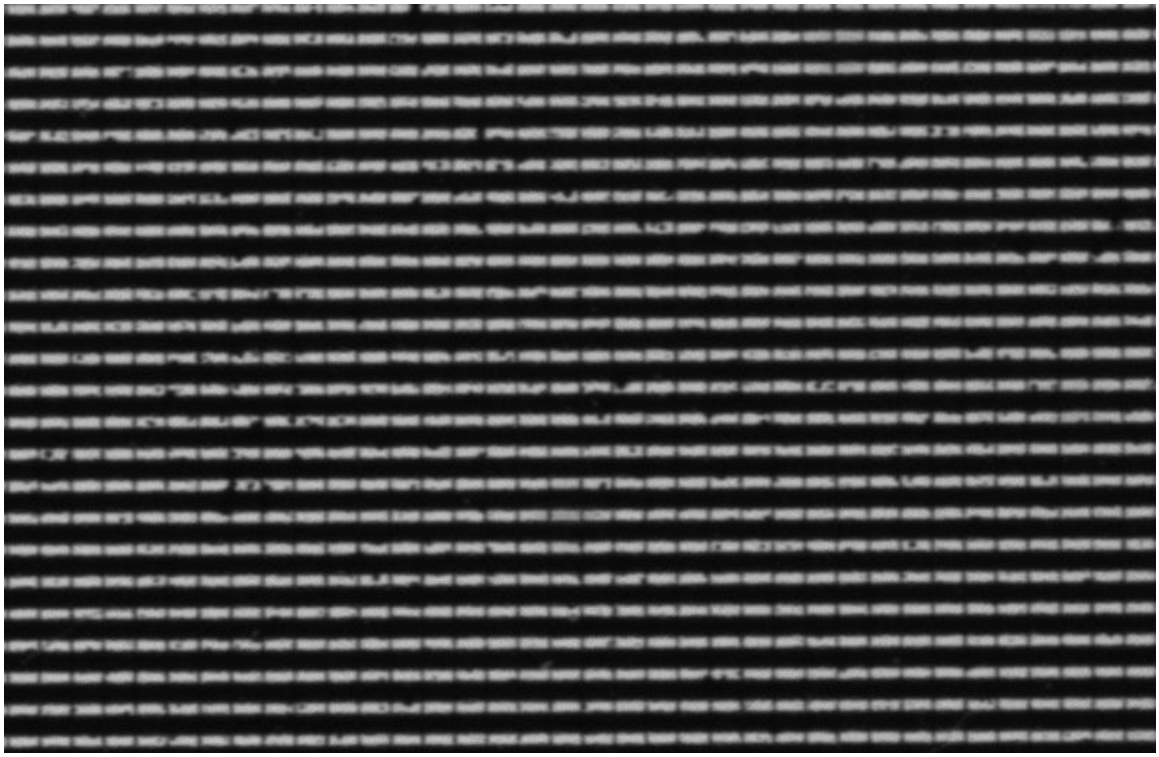
- **Wider viewing angles & improved brightness:** OLEDs can enable a greater artificial contrast ratio (both dynamic range and static, measured in purely dark conditions) and viewing angle compared to LCDs because OLED pixels directly emit light. OLED pixel colours appear correct and unshifted, even as the viewing angle approaches 90 degrees from normal.

- **Better power efficiency:** LCDs filter the light emitted from a backlight, allowing a small fraction of light through so they cannot show true black, while an inactive OLED element does not produce light or consume power.

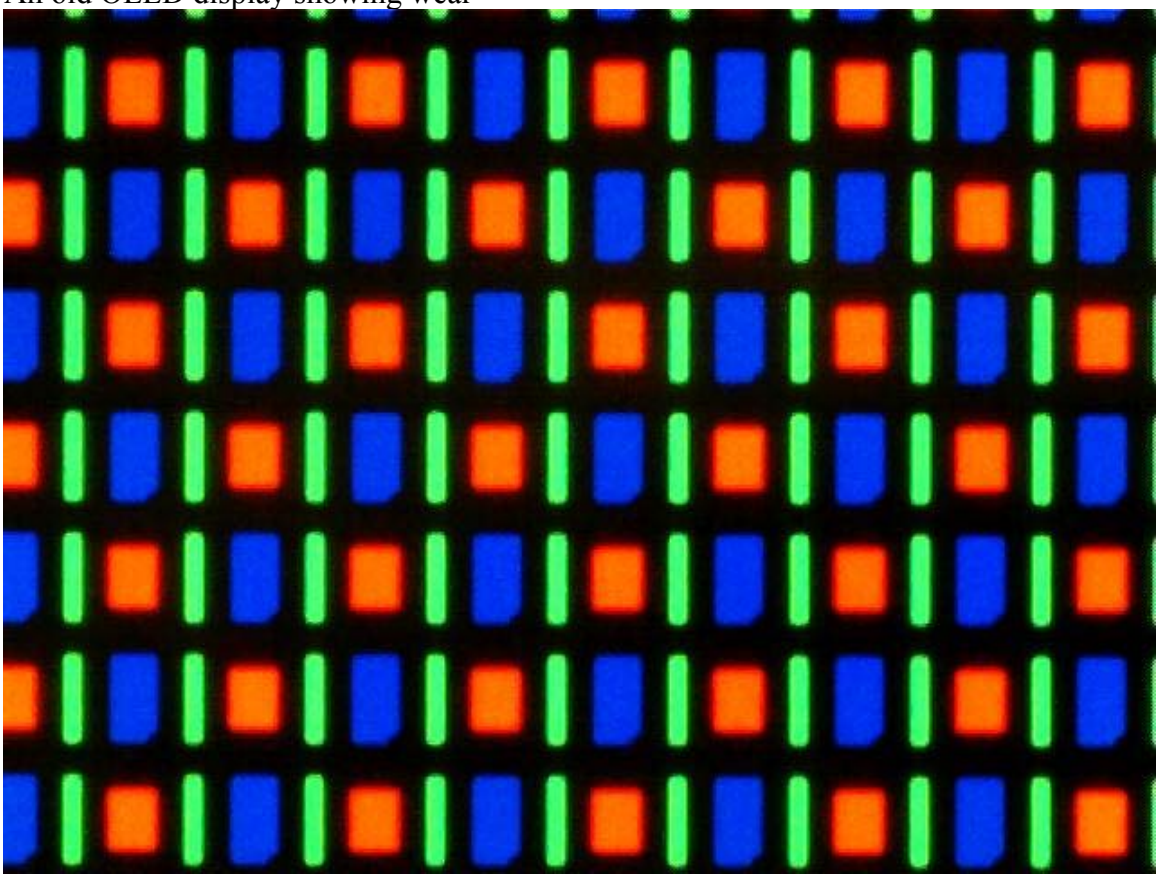
- **Response time:** OLEDs can also have a faster response time than standard LCD screens. Whereas LCD displays are capable of a 1 ms response time or less offering a frame rate of 1,000 Hz or higher, an OLED can theoretically have less than 0.01 ms response time enabling 100,000 Hz refresh rates.

Disadvantages

LEP display showing partial failure



An old OLED display showing wear



Magnified image of the AMOLED screen on the Google Nexus One smartphone using the RGBG system of the PenTile Matrix Family.

- **Lifespan:** The biggest technical problem for OLEDs was the limited lifetime of the organic materials. In particular, blue OLEDs historically have had a lifetime of around 14,000 hours to half original brightness (five years at 8 hours a day) when used for flat-panel displays. This is lower than the typical lifetime of LCD, LED or PDP technology—each currently rated for about 60,000 hours to half brightness, depending on manufacturer and model. However, some manufacturers' displays aim to increase the lifespan of OLED displays, pushing their expected life past that of LCD displays by improving light outcoupling, thus achieving the same brightness at a lower drive current. In 2007, experimental OLEDs were created which can sustain 400 cd/m^2 of luminance for over 198,000 hours for green OLEDs and 62,000 hours for blue OLEDs.

- **Color balance issues:** Additionally, as the OLED material used to produce blue light degrades significantly more rapidly than the materials that produce other colors, blue light output will decrease relative to the other colors of light. This differential color output change will change the color balance of the display and is much more noticeable than a decrease in overall luminance. This can be partially avoided by adjusting colour balance but this may require advanced control circuits and interaction with the user, which is unacceptable for some users. In order to delay the problem, manufacturers bias the colour balance towards blue so that the display initially has an artificially blue tint, leading to complaints of artificial-looking, over-saturated colors. More commonly, though, manufacturers optimize the size of the R, G and B subpixels to reduce the current density through the subpixel in order to equalize lifetime at full luminance. For example, a blue subpixel may be 100% larger than the green subpixel. The red subpixel may be 10% smaller than the green.

- **Efficiency of blue OLEDs:** Improvements to the efficiency and lifetime of blue OLED's is vital to the success of OLED's as replacements for LCD technology. Considerable research has been invested in developing blue OLEDs with high external quantum efficiency as well as a deeper blue color. External quantum efficiency values of 20% and 19% have been reported for red (625 nm) and green (530 nm) diodes, respectively. However, blue diodes (430 nm) have only been able to achieve maximum external quantum efficiencies in the range between 4% to 6%. By calculating the band gap ($E_g = hc/\lambda$), it is clear that the shorter wavelength of the blue OLED results in a larger band gap at 2.9 eV. This leads to higher barriers, causing lower efficiency.

- **Water damage:** Water can damage the organic materials of the displays. Therefore, improved sealing processes are important for practical manufacturing. Water damage may especially limit the longevity of more flexible displays.

- **Outdoor performance:** As an emissive display technology, OLEDs rely completely upon converting electricity to light, unlike most LCDs which are to some extent reflective; e-ink leads the way in efficiency with ~ 33% ambient light reflectivity, enabling the display to be used without any internal light source. The

metallic cathode in an OLED acts as a mirror, with reflectance approaching 80%, leading to poor readability in bright ambient light such as outdoors. However, with the proper application of a circular polarizer and anti-reflective coatings, the diffuse reflectance can be reduced to less than 0.1%. With 10,000 fc incident illumination (typical test condition for simulating outdoor illumination), that yields an approximate photopic contrast of 5:1.

- **Power consumption:** While an OLED will consume around 40% of the power of an LCD displaying an image which is primarily black, for the majority of images it will consume 60–80% of the power of an LCD - however it can use over three times as much power to display an image with a white background such as a document or website. This can lead to disappointing real-world battery life in mobile devices. Websites and tools exist to help disappointed mobile phone users display a black background in places where it would normally be white. For example, the mobile Google search engine can now be used in black to save battery power by using a custom version called Black Google Mobile.

- **Screen burn-in:** Unlike displays with a common light source, the brightness of each OLED pixel fades depending on the content displayed. The varied lifespan of the organic dyes can cause a discrepancy between red, green, and blue intensity. This leads to image persistence, also known as burn-in.

Manufacturers and Commercial Uses

A 3.8 cm (1.5 in) OLED display from a Creative ZEN V media player

OLED technology is used in commercial applications such as displays for mobile phones and portable digital media players, car radios and digital cameras among others. Such portable applications favor the high light output of OLEDs for readability in sunlight and their low power drain. Portable displays are also used intermittently, so the lower lifespan of organic displays is less of an issue. Prototypes have been made of flexible and rollable displays which use OLEDs' unique characteristics. Applications in flexible signs and lighting are also being developed. Philips Lighting have made OLED lighting samples under the brand name 'Lumiblade' available online.

OLEDs have been used in most Motorola and Samsung colour cell phones, as well as some HTC, LG and Sony Ericsson models. Nokia has also recently introduced some OLED products including the N85 and the N86 8MP, both of which feature an AMOLED display. OLED technology can also be found in digital media players such as the Creative ZEN V, the iriver clix, the Zune HD and the Sony Walkman X Series.

The Google and HTC Nexus One smartphone includes an AMOLED screen, as does HTC's own Desire and Legend phones. However due to supply shortages of the

Samsung-produced displays, certain HTC models will use Sony's SLCD displays in the future.

The Google and Samsung Nexus S smartphone includes a Super AMOLED screen, as does Samsung's own Galaxy S phone. However due to supply shortages of the Samsung-produced displays certain countries, such as Russia, will have Nexus S models that use "Super Clear LCD" instead; the same display used by Samsung for its new Wave II S8530.

Other manufacturers of OLED panels include Anwell Technologies Limited, Chi Mei Corporation, LG, and others.

DuPont stated in a press release in May 2010 that they can produce a 50-inch OLED TV in two minutes with a new printing technology. If this can be scaled up in terms of manufacturing, then the total cost of OLED TVs would be greatly reduced. Dupont also states that OLED TVs made with this less expensive technology can last up to 15 years if left on for a normal eight hour day.

Handheld computer manufacturer OQO introduced the smallest Windows netbook computer, including an OLED display, in 2009.

The use of OLEDs may be subject to patents held by Eastman Kodak, DuPont, General Electric, Royal Philips Electronics, numerous universities and others. There are by now literally thousands of patents associated with OLEDs, both from larger corporations and smaller technology companies.

Samsung applications

By 2004 Samsung, South Korea's largest conglomerate, was the world's largest OLED manufacturer, producing 40% of the OLED displays made in the world, and as of 2010 has a 98% share of the global AMOLED market. The company is leading the world OLED industry, generating $100.2 million out of the total $475 million revenues in the global OLED market in 2006. As of 2006, it held more than 600 American patents and more than 2800 international patents, making it the largest owner of AMOLED technology patents.

Samsung SDI announced in 2005 the world's largest OLED TV at the time, at 21 inches (53 cm). This OLED featured the highest resolution at the time, of 6.22 million pixels. In addition, the company adopted active matrix based technology for its low power consumption and high-resolution qualities. This was exceeded in January 2008, when Samsung showcased the world's largest and thinnest OLED TV at the time, at 31 inches and 4.3 mm.

In May 2008, Samsung unveiled an ultra-thin 12.1 inch laptop OLED display concept, with a 1,280×768 resolution with infinite contrast ratio. According to Woo Jong Lee,

Vice President of the Mobile Display Marketing Team at Samsung SDI, the company expected OLED displays to be used in notebook PCs as soon as 2010.

In October 2008, Samsung showcased the world's thinnest OLED display, also the first to be 'flappable' and bendable. It measures just 0.05 mm (thinner than paper), yet a Samsung staff member said that it is "technically possible to make the panel thinner". To achieve this thickness, Samsung etched an OLED panel that uses a normal glass substrate. The drive circuit was formed by low-temperature polysilicon TFTs. Also, low-molecular organic EL materials were employed. The pixel count of the display is 480 × 272. The contrast ratio is 100,000:1, and the luminance is 200 cd/m². The colour reproduction range is 100% of the NTSC standard.

In the same month, Samsung unveiled what was then the world's largest OLED Television at 40-inch with a Full HD resolution of 1920×1080 pixel. In the FPD International, Samsung stated that its 40-inch OLED Panel is the largest size currently possible. The panel has a contrast ratio of 1,000,000:1, a colour gamut of 107% NTSC, and a luminance of 200 cd/m² (peak luminance of 600 cd/m²).

At the Consumer Electronics Show (CES) in January 2010, Samsung demonstrated a laptop computer with a large, transparent OLED display featuring up to 40% transparency and an animated OLED display in a photo ID card.

Samsung's latest AMOLED smartphones use their Super AMOLED trademark, with the Samsung Wave S8500 and Samsung i9000 Galaxy S being launched in June 2010.

Sony applications

Sony XEL-1, the world's first OLED TV. (front)


Sony XEL-1 (side)

The Sony CLIÉ PEG-VZ90 was released in 2004, being the first PDA to feature an OLED screen. Other Sony products to feature OLED screens include the MZ-RH1 portable minidisc recorder, released in 2006 and the Walkman X Series.

At the Las Vegas CES 2007, Sony showcased 11-inch (28 cm, resolution 960×540) and 27-inch (68.5 cm, full HD resolution at 1920×1080) OLED TV models. Both claimed 1,000,000:1 contrast ratios and total thicknesses (including bezels) of 5 mm. In April 2007, Sony announced it would manufacture 1000 11-inch OLED TVs per month for market testing purposes. On October 1, 2007, Sony announced that the 11-inch model, now called the XEL-1, would be released commercially; the XEL-1 was first released in Japan in December 2007.

In May 2007, Sony publicly unveiled a video of a 2.5-inch flexible OLED screen which is only 0.3 millimeters thick. At the Display 2008 exhibition, Sony demonstrated a 0.2 mm thick 3.5 inch display with a resolution of 320×200 pixels and a 0.3 mm thick 11 inch display with 960×540 pixels resolution, one-tenth the thickness of the XEL-1.

In July 2008, a Japanese government body said it would fund a joint project of leading firms, which is to develop a key technology to produce large, energy-saving organic displays. The project involves one laboratory and 10 companies including Sony Corp.

NEDO said the project was aimed at developing a core technology to mass-produce 40 inch or larger OLED displays in the late 2010s.

In October 2008, Sony published results of research it carried out with the Max Planck Institute over the possibility of mass-market bending displays, which could replace rigid LCDs and plasma screens. Eventually, bendable, transparent OLED screens could be stacked to produce 3D images with much greater contrast ratios and viewing angles than existing products.

Sony exhibited a 24.5" prototype OLED 3D television during the Consumer Electronics Show in January 2010.

LG applications

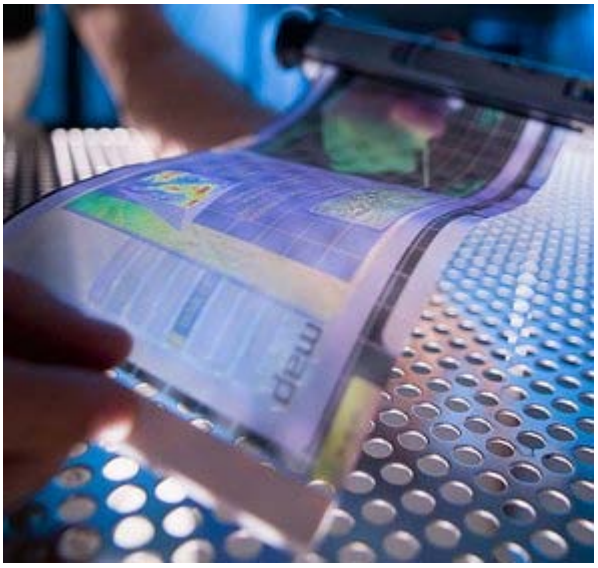
Since 2010 LG sells one 15" OLED TV, the 15EL9500 and has announced a 31" OLED 3D TV for March 2011.

Range of OLED screens sizes in production products

- Media Player - 2.2" to 3.3" - various
- Mobile Phone - 2.6" to 4.1" - various
- Tablet - none
- Laptop - none
- PC monitor - none
- TV - 11" to 15" - Sony XEL-1 and LG 15EL9500

Chapter- 7

Organic Field-Effect Transistor



OFET-based flexible display

An **organic field-effect transistor (OFET)** is a field effect transistor using an organic semiconductor in its channel. OFETs can be prepared either by vacuum evaporation of small molecules, by solution-casting of polymers or small molecules, or by mechanical transfer of a peeled single-crystalline organic layer onto a substrate. These devices have been developed to realize low-cost, large-area electronic products. OFETs have been fabricated with various device geometries. The most commonly used device geometry is bottom gate with top drain- and source electrodes, because this geometry is similar to the thin-film silicon transistor (TFT) using thermally grown Si/SiO_2 oxide as gate dielectric. Organic polymers, such as poly(methyl-methacrylate) (PMMA), can be used as dielectric, too.

In May 2007, Sony reported the first full-color, video-rate, flexible, all plastic display, in which both the thin film transistors and the light emitting pixels were made of organic materials.

History of OFETs

The field-effect transistor (FET) was first proposed by J.E. Lilienfeld, who received a patent for his idea in 1930. He proposed that a field-effect transistor behaves as a capacitor with a conducting channel between a source and a drain electrode. Applied voltage on the gate electrode controls the amount of charge carriers flowing through the system.

The first field-effect transistor was designed and prepared in 1960 by Kahng and Atalla using a metal-oxide-semiconductor. However, rising costs of materials and manufacturing, as well as public interest in more environmentally friendly electronics materials have supported development of organic based electronics in more recent years. In 1987, Koezuka and co-workers reported the first organic field-effect transistor based on a polymer of thiophene molecules. The thiophene polymer is a type of conjugated polymer that is able to conduct charge, eliminating the need to use expensive metal oxide semiconductors. Additionally, other conjugated polymers have been shown to have semi-conducting properties. OFET design has also improved in the past few decades. Many OFETs are now designed based on the thin-film transistor (TFT) model, which allows the devices to use less conductive materials in their design. Improvement on these models in the past few years have been made to field-effect mobility and on-off current ratios.

Materials

One common feature of OFET materials is the inclusion of an aromatic or otherwise conjugated π-electron system, facilitating the delocalization of orbital wavefunctions. Electron withdrawing groups or donating groups can be attached that facilitate hole or electron transport.

OFETs employing many aromatic and conjugated materials as the active semiconducting layer have been reported, including small molecules such as rubrene, tetracene, pentacene, diindenoperylene, perylenediimides, tetracyanoquinodimethane (TCNQ), and polymers such as polythiophenes (especially poly 3-hexylthiophene (P3HT)), polyfluorene, polydiacetylene, poly 2,5-thienylene vinylene, poly p-phenylene vinylene (PPV).

The field is very active, with newly synthesized and tested compounds reported weekly in prominent research journals. Many review articles exist documenting the development of these materials.

Rubrene-based OFETs show the highest carrier mobility 20–40 $cm^2/(V \cdot s)$. Another popular OFET material is pentacene, which has been used since 1980s, but resulted in about 10 times lower mobilities than rubrene. The major problem with pentacene, as well as many other organic conductors, is its rapid oxidation in air to form pentacene-quinone. However if the pentacene is preoxidized, and the thus formed pentacene-quinone is used

as the gate insulator, then the mobility can approach the rubrene values. This pentacene oxidation technique is akin to the silicon oxidation used in the silicon electronics.

Polycrystalline tetrathiafulvalene and its analogues result in mobilities in the range 0.1–1.4 $cm^2/(V \cdot s)$. However, the mobility exceeds 10 $cm^2/(V \cdot s)$ in solution-grown or vapor-transport-grown single crystalline hexamethylene-tetrathiafulvalene (HMTTF). The ON/OFF voltage is different for devices grown by those two techniques, presumably due to the higher processing temperatures using in the vapor transport grows.

All the above-mentioned devices are based on p-type conductivity. N-type organic OFETs are yet poorly developed. They are usually based on perylenediimides or fullerenes or their derivatives, and show electron mobilities below 2 $cm^2/(V \cdot s)$.

Device design of organic field-effect transistors

Three essential components of field-effect transistors are the source, the drain and the gate. Field-effect transistors usually operate as a capacitor. They are composed of two plates. One plate works as a conducting channel between two ohmic contacts, which are called the source and the drain contacts. The other plate works to control the charge induced into the channel, and it is called the gate. The direction of the movement of the carriers in the channel is from the source to the drain. Hence the relationship between these three components is that the gate controls the carrier movement from the source to the drain.

When this capacitor concept is applied to the device design, various devices can be built up based on the difference in the controller-the gate. This can be the gate material, the location of the gate with respect to the channel, how the gate is isolated from the channel, and what type of carrier is induced by the gate voltage into channel (such as electrons in an n-channel device, holes in a p-channel device, and both electrons and holes in a double injection device).

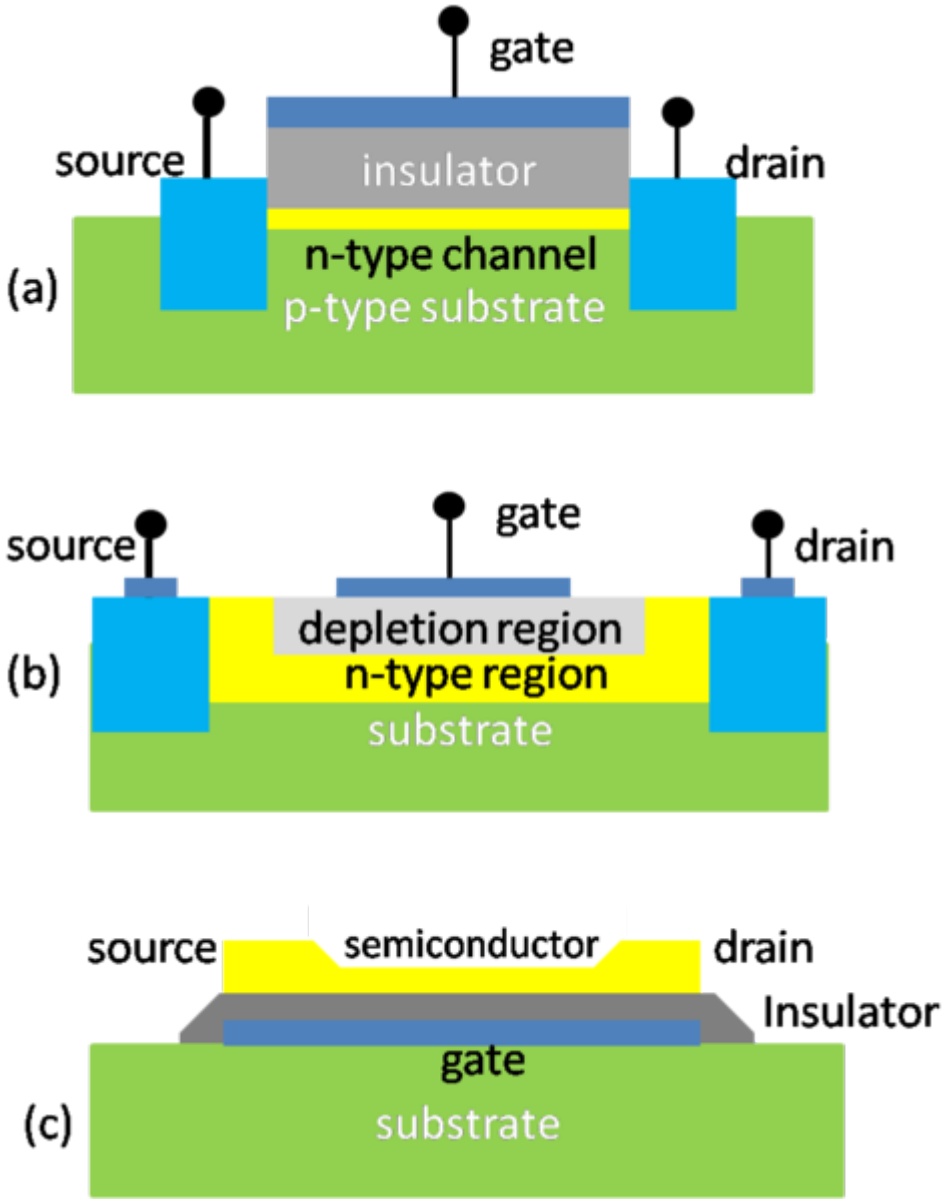


Figure 1 Schematic graphs of three different kinds of Field-Effect Transistor (FET): (a) the metal- insulator -semiconductor FET(MISFET); (b) the metal-semiconductor FET (MESFET); (c) the thin-film transistor (TFT).

Classified by the properties of the carrier, three types of FETs are shown schematically in Figure 1. They are MOSFET (Metal Oxide Semiconductor Field-Effect Transistor), MESFET (MEtal Semiconductor Field-Effect Transistor) and TFT (Thin Film Transistor).

MISFET

The most prominent and widely used FET in modern microelectronics is the MOSFET. There are different kinds in this category, such as MISFET (Metal Insulator

Semiconductor Field-Effect Transistor), and IGFET (Insulator Gate Field-Effect Transistor). The scheme of a MISFET is shown in Figure 1a. The source and the drain are connected by a semiconductor and the gate is separated from the channel by a layer of insulator. If there is no bias (potential difference) applied on the gate, the band bending is induced due to the energy difference of metal conducting band and the semiconductor Fermi-level. Therefore a higher concentration of holes is formed on the interface of the semiconductor and the insulator. When an enough positive bias is applied on the gate contact, the bended band becomes flat. If a larger positive bias is applied, the band bending in the opposite direction occurs and the region close to the insulator-semiconductor interface becomes depleted of holes. Then the depleted region is formed. At an even larger positive bias, the band bending becomes so large that the Fermi-level at the interface of the semiconductor and the insulator becomes closer to the bottom of the conduction band than to the top of the valence band, therefore, it forms an inversion layer of electrons, providing the conducting channel. Finally, it turns the device on.

MESFET

The second type of device is described in Fig.1b. The only difference of this one from the MISFET is that the n-type source and drain are connected by an n-type region. In this case, the depletion region extends all over the n-type channel at zero gate voltage in a normally “off” device (it is similar to the larger positive bias in MISFET case). In the normally “on” device, a portion of the channel is not depleted, and thus leads to passage of a current at zero gate voltage.

TFT

The concept of TFT was first proposed by Paul Weimer in 1962. This is illustrated in Fig. 1c. Here the source and drain electrodes are directly deposited onto the conducting channel (a thin layer of semiconductor) then a thin film of insulator is deposited between the semiconductor and the metal gate contact. This structure suggests that there is no depletion region to separate the device from the substrate. If there is zero bias, the electrons are expelled from the surface due to the Fermi-level energy difference of the semiconductor and the metal. This leads to band bending of semiconductor. In this case, there is no carrier movement between the source and drain. When the positive charge is applied, the accumulation of electrons on the interface leads to the bending of the semiconductor in an opposite way and leads to the lowering of the conduction band with regards to the Fermi-level of the semiconductor. Then a highly conductive channel forms at the interface (shown in Figure 2).

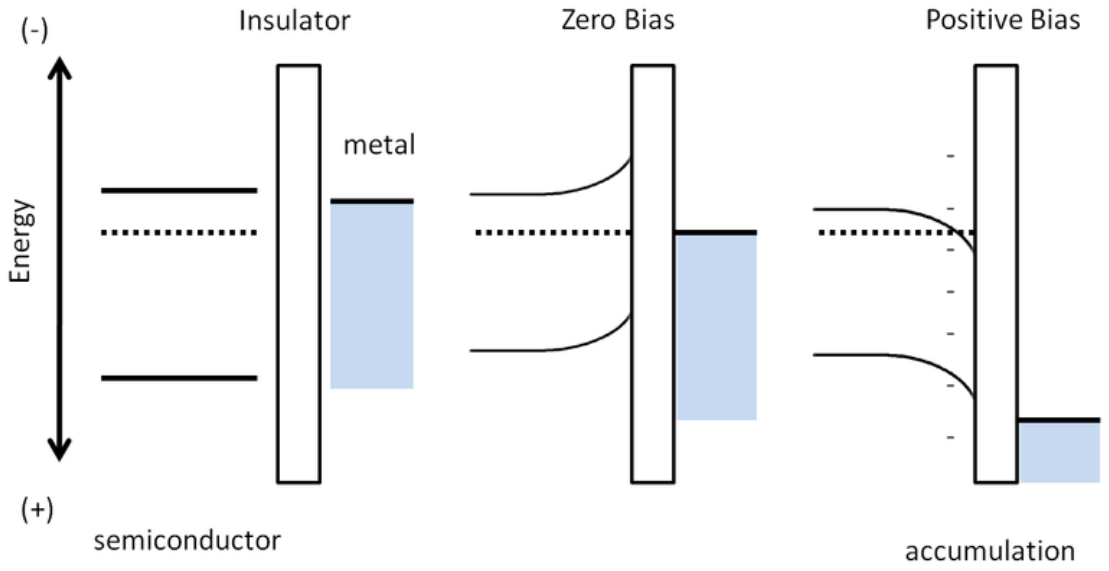
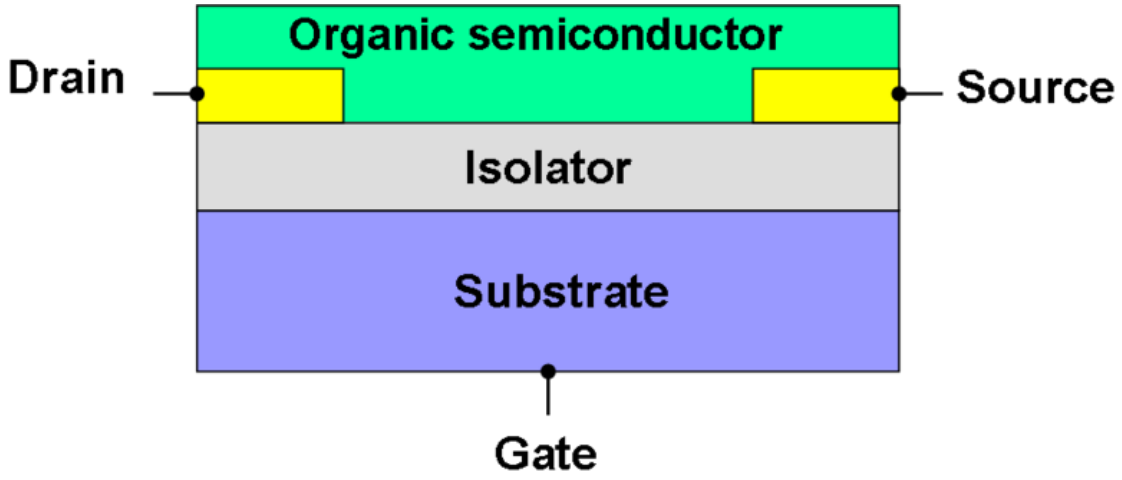


Figure 2: Schematic graphs of band-bending in the TFT device model.

OFET

OFETs adopt the architecture of TFT. With the development of the conducting polymer, the semiconducting properties of small conjugated molecules have been recognized. The interest in OFETs has grown enormously in the past ten years. The reasons for this surge of interest are manifold. The performance of OFETs, which can compete with that of amorphous silicon (a-Si) TFTs with field-effect mobilities of 0.5^{-1} cm^2 V^{-1} s^{-1} and ON/OFF current ratios (which indicate the ability of the device to shut down) of 106–108, has improved significantly. Currently, thin-film OFET mobility values of 5 $cm^2V^{-1}s^{-1}$ in the case of vacuum-deposited small molecules and 0.6 cm^2 V^{-1} s^{-1} for solution-processed polymers have been reported. As a result, there is now a greater industrial interest in using OFETs for applications that are currently incompatible with the use of a-Si or other inorganic transistor technologies. One of their main technological attractions is that all the layers of an OFET can be deposited and patterned at room temperature by a combination of low-cost solution-processing and direct-write printing, which makes them ideally suited for realization of low-cost, large-area electronic functions on flexible substrates.

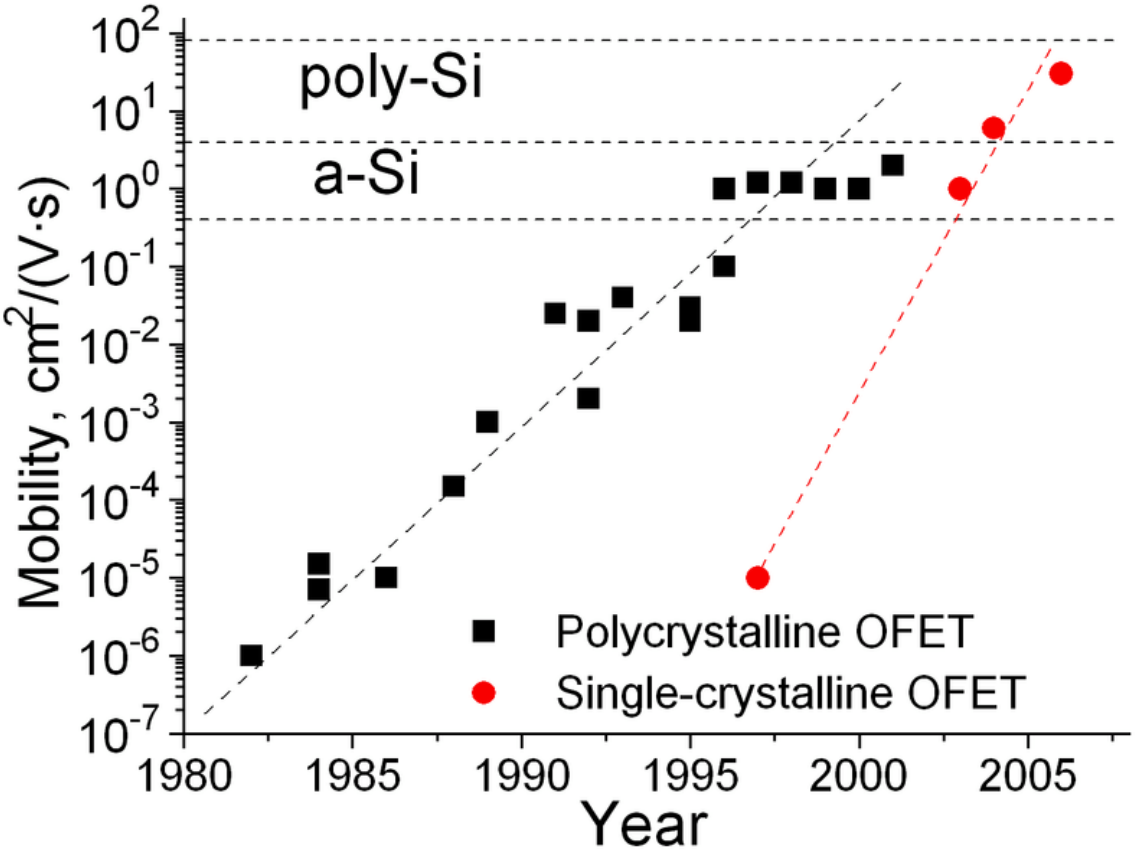
Device preparation



OFET schematic

Thermally oxidized silicon is a traditional substrate for OFETs where the silicon dioxide serves as the gate insulator. The active FET layer is usually deposited onto this substrate using either (i) thermal evaporation, (ii) coating from organic solution, or (iii) electrostatic lamination. The first two techniques result in polycrystalline active layers; they are much easier to produce, but result in relatively poor transistor performance. Numerous variations of the solution coating technique (ii) are known, including dip-coating, spin-coating, inkjet printing and screen printing. The electrostatic lamination technique is based on manual peeling of a thin layer off a single organic crystal; it results in a superior single-crystalline active layer, yet it is more tedious. The thickness of the gate oxide and the active layer is below one micrometer.

Carrier transport



Evolution of carrier mobility in organic field-effect transistor.

The carrier transport in OFET is specific for two-dimensional (2D) carrier propagation through the device. Various experimental techniques were used for this study, such as Haynes - Shockley experiment on the transit times of injected carriers, time-of-flight (TOF) experiment for the determination of carrier mobility, pressure-wave propagation experiment for probing electric-field distribution in insulators, organic monolayer experiment for probing orientational dipolar changes, optical time-resolved second harmonic generation (TRM-SHG), etc. Whereas carriers propagate through polycrystalline OFETs in a diffusion-like (trap-limited) manner, they move through the conduction band in the best single-crystalline OFETs.

The most important parameter of OFET carrier transport is carrier mobility. Its evolution over the years of OFET research is shown in the graph for polycrystalline and single crystalline OFETs. The horizontal lines indicate the comparison guides to the main OFET competitors – amorphous (a-Si) and polycrystalline silicon. The graph reveals that the mobility in polycrystalline OFETs is comparable to that of a-Si whereas mobility in rubrene-based OFETs (20–40 $cm^2/(V \cdot s)$) approaches that of best poly-silicon devices.

Light-emitting OFETs

Because an electric current flows through such a transistor, it can be used as a light-emitting device, thus integrating current modulation and light emission. In 2003, a German group reported the first organic light-emitting field-effect transistor (OLET). The device structure comprises interdigitated gold source- and drain electrodes and a polycrystalline tetracene thin film. Both, positive charges (holes) as well as negative charges (electrons) are injected from the gold contacts into this layer leading to electroluminescence from the tetracene.

Chapter- 8

Introduction to Microelectromechanical Systems

Microelectromechanical systems (**MEMS**) (also written as *micro-electro-mechanical*, *MicroElectroMechanical* or *microelectronic and microelectromechanical systems*) is the technology of very small mechanical devices driven by electricity; it merges at the nano-scale into nanoelectromechanical systems (NEMS) and nanotechnology. MEMS are also referred to as micromachines (in Japan), or *Micro Systems Technology - MST* (in Europe).

MEMS are separate and distinct from the hypothetical vision of molecular nanotechnology or molecular electronics. MEMS are made up of components between 1 to 100 micrometres in size (i.e. 0.001 to 0.1 mm) and MEMS devices generally range in size from 20 micrometres (20 millionths of a metre) to a millimetre. They usually consist of a central unit that processes data, the microprocessor and several components that interact with the outside such as microsensors. At these size scales, the standard constructs of classical physics are not always useful. Because of the large surface area to volume ratio of MEMS, surface effects such as electrostatics and wetting dominate volume effects such as inertia or thermal mass.

The potential of very small machines was appreciated before the technology existed that could make them—see, for example, Richard Feynman's famous 1959 lecture There's Plenty of Room at the Bottom. MEMS became practical once they could be fabricated using modified semiconductor device fabrication technologies, normally used to make electronics. These include molding and plating, wet etching (KOH, TMAH) and dry etching (RIE and DRIE), electro discharge machining (EDM), and other technologies capable of manufacturing small devices. An early example of a MEMS device is the resonistor – an electromechanical monolithic resonator.

MEMS description

MEMS technology can be implemented using a number of different materials and manufacturing techniques, depending on target device and market sector.

Materials for MEMS manufacturing

Silicon

Silicon is the material used to create most integrated circuits used in consumer electronics in the modern world. The economies of scale, ready availability of cheap high-quality materials and ability to incorporate electronic functionality make silicon attractive for a wide variety of MEMS applications. Silicon also has significant advantages engendered through its material properties. In single crystal form, silicon is an almost perfect Hookean material, meaning that when it is flexed there is virtually no hysteresis and hence almost no energy dissipation. As well as making for highly repeatable motion, this also makes silicon very reliable as it suffers very little fatigue and can have service lifetimes in the range of billions to trillions of cycles without breaking. The basic techniques for producing all silicon based MEMS devices are deposition of material layers, patterning of these layers by photolithography and then etching to produce the required shapes.

Polymers

Even though the electronics industry provides an economy of scale for the silicon industry, crystalline silicon is still a complex and relatively expensive material to produce. Polymers on the other hand can be produced in huge volumes, with a great variety of material characteristics. MEMS devices can be made from polymers by processes such as injection molding, embossing or stereolithography and are especially well suited to microfluidic applications such as disposable blood testing cartridges.

Metals

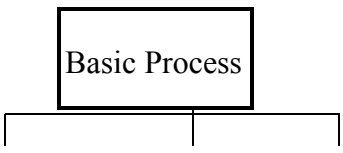
Metals can also be used to create MEMS elements. While metals do not have some of the advantages displayed by silicon in terms of mechanical properties, when used within their limitations, metals can exhibit very high degrees of reliability.

Metals can be deposited by electroplating, evaporation, and sputtering processes.

Commonly used metals include gold, nickel, aluminium, copper, chromium, titanium, tungsten, platinum, and silver.

MEMS basic processes

This chart is not complete:



Deposition	Patterning	Etching

Deposition processes

One of the basic building blocks in MEMS processing is the ability to deposit thin films of material with a thickness anywhere between a few nanometres to about 100 micrometres.

Physical deposition

There is a type of physical deposition.

Physical vapor deposition (PVD)

Sputtering

Evaporation

Chemical deposition

There are 2 types of chemical deposition.

Chemical vapor deposition

LPCVD : Low Pressure CVD PECVD : Plasma Enhanced CVD

Thermal oxidation

Patterning

Patterning in MEMS is the transfer of a pattern into a material.

Lithography

Lithography in MEMS context is typically the transfer of a pattern into a photosensitive material by selective exposure to a radiation source such as light. A photosensitive material is a material that experiences a change in its physical properties when exposed to a radiation source. If a photosensitive material is selectively exposed to radiation (e.g. by masking some of the radiation) the pattern of the radiation on the material is transferred to the material exposed, as the properties of the exposed and unexposed regions differs.

This exposed region can then be removed or treated providing a mask for the underlying substrate. Photolithography is typically used with metal or other thin film deposition, wet and dry etching.

Photolithography

KrF ArF Immersion EUV

Electron beam lithography

Ion beam lithography

X-ray lithography

Diamond patterning

Etching processes

There are two basic categories of etching processes: wet etching and dry etching. In the former, the material is dissolved when immersed in a chemical solution. In the latter, the material is sputtered or dissolved using reactive ions or a vapor phase etchant. for a somewhat dated overview of MEMS etching technologies.

Wet etching

Wet chemical etching consists in selective removal of material by dipping a substrate into a solution that dissolves it. The chemical nature of this etching process provides a good selectivity, which means the etching rate of the target material is considerably higher than the mask material if selected carefully.

Isotropic etching

Etching progresses at the same speed in all directions. Long and narrow holes in a mask will produce v-shaped grooves in the silicon. The surface of these grooves can be atomically smooth if the etch is carried out correctly, with dimensions and angles being extremely accurate.

Anisotropic etching

Some single crystal materials, such as silicon, will have different etching rates depending on the crystallographic orientation of the substrate. This is known as anisotropic etching and one of the most common examples is the etching of silicon in KOH (potassium hydroxide), where Si <111> planes etch approximately 100 times slower than other planes (crystallographic orientations). Therefore, etching a rectangular hole in a (100)-Si wafer results in a pyramid shaped etch pit with 54.7° walls, instead of a hole with curved sidewalls as with isotropic etching.

HF etching

Hydrofluoric acid is commonly used as an aqueous etchant for silicon dioxide (SiO_2, also known as BOX for SOI), usually in 49% concentrated form, 5:1, 10:1 or 20:1 BOE (buffered oxide etchant) or BHF (Buffered HF). They were first used in medieval times for glass etching. It was used in IC fabrication for patterning the gate oxide until the process step was replaced by RIE.

Hydrofluoric acid is considered one of the more dangerous acids in the cleanroom. It penetrates the skin upon contact and it diffuses straight to the bone. Therefore the damage is not felt until it is too late.

Electrochemical etching

Electrochemical etching (ECE) for dopant-selective removal of silicon is a common method to automate and to selectively control etching. An active p-n diode junction is required, and either type of dopant can be the etch-resistant ("etch-stop") material. Boron is the most common etch-stop dopant. In combination with wet anisotropic etching as described above, ECE has been used successfully for controlling silicon diaphragm thickness in commercial piezoresistive silicon pressure sensors. Selectively doped regions can be created either by implantation, diffusion, or epitaxial deposition of silicon.

Dry etching

Vapor etching

Xenon difluoride etching

Xenon difluoride (XeF_2) is a dry vapor phase isotropic etch for silicon originally applied for MEMS in 1995 at University of California, Los Angeles. Primarily used for releasing metal and dielectric structures by undercutting silicon, XeF_2 has the advantage of a stiction-free release unlike wet etchants. Its etch selectivity to silicon is very high, allowing it to work with photoresist, SiO_2, silicon nitride, and various metals for masking. Its reaction to silicon is "plasmaless", is purely chemical and spontaneous and is often operated in pulsed mode. Models of the etching action are available,, and university laboratories and various commercial tools offer solutions using this approach.

Plasma etching

Sputtering

Reactive ion etching (RIE)

In reactive ion etching (RIE), the substrate is placed inside a reactor, and several gases are introduced. A plasma is struck in the gas mixture using an RF power source, which breaks the gas molecules into ions. The ions accelerate towards, and react with, the

surface of the material being etched, forming another gaseous material. This is known as the chemical part of reactive ion etching. There is also a physical part, which is similar to the sputtering deposition process. If the ions have high enough energy, they can knock atoms out of the material to be etched without a chemical reaction. It is a very complex task to develop dry etch processes that balance chemical and physical etching, since there are many parameters to adjust. By changing the balance it is possible to influence the anisotropy of the etching, since the chemical part is isotropic and the physical part highly anisotropic the combination can form sidewalls that have shapes from rounded to vertical. RIE can be deep (Deep RIE or deep reactive ion etching (DRIE)).

Deep RIE (DRIE) is a special subclass of RIE that is growing in popularity. In this process, etch depths of hundreds of micrometres are achieved with almost vertical sidewalls. The primary technology is based on the so-called "Bosch process", named after the German company Robert Bosch, which filed the original patent, where two different gas compositions alternate in the reactor. Currently there are two variations of the DRIE. The first variation consists of three distinct steps (the Bosch Process as used in the UNAXIS tool) while the second variation only consists of two steps (ASE used in the STS tool). In the 1st Variation, the etch cycle is as follows: (i) SF_6 isotropic etch; (ii) C_4F_8 passivation; (iii) SF_6 anisoptropic etch for floor cleaning. In the 2nd variation, steps (i) and (iii) are combined.

Both variations operate similarly. The C_4F_8 creates a polymer on the surface of the substrate, and the second gas composition (SF_6 and O_2) etches the substrate. The polymer is immediately sputtered away by the physical part of the etching, but only on the horizontal surfaces and not the sidewalls. Since the polymer only dissolves very slowly in the chemical part of the etching, it builds up on the sidewalls and protects them from etching. As a result, etching aspect ratios of 50 to 1 can be achieved. The process can easily be used to etch completely through a silicon substrate, and etch rates are 3–6 times higher than wet etching.

MEMS manufacturing technologies

Bulk micromachining

Bulk micromachining is the oldest paradigm of silicon based MEMS. The whole thickness of a silicon wafer is used for building the micro-mechanical structures. Silicon is machined using various etching processes. Anodic bonding of glass plates or additional silicon wafers is used for adding features in the third dimension and for hermetic encapsulation. Bulk micromachining has been essential in enabling high performance pressure sensors and accelerometers that have changed the shape of the sensor industry in the 80's and 90's.

Surface micromachining

Surface micromachining uses layers deposited on the surface of a substrate as the structural materials, rather than using the substrate itself. Surface micromachining was

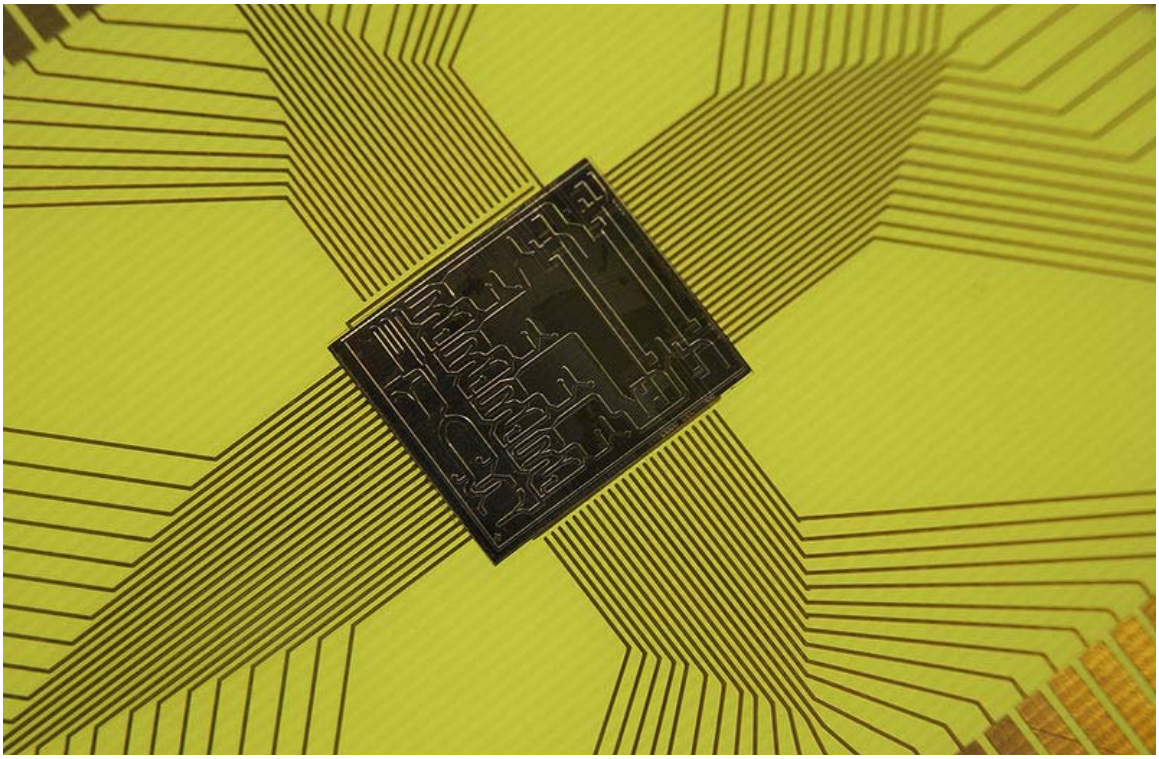
created in the late 1980s to render micromachining of silicon more compatible with planar integrated circuit technology, with the goal of combining MEMS and integrated circuits on the same silicon wafer. The original surface micromachining concept was based on thin polycrystalline silicon layers patterned as movable mechanical structures and released by sacrificial etching of the underlying oxide layer. Interdigital comb electrodes were used to produce in-plane forces and to detect in-plane movement capacitively. This MEMS paradigm has enabled the manufacturing of low cost accelerometers for e.g. automotive air-bag systems and other applications where low performance and/or high g-ranges are sufficient. Analog Devices have pioneered the industrialization of surface micromachining and have realized the co-integration of MEMS and integrated circuits.

High aspect ratio (HAR) silicon micromachining

Both bulk and surface silicon micromachining are used in the industrial production of sensors, ink-jet nozzles, and other devices. But in many cases the distinction between these two has diminished. A new etching technology, deep reactive-ion etching, has made it possible to combine good performance typical of bulk micromachining with comb structures and in-plane operation typical of surface micromachining. While it is common in surface micromachining to have structural layer thickness in the range of 2 μm, in HAR silicon micromachining the thickness can be from 10 to 100 μm. The materials commonly used in HAR silicon micromachining are thick polycrystalline silicon, known as epi-poly, and bonded silicon-on-insulator (SOI) wafers although processes for bulk silicon wafer also have been created (SCREAM). Bonding a second wafer by glass frit bonding, anodic bonding or alloy bonding is used to protect the MEMS structures. Integrated circuits are typically not combined with HAR silicon micromachining. The consensus of the industry at the moment seems to be that the flexibility and reduced process complexity obtained by having the two functions separated far outweighs the small penalty in packaging. A comparison of different high-aspect-ratio microstructure technologies can be found in the HARMST article.

A forgotten history regarding surface micromachining revolved around the choice of polysilicon A or B. Fine grained (<300A grain size, US4897360), post strain annealed pure polysilicon was advocated by Prof Henry Guckel (U. Wisconsin); while a larger grain, doped stress controlled polysilicon was advocated by the UC Berkeley group.

Applications



microelectromechanical systems chip, sometimes called "lab on a chip"

In one viewpoint MEMS application is categorized by type of use.

- Sensor
- Actuator
- Structure

In another view point MEMS applications are categorized by the field of application (commercial applications include):

- Inkjet printers, which use piezoelectrics or thermal bubble ejection to deposit ink on paper.
- Accelerometers in modern cars for a large number of purposes including airbag deployment in collisions.
- Accelerometers in consumer electronics devices such as game controllers (Nintendo Wii), personal media players / cell phones (Apple iPhone, various Nokia mobile phone models, various HTC PDA models) and a number of Digital Cameras (various Canon Digital IXUS models). Also used in PCs to park the hard disk head when free-fall is detected, to prevent damage and data loss.
- MEMS gyroscopes used in modern cars and other applications to detect yaw; e.g., to deploy a roll over bar or trigger dynamic stability control

- Silicon pressure sensors e.g., car tire pressure sensors, and disposable blood pressure sensors
- Displays e.g., the DMD chip in a projector based on DLP technology, which has a surface with several hundred thousand micromirrors
- Optical switching technology, which is used for switching technology and alignment for data communications
- Bio-MEMS applications in medical and health related technologies from Lab-On-Chip to MicroTotalAnalysis (biosensor, chemosensor)
- Interferometric modulator display (IMOD) applications in consumer electronics (primarily displays for mobile devices), used to create interferometric modulation - reflective display technology as found in mirasol displays

Companies with strong MEMS programs come in many sizes. The larger firms specialize in manufacturing high volume inexpensive components or packaged solutions for end markets such as automobiles, biomedical, and electronics. The successful small firms provide value in innovative solutions and absorb the expense of custom fabrication with high sales margins. In addition, both large and small companies work in R&D to explore MEMS technology.

Research and development

Researchers in MEMS use various engineering software tools to take a design from concept to simulation, prototyping and testing. Finite element analysis is often used in MEMS design. Simulation of dynamics, heat, and electrical domains, among others, can be performed by ANSYS, COMSOL and CoventorWare-ANALYZER. Other software, such as CoventorWare-ARCHITECT and MEMS-PRO, is used to produce a design layout suitable for delivery to a fabrication firm and even simulate the MEMS embedded in a system. Once prototypes are on-hand, researchers can test the specimens using various instruments, including laser doppler scanning vibrometers, microscopes, and stroboscopes.

Industry structure

The global market for micro-electromechanical systems, which includes products such as automobile airbag systems, display systems and inkjet cartridges totaled $40 billion in 2006 according to Global MEMS/Microsystems Markets and Opportunities, a research report from SEMI and Yole Developpement and is forecasted to reach $72 billion by 2011.

MEMS devices are defined as die-level components of first-level packaging, and include pressure sensors, accelerometers, gyroscopes, microphones, digital mirror displays, micro fluidic devices, etc. The materials and equipment used to manufacture MEMS devices topped $1 billion worldwide in 2006. Materials demand is driven by substrates, making up over 70 percent of the market, packaging coatings and increasing use of chemical mechanical planarization (CMP). While MEMS manufacturing continues to be

dominated by used semiconductor equipment, there is a migration to 200 mm lines and select new tools, including etch and bonding for certain MEMS applications.

Chapter- 9

Photolithography

Photolithography (or "optical lithography") is a process used in microfabrication to selectively remove parts of a thin film or the bulk of a substrate. It uses light to transfer a geometric pattern from a photo mask to a light-sensitive chemical "photoresist", or simply "resist," on the substrate. A series of chemical treatments then engraves the exposure pattern into the material underneath the photo resist. In complex integrated circuits, for example a modern CMOS, a wafer will go through the photolithographic cycle up to 50 times.

Photolithography shares some fundamental principles with photography in that the pattern in the etching resist is created by exposing it to light, either using a projected image or an optical mask. This procedure is comparable to a high precision version of the method used to make printed circuit boards. Subsequent stages in the process have more in common with etching than to lithographic printing. It is used because it affords exact control over the shape and size of the objects it creates, and because it can create patterns over an entire surface simultaneously. Its main disadvantages are that it requires a flat substrate to start with, it is not very effective at creating shapes that are not flat, and it can require extremely clean operating conditions.

Basic procedure

The wafer track portion of an aligner that uses 365 nm ultraviolet light.

A single iteration of photolithography combines several steps in sequence. Modern cleanrooms use automated, robotic wafer track systems to coordinate the process. The procedure described here omits some advanced treatments, such as thinning agents or edge-bead removal.

Cleaning

If organic or inorganic contaminations are present on the water surface, they are usually removed by wet chemical treatment, e.g. the RCA clean procedure based on solutions containing hydrogen peroxide.

Preparation

The wafer is initially heated to a temperature sufficient to drive off any moisture that may be present on the wafer surface. Wafers that have been in storage must be chemically cleaned to remove contamination. A liquid or gaseous "adhesion promoter", such as Bis(trimethylsilyl)amine ("hexamethyldisilazane", HMDS), is applied to promote

adhesion of the photoresist to the wafer. The phrase "adhesion promoter" is a misnomer, as the surface layer of silicon dioxide on the wafer reacts with the agent to form Methylated Silicon-hydroxide, a highly water repellent layer not unlike the layer of wax on a car's paint. This water repellent layer prevents the aqueous developer from penetrating between the photoresist layer and the wafer's surface, thus preventing so-called lifting of small photoresist structures in the (developing) pattern.

Photoresist application

The wafer is covered with photoresist by spin coating. A viscous, liquid solution of photoresist is dispensed onto the wafer, and the wafer is spun rapidly to produce a uniformly thick layer. The spin coating typically runs at 1200 to 4800 rpm for 30 to 60 seconds, and produces a layer between 0.5 and 2.5 micrometres thick. The spin coating process results in a uniform thin layer, usually with uniformity of within 5 to 10 nanometres. This uniformity can be explained by detailed fluid-mechanical modelling, which shows that the resist moves much faster at the top of the layer than at the bottom, where viscous forces bind the resist to the wafer surface. Thus, the top layer of resist is quickly ejected from the wafer's edge while the bottom layer still creeps slowly radially along the wafer. In this way, any 'bump' or 'ridge' of resist is removed, leaving a very flat layer. Final thickness is also determined by the evaporation of liquid solvents from the resist. For very small, dense features (<125 or so nm), thinner resist thicknesses (<0.5 micrometres) are needed to overcome collapse effects at high aspect ratios; typical aspect ratios are <4:1.

The photo resist-coated wafer is then prebaked to drive off excess photoresist solvent, typically at 90 to 100 °C for 30 to 60 seconds on a hotplate.

Exposure and developing

After prebaking, the photoresist is exposed to a pattern of intense light. Optical lithography typically uses ultraviolet light (see below). Positive photoresist, the most common type, becomes soluble in the basic developer when exposed; exposed negative photoresist becomes insoluble in the (organic) developer. This chemical change allows some of the photoresist to be removed by a special solution, called "developer" by analogy with photographic developer. To learn more about the process of exposure and development of positive resist, see: Ralph Dammel, "Diazonaphtoquinone-based resists", SPIE Optical Engineering Press, Vol TT11 (1993).

A PEB (post-exposure bake) is performed before developing, typically to help reduce standing wave phenomena caused by the destructive and constructive interference patterns of the incident light. In DUV (deep ultraviolet, or 248 nm exposure wavelength) lithography, CAR (chemically amplified resist) chemistry is used. This process is much more sensitive to PEB time, temperature, and delay, as most of the "exposure" reaction (creating acid, making the polymer soluble in the basic developer) actually occurs in the PEB.

The develop chemistry is delivered on a spinner, much like photoresist. Developers originally often contained sodium hydroxide (NaOH). However, sodium is considered an extremely undesirable contaminant in MOSFET fabrication because it degrades the insulating properties of gate oxides (specifically, sodium ions can migrate in and out of the gate, changing the threshold voltage of the transistor and making it harder or easier to turn the transistor on over time). Metal-ion-free developers such as tetramethylammonium hydroxide (TMAH) are now used.

The resulting wafer is then "hard-baked" if a non-chemically amplified resist was used, typically at 120 to 180 °C for 20 to 30 minutes. The hard bake solidifies the remaining photoresist, to make a more durable protecting layer in future ion implantation, wet chemical etching, or plasma etching.

Etching

Etching tanks used to perform Piranha, Hydrofluoric acid or RCA clean on 4-inch wafer batches at LAAS technological facility in Toulouse, France.

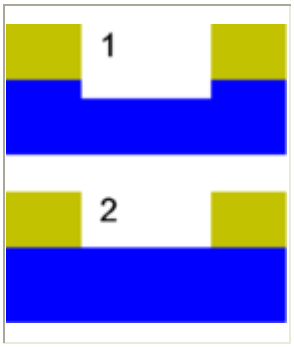
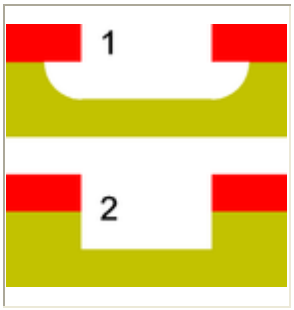
Etching is used in microfabrication to chemically remove layers from the surface of a wafer during manufacturing. Etching is a critically important process module, and every wafer undergoes many etching steps before it is complete.

For many etch steps, part of the wafer is protected from the etchant by a "masking" material which resists etching. In some cases, the masking material is a photoresist which has been patterned using photolithography. Other situations require a more durable mask, such as silicon nitride.

Figures of merit

If the etch is intended to make a cavity in a material, the depth of the cavity may be controlled approximately using the etching time and the known etch rate. More often, though, etching must entirely remove the top layer of a multilayer structure, without damaging the underlying or masking layers. The etching system's ability to do this depends on the ratio of etch rates in the two materials (*selectivity*).

Some etches undercut the masking layer and form cavities with sloping sidewalls. The distance of undercutting is called *bias*. Etchants with large bias are called *isotropic*, because they erode the substrate equally in all directions. Modern processes greatly prefer anisotropic etches, because they produce sharp, well-controlled features.

Selectivity		Yellow: layer to be removed; blue: layer to remain 1. A poorly selective etch removes the top layer, but also attacks the underlying material. 2. A highly selective etch leaves the underlying material unharmed.
Isotropy		Red: masking layer; yellow: layer to be removed 1. A perfectly isotropic etch produces round sidewalls. 2. A perfectly anisotropic etch produces vertical sidewalls.

Etching media and technology

The two fundamental types of etchants are liquid-phase ("wet") and plasma-phase ("dry"). Each of these exists in several varieties.

Wet etching

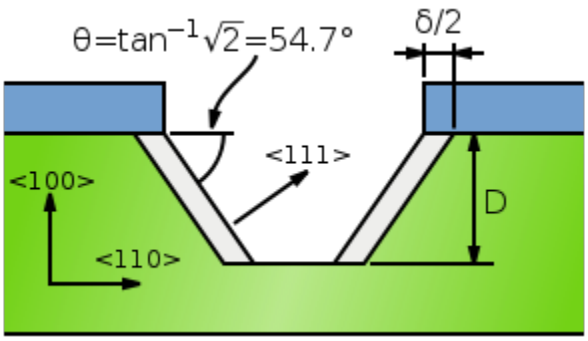
The first etching processes used liquid-phase ("wet") etchants. The wafer can be immersed in a bath of etchant, which must be agitated to achieve good process control. For instance, buffered hydrofluoric acid (BHF) is used commonly to etch silicon dioxide over a silicon substrate.

Different specialised etchants can be used to characterise the surface etched.

Wet etchants are usually isotropic, which leads to large bias when etching thick films. They also require the disposal of large amounts of toxic waste. For these reasons, they are seldom used in state-of-the-art processes. However, the photographic developer used for photoresist resembles wet etching.

As an alternative to immersion, single wafer machines use the Bernoulli principle to employ a gas (usually, pure nitrogen) to cushion and protect one side of the wafer while etchant is applied to the other side. It can be done to either the front side or back side. The etch chemistry is dispensed on the top side when in the machine and the bottom side is not affected. This etch method is particularly effective just before "backend" processing (BEOL), where wafers are normally very much thinner after wafer backgrinding, and very sensitive to thermal or mechanical stress. Etching a thin layer of even a few micrometres will remove microcracks produced during backgrinding resulting in a the wafer having dramatically increased strength and flexibility without breaking.

Anisotropic wet etching



An anisotropic wet etch on a silicon wafer creates a cavity with a trapezoidal cross-section. The bottom of the cavity is a <100> plane, and the sides are <111> planes. The blue material is an etch mask, and the green material is silicon.

Some wet etchants etch crystalline materials at very different rates depending upon which crystal face is exposed. In single-crystal materials (e.g. silicon wafers), this effect can allow very high anisotropy, as shown in the figure.

Several anisotropic wet etchants are available for silicon. For instance, potassium hydroxide (KOH) can achieve selectivity of 400 between <100> and <111> planes.

Another option is EDP (an aqueous solution of ethylene diamine and pyrocatechol), which also displays high selectivity for p-type doping. Neither of these etchants may be used on wafers that contain CMOS integrated circuits. Both of them etch aluminum, commonly used as a metallization (wiring) material. KOH introduces mobile potassium ions into silicon dioxide, and EDP is highly corrosive and carcinogenic. Tetramethylammonium hydroxide (TMAH) presents a safer alternative, although it has even worse selectivity between <100> and <111> planes in silicon than does EDP.

Etching a rectangular hole in a (100)-Si wafer will result in a pyramid shaped etch pit. The wall will be flat and angled (as opposed to curved in isotropic etching), and have an angle to the surface of the wafer of:

$$\tan^{-1}\sqrt{2} = 54.7^\circ$$

If the etching is stopped before the pyramid is formed, a frustum will be formed. The undercut, δ, under the resist mask is given by:

$$\delta = \frac{\sqrt{6}D}{S} = \frac{\sqrt{6}R_{100}T}{R_{100}/R_{111}} = \sqrt{6}TR_{111},$$

where R_{xxx} is the etch rate in the <xxx> direction, T is the etch time, D is the etch depth and S is the anisotropy of the material and etchant.

Plasma etching

Modern VLSI processes avoid wet etching, and use *plasma etching* instead. Plasma etchers can operate in several modes by adjusting the parameters of the plasma. Ordinary plasma etching operates between 0.1 and 5 Torr. (This unit of pressure, commonly used in vacuum engineering, equals approximately 133.3 pascals.) The plasma produces energetic free radicals, neutrally charged, that react at the surface of the wafer. Since neutral particles attack the wafer from all angles, this process is isotropic.

The source gas for the plasma usually contains small molecules rich in chlorine or fluorine. For instance, carbon tetrachloride (CCl_4) etches silicon and aluminium, and trifluoromethane etches silicon dioxide and silicon nitride. A plasma containing oxygen is used to oxidize ("ash") photoresist and facilitate its removal.

Ion milling, or *sputter etching*, uses lower pressures, often as low as 10^{-4} Torr (10 mPa). It bombards the wafer with energetic ions of noble gases, often Ar^+, which knock atoms from the substrate by transferring momentum. Because the etching is performed by ions, which approach the wafer approximately from one direction, this process is highly anisotropic. On the other hand, it tends to display poor selectivity. *Reactive-ion etching* (RIE) operates under conditions intermediate between sputter and plasma etching (between 10^{-3} and 10^{-1} Torr). *Deep reactive-ion etching* (DRIE) modifies the RIE technique to produce deep, narrow features.

Photoresist removal

After a photoresist is no longer needed, it must be removed from the substrate. This usually requires a liquid "resist stripper", which chemically alters the resist so that it no longer adheres to the substrate. Alternatively, photoresist may be removed by a plasma containing oxygen, which oxidizes it. This process is called ashing, and resembles dry etching.

Exposure ("printing") systems

Exposure systems typically produce an image on the wafer using a photomask. The light shines through the photomask, which blocks it in some areas and lets it pass in others. (Maskless lithography projects a precise beam directly onto the wafer without using a mask, but it is not widely used in commercial processes.) Exposure systems may be classified by the optics that transfer the image from the mask to the wafer.

Contact and proximity

A contact printer, the simplest exposure system, puts a photomask in direct contact with the wafer and exposes it to a uniform light. A proximity printer puts a small gap between the photomask and wafer. In both cases, the mask covers the entire wafer, and simultaneously patterns every die.

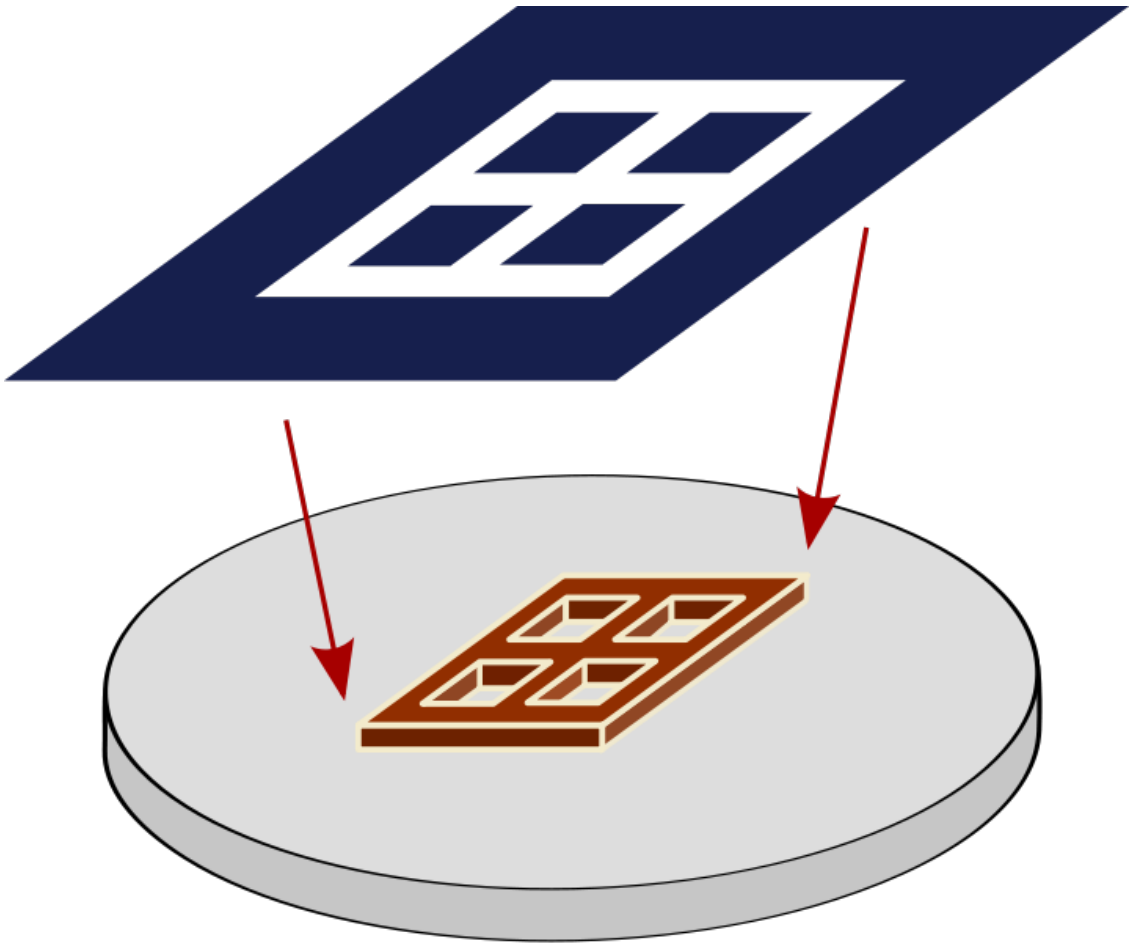
Contact printing is liable to damage both the mask and the wafer, and this was the primary reason it was abandoned for high volume production. Both contact and proximity lithography require the light intensity to be uniform across an entire wafer, and the mask to align precisely to features already on the wafer. As modern processes use increasingly large wafers, these conditions become increasingly difficult.

Research and prototyping processes often use contact lithography, because it uses inexpensive hardware and can achieve high optical resolution. The resolution is approximately the square root of the product of the wavelength and the gap distance. Hence, contact printing offers the best resolution, because its gap distance is approximately zero (neglecting the thickness of the photoresist itself). In addition, nanoimprint lithography may revive interest in this familiar technique, especially since the cost of ownership is expected to below.

Projection

Very-large-scale integration lithography uses projection systems. Unlike contact or proximity masks, which cover an entire wafer, projection masks (known as "reticles") show only one die or an array of die (known as a "field"). Projection exposure systems (steppers) project the mask onto the wafer many times to create the complete pattern.

Photomasks



A schematic illustration of a photomask (top) and an integrated circuit created using that mask (bottom).

A **photomask** is an opaque plate with holes or transparencies that allow light to shine through in a defined pattern. They are commonly used in photolithography.

Overview

Lithographic photomasks are typically transparent fused silica blanks covered with a pattern defined with a chrome metal absorbing film. Photomasks are used at wavelengths of 365 nm, 248 nm, and 193 nm. Photomasks have also been developed for other forms of radiation such as 157 nm, 13.5 nm (EUV), X-ray and electrons and ions, but these require entirely new materials for the substrate and the pattern film.

A set of photomasks, each defining a pattern layer in integrated circuit fabrication, is fed into a photolithography stepper or scanner and individually selected for exposure. In

double patterning techniques, a photomask would correspond to a subset of the layer pattern.

In photolithography for the mass production of integrated circuit devices, the more correct term is usually **photoreticle** or simply **reticle**. In the case of a photomask, there is a one-to-one correspondence between the mask pattern and the wafer pattern. This was the standard for the 1:1 mask aligners that were succeeded by steppers and scanners with reduction optics. As used in steppers and scanners, the reticle commonly contains only one layer of the chip. (However, some photolithography fabs utilize reticles with more than one layer patterned onto the same mask). The pattern is projected and shrunk by four or five times onto the wafer surface. To achieve complete wafer coverage, the wafer is repeatedly 'stepped' from position to position under the optical column until full exposure is achieved.

Features 150 nm or below in size generally require phase-shifting to enhance the image quality to acceptable values. This can be achieved in many ways, but the two most common methods are to use an attenuated phase-shifting background film on the mask to increase the contrast of small intensity peaks, or to etch the exposed quartz so that the edge between the etched and unetched areas can be used to image nearly zero intensity. In the second case, unwanted edges would need to be trimmed out with another exposure. The former method is *attenuated phase-shifting*, and is often considered a weak enhancement, requiring special illumination for the most enhancement, while the latter method is known as *alternating-aperture phase-shifting*, and is the most popular strong enhancement technique.

As leading-edge semiconductor features shrink, photomask features which are 4× larger must inevitably shrink as well. This could pose challenges as the absorber film will need to become thinner, and hence less opaque. A recent study by IMEC has found that thinner absorbers degrade image contrast and hence contribute to line-edge roughness, using state-of-the-art photolithography tools. One possibility is to eliminate absorbers altogether and use 'chromeless' masks, relying solely on phase-shifting for imaging.

The emergence of immersion lithography has a strong impact on photomask requirements. The commonly used attenuated phase-shifting mask is more sensitive to the higher incidence angles applied in "hyper-NA" lithography, due to the longer optical path through the patterned film.

Mask Error Enhancement Factor

Leading edge photomasks contain (pre-corrected) images of the final chip patterns magnified by 4x. This magnification factor has been a key benefit in reducing pattern sensitivity to imaging errors. However, as features continue to shrink, two trends come into play: 1) the mask error factor begins to exceed 1, i.e., the dimension error on the wafer may be more than 1/4 the dimension error on the mask, and 2) the mask feature is becoming smaller, and the dimension tolerance is approaching a few nanometers. As an example, a 25 nm wafer pattern should correspond to a 100 nm mask pattern, but the

wafer tolerance could be 1.25 nm (5% spec) which translates into 5 nm on the photomask. The variation of electron beam scattering in directly writing the photomask pattern can easily well exceed this.

Pellicles

Particle contamination can be a significant problem in semiconductor manufacturing. A photomask is protected from particles by a pellicle – a thin transparent film stretched over a frame that is glued over one side of the photomask. The pellicle is far enough away from the mask patterns so that moderate-to-small sized particles that land on the pellicle will be too far out of focus to print. Although they are designed to keep particles away, pellicles become a part of the imaging system and their optical properties need to be taken into account.

Leading commercial photomask manufacturers

The SPIE Annual Conference, Photomask Technology reports the SEMATECH Mask Industry Assessment which includes current industry analysis and the results of their annual photomask manufacturers survey. The following companies are listed in order of their global market share (2009 info).

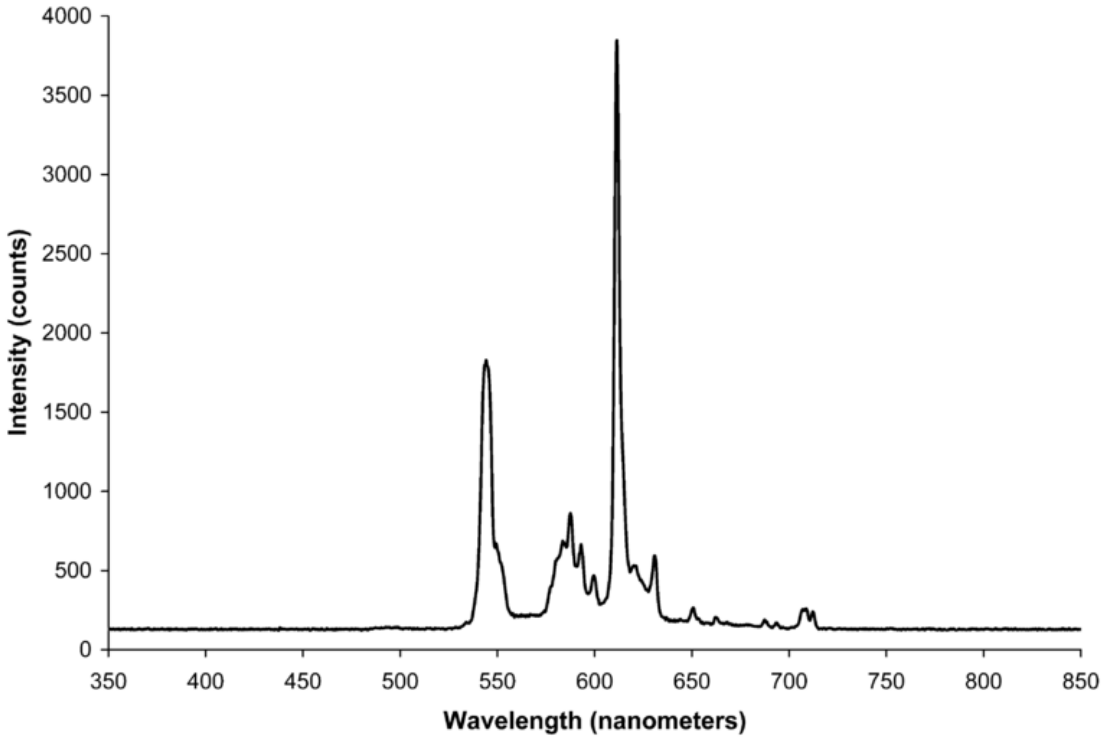
- Dai Nippon Printing Company
- Toppan Photomasks
- Photronics
- Hoya Japan
- Taiwan Mask Corporation
- Compugraphics Photomask Solutions

Major chipmakers such as Intel (Intel Mask Operations), AMD, IBM, NEC, TSMC, Samsung, Micron Technology, etc., have their own large maskmaking facilities.

- Photo Sciences, Inc. is the largest privately held photomask manufacturer in the United States, 2009.

The cost to set up a modern 45 nm process mask shop is US$200-500 million, a very high threshold for entering this market. The purchase price of a photomask can range from $1,000 to $100,000 for a single, high end phase shift mask; as many as 30 masks (of varying price) may be required to form a complete mask set.

Resolution in projection systems



The filtered fluorescent lighting in photolithography cleanrooms contains no ultraviolet or blue light in order to avoid exposing photoresists. The spectrum of light emitted by such fixtures gives virtually all such spaces a bright yellow color.

The ability to project a clear image of a small feature onto the wafer is limited by the wavelength of the light that is used, and the ability of the reduction lens system to capture enough diffraction orders from the illuminated mask. Current state-of-the-art photolithography tools use deep ultraviolet (DUV) light with wavelengths of 248 and 193 nm, which allow minimum feature sizes down to 50 nm.

The minimum feature size that a projection system can print is given approximately by:

$$CD = k_1 \cdot \frac{\lambda}{NA}$$

where

CD is the **minimum feature size** (also called the **critical dimension**, *target design rule*). It is also common to write 2 *times* the *half-pitch*.

k_1(commonly called *k1 factor*) is a coefficient that encapsulates process-related factors, and typically equals 0.4 for production. The minimum feature size can be reduced by decreasing this coefficient through Computational lithography.

$\lambda$is the wavelength of light used

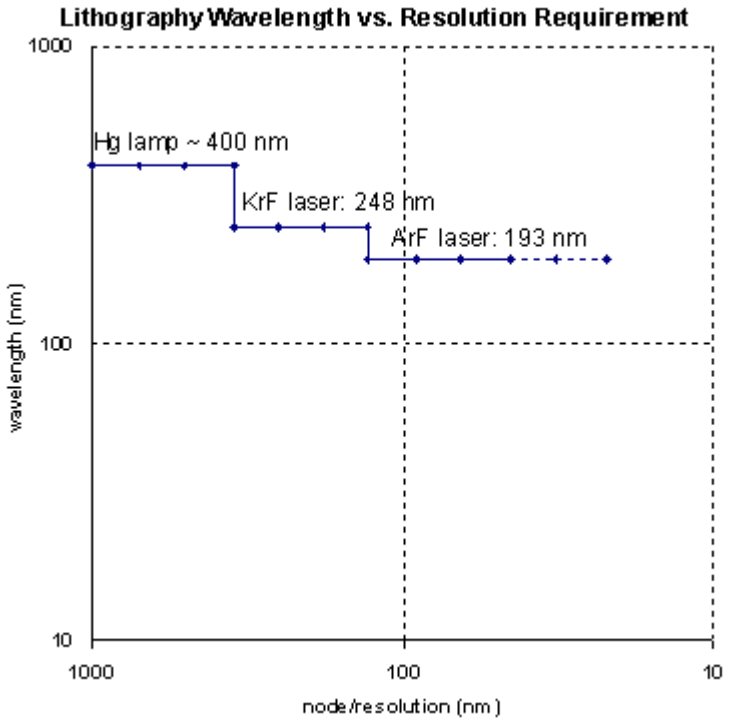
$NA$is the numerical aperture of the lens as seen from the wafer

According to this equation, minimum feature sizes can be decreased by decreasing the wavelength, and increasing the numerical aperture (to achieve a tighter focused beam and a smaller spot size). However, this design method runs into a competing constraint. In modern systems, the depth of focus is also a concern:

$$D_F = k_2 \cdot \frac{\lambda}{NA^2}$$

Here, $k_2$is another process-related coefficient. The depth of focus restricts the thickness of the photoresist and the depth of the topography on the wafer. Chemical mechanical polishing is often used to flatten topography before high-resolution lithographic steps.

Light sources

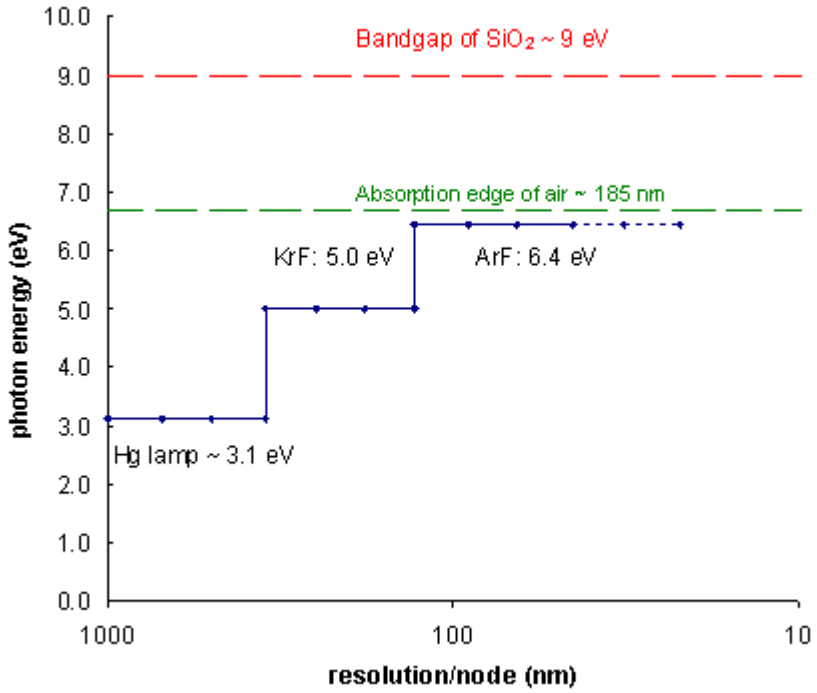


The evolution of lithography wavelength corresponding to different light sources. It is worth noting that the same light source may be used for several technology generations.

Historically, photolithography has used ultraviolet light from gas-discharge lamps using mercury, sometimes in combination with noble gases such as xenon. These lamps produce light across a broad spectrum with several strong peaks in the ultraviolet range. This spectrum is filtered to select a single spectral line, usually the "g-line" (436 nm) or "i-line" (365 nm).

More recently, lithography has moved to "deep ultraviolet", produced by excimer lasers. (In lithography, wavelengths below 300 nm are called "deep UV".) Krypton fluoride produces a 248-nm spectral line, and argon fluoride a 193-nm line. Excimer laser light sources for photolithography applications were first introduced by Cymer Inc. in the late 1980s for 248-nm wavelength lithography (KrF), and in 2001 for 193-nm wavelength lithography (ArF). Generally, changing wavelength is not a trivial matter, as the method of generating the new wavelength is completely different, and the absorption characteristics of materials change. For example, air begins to absorb significantly around the 193 nm wavelength; moving to sub-193 nm wavelengths would require installing vacuum pump and purge equipment on the lithography tools (a significant challenge). Furthermore, insulating materials such as silicon dioxide (SiO_2), when exposed to photons with energy greater than the band gap, release free electrons and holes which subsequently cause adverse charging.

Optical lithography has been extended to feature sizes below 50 nm using 193 nm and liquid immersion techniques. Also termed immersion lithography, this enables the use of optics with numerical apertures exceeding 1.0. The liquid used is typically ultra-pure, deionised water, which provides for a refractive index above that of the usual air gap between the lens and the wafer surface. The water is continually circulated to eliminate thermally-induced distortions. Water will only allow *NA*'s of up to ~1.4, but materials with higher refractive indices will allow the effective *NA* to be increased further.



Changing the lithography wavelength is significantly limited by absorption. Air absorbs below ~ 185 nm.

Experimental tools using 157 nm wavelength DUV in a manner similar to current exposure systems have been built. These were once targeted to succeed 193 nm at the 65 nm feature size node but have now all but been eliminated by the introduction of immersion lithography. This was due to persistent technical problems with the 157 nm technology and economic considerations that provided strong incentives for the continued use of 193 nm technology. High-index immersion lithography is the newest extension of 193 nm lithography to be considered. In 2006, features less than 30 nm were demonstrated by IBM using this technique.

Experimental methods

Photolithography has been defeating predictions of its demise for many years. For instance, it was predicted that features smaller than 1 micrometre could not be printed optically. Modern techniques already print features with dimensions a fraction of the wavelength of light used - an amazing optical feat. Current research is exploring new tricks in the ultraviolet regime, as well as alternatives to conventional UV, such as electron beam lithography, X-ray lithography, extreme ultraviolet lithography, ion projection lithography, and immersion lithography.

Chapter- 10

Electron Beam Lithography

Electron beam lithography (often abbreviated as **e-beam lithography**) is the practice of scanning a beam of electrons in a patterned fashion across a surface covered with a film (called the resist), ("exposing" the resist) and of selectively removing either exposed or non-exposed regions of the resist ("developing"). The purpose, as with photolithography, is to create very small structures in the resist that can subsequently be transferred to the substrate material, often by etching. It was developed for manufacturing integrated circuits, and is also used for creating nanotechnology architectures.

The primary advantage of electron beam lithography is that it is one of the ways to beat the diffraction limit of light and make features in the nanometer regime. This form of maskless lithography has found wide usage in photomask-making used in photolithography, low-volume production of semiconductor components, and research & development.

On the other hand, the key limitation of electron beam lithography is throughput, i.e., the very long time it takes to expose an entire silicon wafer or glass substrate. A long exposure time leaves the user vulnerable to beam drift or instability which may occur during the exposure. Also, the turn-around time for reworking or re-design is lengthened unnecessarily if the pattern is not being changed the second time.

Electron beam lithography systems

Electron beam lithography systems used in commercial applications are dedicated e-beam writing systems that are very expensive (>$4M USD). For research applications, it is very common to convert an electron microscope into an electron beam lithography system using a relatively low cost accessory (<USD 100k). Such converted systems have produced linewidths of ~20 nm since at least 1990, while current dedicated systems have produced linewidths on the order of 10 nm or smaller.

Electron beam lithography systems can be classified according to both beam shape and beam deflection strategy. Older systems used Gaussian-shaped beams and scanned these

beams in a raster fashion. Newer systems use shaped beams, which may be deflected to various positions in the writing field (this is also known as **vector scan**).

Electron sources

Lower resolution systems can use thermionic sources, which are usually formed from LaB_6. However, systems with higher resolution requirements need to use field electron emission sources, such as heated W/ZrO_2 for lower energy spread and enhanced brightness. Thermal field emission sources are preferred over cold emission sources, in spite of their slightly larger beam size, because the former offer better stability over typical writing times of several hours.

Lenses

Both electrostatic and magnetic lenses may be used. However, electrostatic lenses have more aberrations and so are not used for fine focusing. There is no current mechanism to make achromatic electron beam lenses, so extremely narrow dispersions of the electron beam energy are needed for finest focusing.

Stage, stitching and alignment

Typically, for very small beam deflections electrostatic deflection 'lenses' are used, larger beam deflections require electromagnetic scanning. Because of the inaccuracy and because of the finite number of steps in the exposure grid the writing field is of the order of 100 micrometre - 1 mm. Larger patterns require stage moves. An accurate stage is critical for stitching (tiling writing fields exactly against each other) and pattern overlay (aligning a pattern to a previously made one).

Electron beam write time

The minimum time to expose a given area for a given dose is given by the following formula:

$$D \cdot A = T \cdot I$$

where T is the time to expose the object (can be divided into exposure time/step size), I is the beam current, D is the dose and A is the area exposed.

For example, assuming an exposure area of 1 cm^2, a dose of 10^{-3} Coulombs/cm^2, and a beam current of 10^{-9} Amperes, the resulting minimum write time would be 10^6 seconds (about 12 days). This minimum write time does not include time for the stage to move back and forth, as well as time for the beam to be blanked (blocked from the wafer during deflection), as well as time for other possible beam corrections and adjustments in the middle of writing. To cover the 700 cm^2 surface area of a 300 mm silicon wafer, the minimum write time would extend to $7*10^8$ seconds, about 22 years. This is a factor of about 10 million times slower than current optical lithography tools. It is clear that

throughput is a serious limitation for electron beam lithography, especially when writing dense patterns over a large area.

E-beam lithography is not suitable for high-volume manufacturing because of its limited throughput. The smaller field of electron beam writing makes for very slow pattern generation compared with photolithography (the current standard) because more exposure fields must be scanned to form the final pattern area ($\leq mm^2$ for electron beam vs. $\geq 40\ mm^2$ for an optical mask projection scanner). The stage moves in between field scans. The electron beam field is small enough that a rastering or serpentine stage motion is needed to pattern a 26 mm X 33 mm area for example, whereas in a photolithography scanner only a one-dimensional motion of a 26 mm X 2 mm slit field would be required.

Currently an optical maskless lithography tool is much faster than an electron beam tool used at the same resolution for photomask patterning.

Defects in electron-beam lithography

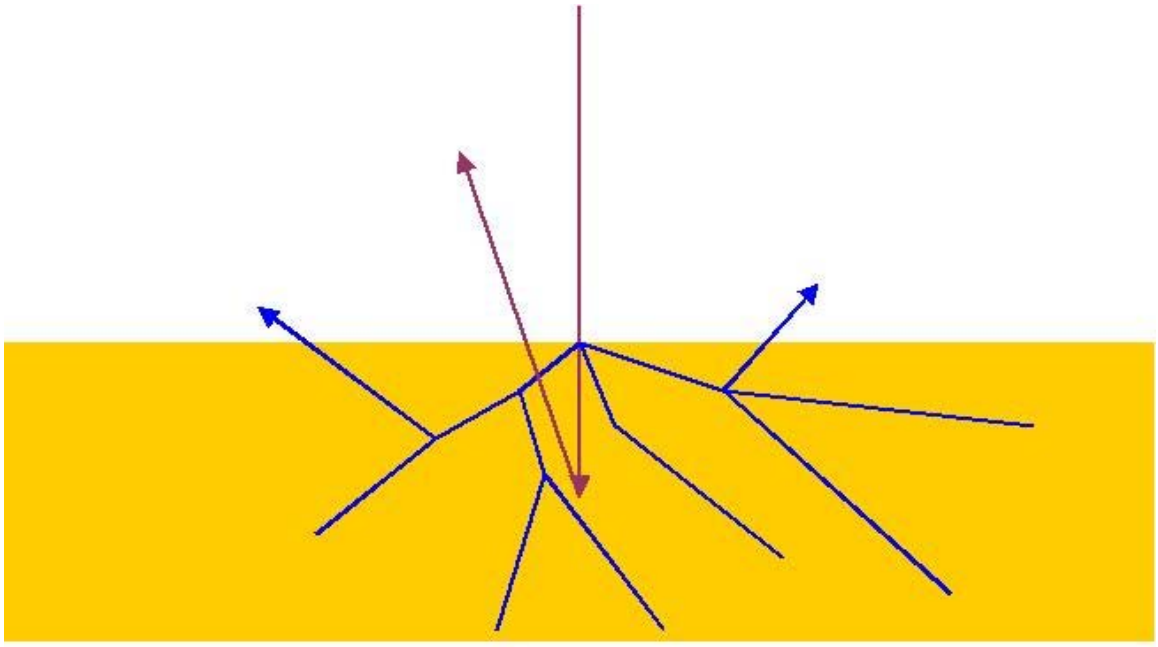
Despite the high resolution of electron-beam lithography, the generation of defects during electron-beam lithography is often not considered by users. Defects may be classified into two categories: data-related defects, and physical defects.

Data-related defects may be classified further into two sub-categories. **Blanking** or **deflection errors** occur when the electron beam is not deflected properly when it is supposed to, while **shaping errors** occur in variable-shaped beam systems when the wrong shape is projected onto the sample. These errors can originate either from the electron optical control hardware or the input data that was taped out. As might be expected, larger data files are more susceptible to data-related defects.

Physical defects are more varied, and can include sample charging (either negative or positive), backscattering calculation errors, dose errors, fogging (long-range reflection of backscattered electrons), outgassing, contamination, beam drift and particles. Since the write time for electron beam lithography can easily exceed a day, "randomly occurring" defects are more likely to occur. Here again, larger data files can present more opportunities for defects.

Photomask defects largely originate during the electron beam lithography used for pattern definition.

Electron energy deposition in matter



Electron trajectories in resist: An incident electron (purple) produces secondary electrons (blue). Sometimes, the incident electron may itself be backscattered as shown here and leave the surface of the resist (amber).

The primary electrons in the incident beam lose energy upon entering a material through inelastic scattering or collisions with other electrons. In such a collision the momentum transfer from the incident electron to an atomic electron can be expressed as $dp = 2e^2 / bv$, where b is the distance of closest approach between the electrons, and v is the incident electron velocity. The energy transferred by the collision is given by $T = (dp)^2 / 2m = e^4 / Eb^2$, where m is the electron mass and E is the incident electron energy, given by $E = (1 / 2)mv^2$. By integrating over all values of T between the lowest binding energy, E_0 and the incident energy, one obtains the result that the total cross section for collision is inversely proportional to the incident energy E, and proportional to $1/E_0 – 1/E$. Generally, $E >> E_0$, so the result is essentially inversely proportional to the binding energy.

By using the same integration approach, but over the range $2E_0$ to E, one obtains by comparing cross-sections that half of the inelastic collisions of the incident electrons produce electrons with kinetic energy greater than E_0. These secondary electrons are capable of breaking bonds (with binding energy E_0) at some distance away from the original collision. Additionally, they can generate additional, lower energy electrons, resulting in an electron cascade. Hence, it is important to recognize the significant contribution of secondary electrons to the spread of the energy deposition.

In general, for a molecule AB:

$$e^- + AB \rightarrow AB^- \rightarrow A + B^-$$

This reaction, also known as "electron attachment" or "dissociative electron attachment" is most likely to occur after the electron has essentially slowed to a halt, since it is easiest to capture at that point. The cross-section for electron attachment is inversely proportional to electron energy at high energies, but approaches a maximum limiting value at zero energy. On the other hand, it is already known that the mean free path at the lowest energies (few to several eV or less, where dissociative attachment is significant) is well over 10 nm, thus limiting the ability to consistently achieve resolution at this scale.

Resolution capability

With today's electron optics, electron beam widths can routinely go down to a few nm. This is limited mainly by aberrations and space charge. However, the feature resolution limit is determined not by the beam size but by forward scattering (or effective beam broadening) in the photoresist while the pitch resolution limit is determined by secondary electron travel in the photoresist. This point is driven home by the 2007 demonstration of double patterning using electron beam lithography in the fabrication of 15 nm half-pitch zone plates. Although a 15 nm feature was resolved, a 30 nm pitch was still difficult to do, due to secondary electrons scattering from the adjacent feature. The use of double patterning allowed the spacing between features to be wide enough for the secondary electron scattering to be significantly reduced. The forward scattering can be decreased by using higher energy electrons or thinner photoresist, but the generation of secondary electrons is inevitable. It is now recognized that for insulating materials like PMMA, low energy electrons can travel quite a far distance (several nm is possible). This is due to the fact that below the ionization potential the only energy loss mechanism is mainly through phonons and polarons, although the latter is basically an ionic lattice effect. In fact, polaron hopping could extend as far as 20 nm. The travel distance of secondary electrons is not a fundamentally derived physical value, but a statistical parameter often determined from many experiments or Monte Carlo simulations down to < 1 eV. This is necessary since the energy distribution of secondary electrons peaks well below 10 eV. Hence, the resolution limit is not usually cited as a well-fixed number as with an optical diffraction-limited system. Repeatability and control at the practical resolution limit often require considerations not related to image formation, e.g., photoresist development and intermolecular forces.

Scattering

In addition to producing secondary electrons, primary electrons from the incident beam with sufficient energy to penetrate the photoresist can be multiply scattered over large distances from underlying films and/or the substrate. This leads to exposure of areas at a significant distance from the desired exposure location. For thicker electrons, as the primary electrons move forward, they have an increasing opportunity to scatter laterally from the beam-defined location. This scattering is called **forward scattering**. Sometimes the primary electrons are scattered at angles exceeding 90 degrees, i.e., they no longer advance further into the resist. These electrons are called **backscattered electrons** and

have the same effect as long-range flare in optical projection systems. A large enough dose of backscattered electrons can lead to complete exposure of resist over an area much larger than defined by the beam spot.

Proximity effect

The smallest features produced by electron beam lithography have generally been isolated features, as nested features exacerbate the proximity effect, whereby electrons from exposure of an adjacent region spill over into the exposure of the currently written feature, effectively enlarging its image, and reducing its contrast, i.e., difference between maximum and minimum intensity. Hence, nested feature resolution is harder to control. For most resists, it is difficult to go below 25 nm lines and spaces, and a limit of 20 nm lines and spaces has been found. In actuality, though, the range of secondary electron scattering is quite far, sometimes exceeding 100 nm, but becoming very significant below 30 nm.

The proximity effect is also manifest by secondary electrons leaving the top surface of the resist and then returning some tens of nanometers distance away.

Proximity effects (due to electron scattering) can be addressed by solving the inverse problem and calculating the exposure function *E(x,y)* that leads to a dose distribution as close as possible to the desired dose *D(x,y)* when convolved by the scattering distribution point spread function *PSF(x,y)*. However, it must be remembered that an error in the applied dose (e.g., from shot noise) would cause the proximity effect correction to fail.

Charging

Since electrons are charged particles, they tend to charge the substrate negatively unless they can quickly gain access to a path to ground. For a high-energy beam incident on a silicon wafer, virtually all the electrons stop in the wafer where they can follow a path to ground. However, for a quartz substrate such as a photomask, the embedded electrons will take a much longer time to move to ground. Often the negative charge acquired by a substrate can be compensated or even exceeded by a positive charge on the surface due to secondary electron emission into the vacuum. The presence of a thin conducting layer above or below the resist is generally of limited use for high energy (50 keV or more) electron beams, since most electrons pass through the layer into the substrate. The charge dissipation layer is generally useful only around or below 10 keV, since the resist is thinner and most of the electrons either stop in the resist or close to the conducting layer. However, they are of limited use due to their high sheet resistance, which can lead to ineffective grounding.

The range of low-energy secondary electrons (the largest component of the free electron population in the resist-substrate system) which can contribute to charging is not a fixed number but can vary from 0 to as high as 50 nm. Hence, resist-substrate charging is not repeatable and is difficult to compensate consistently. Negative charging deflects the

electron beam away from the charged area while positive charging deflects the electron beam toward the charged area.

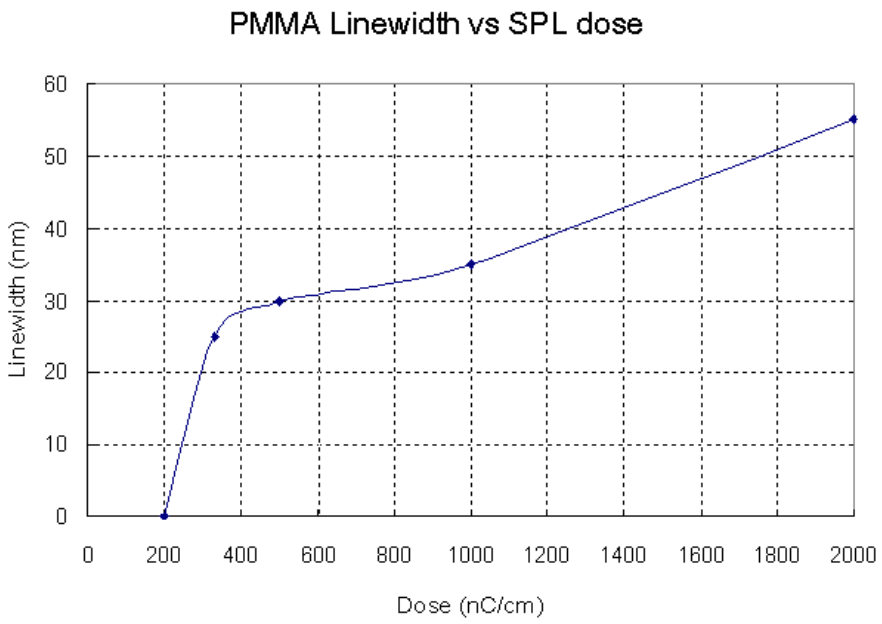
Electron beam resist performance

A study performed at the Naval Research Laboratory indicated that low-energy (10−50 eV) electrons were able to damage ~30 nm thick PMMA films. The damage was manifest as a loss of material.

- For the popular electron-beam resist ZEP-520, a pitch resolution limit of 60 nm (30 nm lines and spaces), independent of thickness and beam energy, was found.
- A 20 nm resolution had also been demonstrated using a 3 nm 100 keV electron beam and PMMA resist. 20 nm unexposed gaps between exposed lines showed inadvertent exposure by secondary electrons.
- Hydrogen silsesquioxane (HSQ) is a negative resist that is capable of forming sub-30 nm lines in very thin layers, but is itself similar to porous, hydrogenated SiO2. It may be used to etch silicon but not silicon dioxide or other similar dielectrics.

New frontiers in electron-beam lithography

To get around the secondary electron generation, it will be imperative to use low-energy electrons as the primary radiation to expose photoresist. Ideally, these electrons should have energies on the order of not much more than several eV in order to expose the photoresist without generating any secondary electrons, since they will not have sufficient excess energy. Such exposure has been demonstrated using a scanning tunneling microscope as the electron beam source. The data suggest that electrons with energies as low as 12 eV can penetrate 50 nm thick polymer photoresist. The drawback to using low energy electrons is that it is hard to prevent spreading of the electron beam in the photoresist. Low energy electron optical systems are also hard to design for high resolution. Coulomb inter-electron repulsion always becomes more severe for lower electron energy.



Scanning probe lithography. A scanning probe can be used for low-energy electron beam lithography, offering sub-100 nm resolution, determined by the dose of low-energy electrons.

Another alternative in electron-beam lithography is to use extremely high electron energies (at least 100 keV) to essentially "drill" or sputter the material. This phenomenon has been observed frequently in transmission electron microscopy. However, this is a very inefficient process, due to the inefficient transfer of momentum from the electron beam to the material. As a result it is a slow process, requiring much longer exposure times than conventional electron beam lithography. Also high energy beams always bring up the concern of substrate damage.

Interference lithography using electron beams is another possible path for patterning arrays with nanometer-scale periods. A key advantage of using electrons over photons in interferometry is the much shorter wavelength for the same energy.

Despite the various intricacies and subtleties of electron beam lithography at different energies, it remains the most practical way to concentrate the most energy into the smallest area.

There has been significant interest in the development of multiple electron beam approaches to lithography in order to increase throughput. This work has been supported by SEMATECH and start-up companies such as Multibeam Corporation and Mapper. However, the degree of parallelism required to be competitive would need to be very high (at least 10 million, as estimated above); this is far in excess of most scheduled demonstrations.

Chapter- 11

Other Systems and Processes of Microelectromechanical Systems

Surface micromachining

Surface micromachining is a process used to produce micromachinery or Microelectromechanical systems.

Unlike Bulk micromachining, where a silicon substrate (wafer) is selectively etched to produce structures, surface micromachining builds microstructures by deposition and etching of different structural layers on top of the substrate.. Generally *polysilicon* is commonly used as one of the layers and *silicon dioxide* is used as a *sacrificial layer* which is removed or etched out to create the necessary void in the thickness direction. Added layers are generally very thin with their size varying from 2-5 Micro metres. The main advantage of this machining process is the possibility of realizing monolithic microsystems in which the electronic and the mechanical components(functions) are built in on the same substrate. The surface micromachined components are smaller compared to their counterparts, the bulk micromachined ones.

As the structures are built on top of the substrate and not inside it, the substrate's properties are not as important as in bulk micromachining, and the expensive silicon wafers can be replaced by cheaper substrates, such as glass or plastic. The size of the substrates can also be much larger than a silicon wafer, and surface micromachining is used to produce TFTs on large area glass substrates for flat panel displays. This technology can also be used for the manufacture of thin film solar cells, which can be deposited on glass, but also on PET substrates or other non-rigid materials.

Fabrication Process

Micromachining starts with a silicon wafer or other substrate and grows layers on top. These layers are selectively etched by photolithography and either a wet etch involving an acid or a dry etch involving an ionized gas, or plasma. Dry etching can combine

chemical etching with physical etching, or ion bombardment of the material. Surface micromachining can involve as many layers as is needed with a different mask (producing a different pattern) on each layer. Modern integrated circuit fabrication uses this technique and can use dozens of layers, approaching 100. Micromachining is a younger technology and usually uses no more than 5 or 6 layers. Surface micromachining uses developed technology (although sometimes not enough for demanding applications)which is very repeatable for volume production.

Sacrificial Layers

Complicated components, such as movable parts, are built using a *sacrificial layer*. For example, a suspended cantilever can be built by depositing and structuring a sacrificial layer, which is then selectively removed at the locations where the future beams must be attached to the substrate (i.e. the anchor points). The *structural layer* is then deposited on top of the polymer and structured to define the beams. Finally, the sacrificial layer is removed to release the beams, using a selective etch process that will not damage the structural layer.

There are many possible combinations of structural/sacrificial layer. The combination chosen depends on the process. For example it is important for the structural layer not to be damaged by the process used to remove the sacrificial layer.

Bulk micromachining

Bulk micromachining is a process used to produce micromachinery or microelectromechanical systems (MEMS).

Unlike surface micromachining, which uses a succession of thin film deposition and selective etching, bulk micromachining defines structures by selectively etching inside a substrate. Whereas surface micromachining creates structures *on top* of a substrate, bulk micromachining produces structures *inside* a substrate.

Usually, silicon wafers are used as substrates for bulk micromachining, as they can be anisotropically wet etched, forming highly regular structures. Wet etching typically uses alkaline liquid solvents, such as potassium hydroxide (KOH) or tetramethylammonium hydroxide (TMAH) to dissolve silicon which has been left exposed by the photolithography masking step. These alkali solvents dissolve the silicon in a highly anisotropic way, with some crystallographic orientations dissolving up to 1000 times faster than others. Such an approach is often used with very specific crystallographic orientations in the raw silicon to produce V-shaped grooves. The surface of these grooves can be atomically smooth if the etch is carried out correctly, and the dimensions and angles can be precisely defined.

Bulk micromachining starts with a silicon wafer or other substrates which is selectively etched, using photolithography to transfer a pattern from a mask to the surface. Like surface micromachining, bulk micromachining can be performed with wet or dry etches, although the most common etch in silicon is the anisotropic wet etch. This etch takes advantage of the fact that silicon has a crystal structure, which means its atoms are all arranged periodically in lines and planes. Certain planes have weaker bonds and are more susceptible to etching. The etch results in pits that have angled walls, with the angle being a function of the crystal orientation of the substrate. This type of etching is inexpensive and is generally used in early, low-budget research.

Deep reactive-ion etching

Deep reactive-ion etching (**DRIE**) is a highly anisotropic etch process used to create deep, steep-sided holes and trenches in wafers, with aspect ratios of 20:1 or more. It was developed for microelectromechanical systems (MEMS), which require these features, but is also used to excavate trenches for high-density capacitors for DRAM and more recently for creating through wafer via's (TSV)'s in advanced 3D wafer level packaging technology .

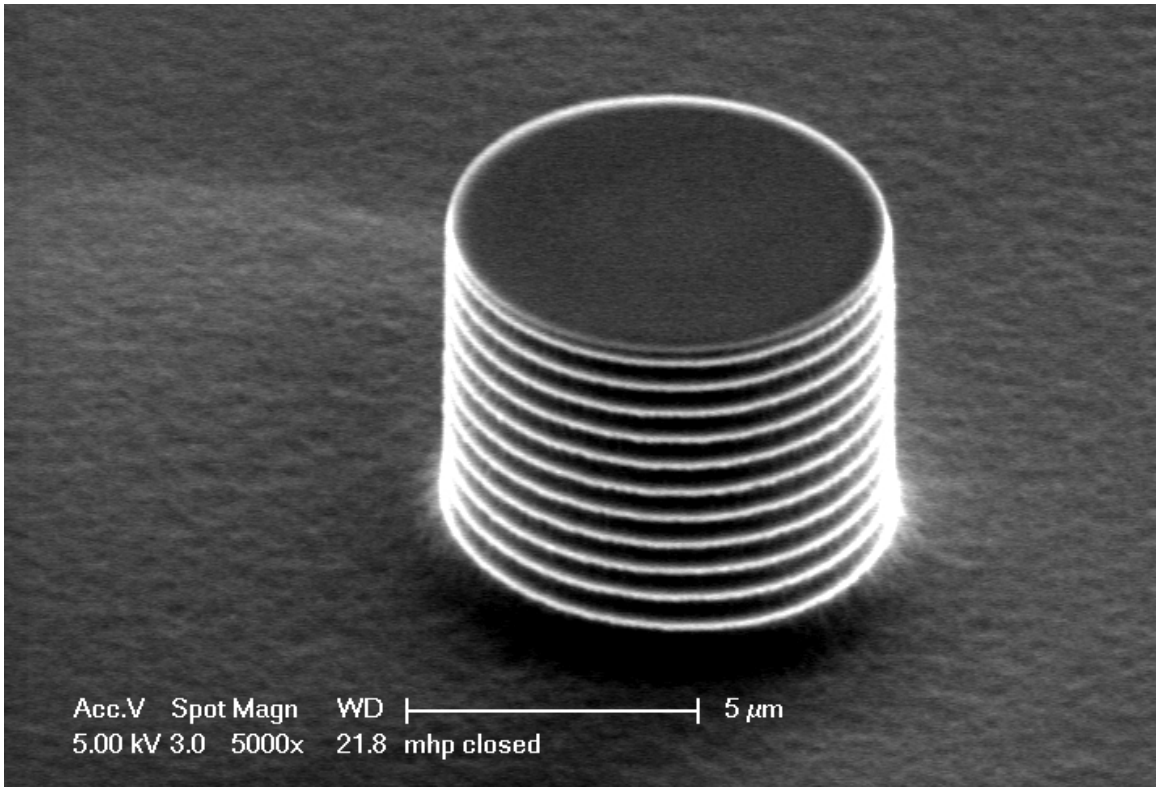
There are two main technologies for high-rate DRIE: cryogenic and Bosch, although the Bosch process is the only recognised production technique. Both Bosch and cryo processes can fabricate 90° (truly vertical) walls, but often the walls are slightly tapered, e.g. 88° or 92° ("retrograde").

Another mechanism is sidewall passivation: SiO_xF_y functional groups (which originate from sulphur hexafluoride and oxygen etch gases) condensate on the sidewalls, and protect them from lateral etching. As a combination of these processes deep vertical structures can be made.

Cryogenic process

In cryo-DRIE, the wafer is chilled to −110 °C (163 K). The low temperature slows down the chemical reaction that produces isotropic etching. However, ions continue to bombard upward-facing surfaces and etch them away. This process produces trenches with highly vertical sidewalls. The primary issues with cryo-DRIE is that the standard masks on substrates crack under the extreme cold, plus etch by-products have a tendency of depositing on the nearest cold surface, i.e. the substrate or electrode.

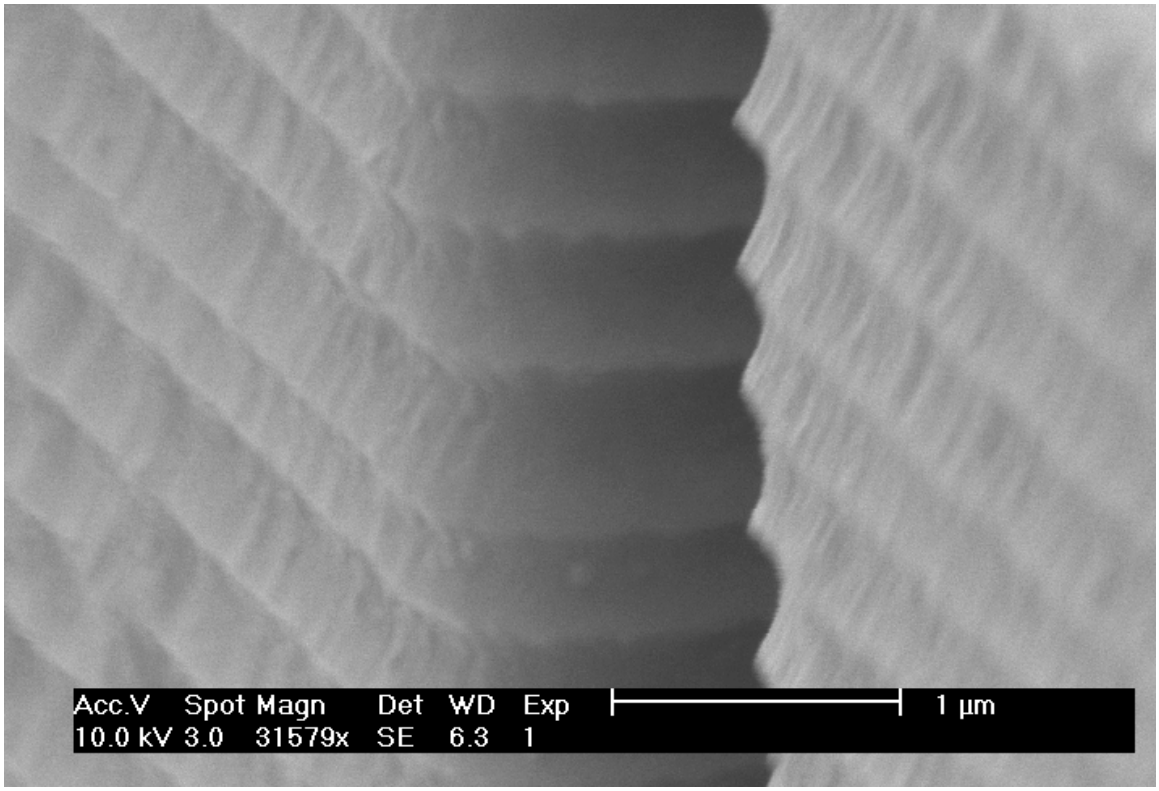
Bosch process



A silicon micro-pillar fabricated using the Bosch process

The Bosch process, also known as pulsed or time-multiplexed etching, alternates repeatedly between two modes to achieve nearly vertical structures.

1. A standard, nearly isotropic plasma etch. The plasma contains some ions, which attack the wafer from a nearly vertical direction. (For silicon, this often uses sulfur hexafluoride [SF_6].)
2. Deposition of a chemically inert passivation layer. (For instance, C_4F_8 source gas yields a substance similar to Teflon.)



Undulating sidewall of a silicon structure created using the Bosch process

Each phase lasts for several seconds. The passivation layer protects the entire substrate from further chemical attack and prevents further etching. However, during the etching phase, the directional ions that bombard the substrate attack the passivation layer at the bottom of the trench (but not along the sides). They collide with it and sputter it off, exposing the substrate to the chemical etchant.

These etch/deposit steps are repeated many times over resulting in a large number of very small isotropic etch steps taking place only at the bottom of the etched pits. To etch through a 0.5 mm silicon wafer, for example, 100–1000 etch/deposit steps are needed. The two-phase process causes the sidewalls to undulate with an amplitude of about 100–500 nm. The cycle time can be adjusted: short cycles yield smoother walls, and long cycles yield a higher etch rate.

Applications

RIE "deepness" depends on application:

- in DRAM memory circuits, capacitor trenches may be 10–20 µm deep,
- in MEMS, DRIE is used for anything from a few micrometers to 0.5 mm.

What distinguishes DRIE from RIE is etch depth: Practical etch depths for RIE (as used in IC manufacturing)would be limited to around 10 µm at a rate up to 1 µm/min, while

DRIE can etch features much greater, up to 600 μm or more with rates up to 20 μm/min or more in some applications.

DRIE of glass requires high plasma power, which makes it difficult to find suitable mask materials for truly deep etching. Polysilicon and nickel are used for 10–50 μm etched depths. In DRIE of polymers, Bosch process with alternating steps of SF_6 etching and C_4F_8 passivation take place. Metal masks can be used however are expensive to use in that several additional photo and deposition steps are always required. Metal masks are not necessary however on various substrates (Si [up to 800 μm], InP [up to 40 μm] or glass [up to 12 μm]) if using chemically amplified negative resists from providers such as Futurrex, Inc. or others.

Gallium ion implantion can be used as etch mask in cryo-DRIE. Combined nanofabrication process of focused ion beam and cryo-DRIE was first reported by N Chekurov *et al* in their article "The fabrication of silicon nanostructures by local gallium implantation and cryogenic deep reactive ion etching" (Nanotechnology, 2009).

Reactive-ion etching

Reactive ion etching (**RIE**) is an etching technology used in microfabrication. It uses chemically reactive plasma to remove material deposited on wafers. The plasma is generated under low pressure (vacuum) by an electromagnetic field. High-energy ions from the plasma attack the wafer surface and react with it.

Equipment

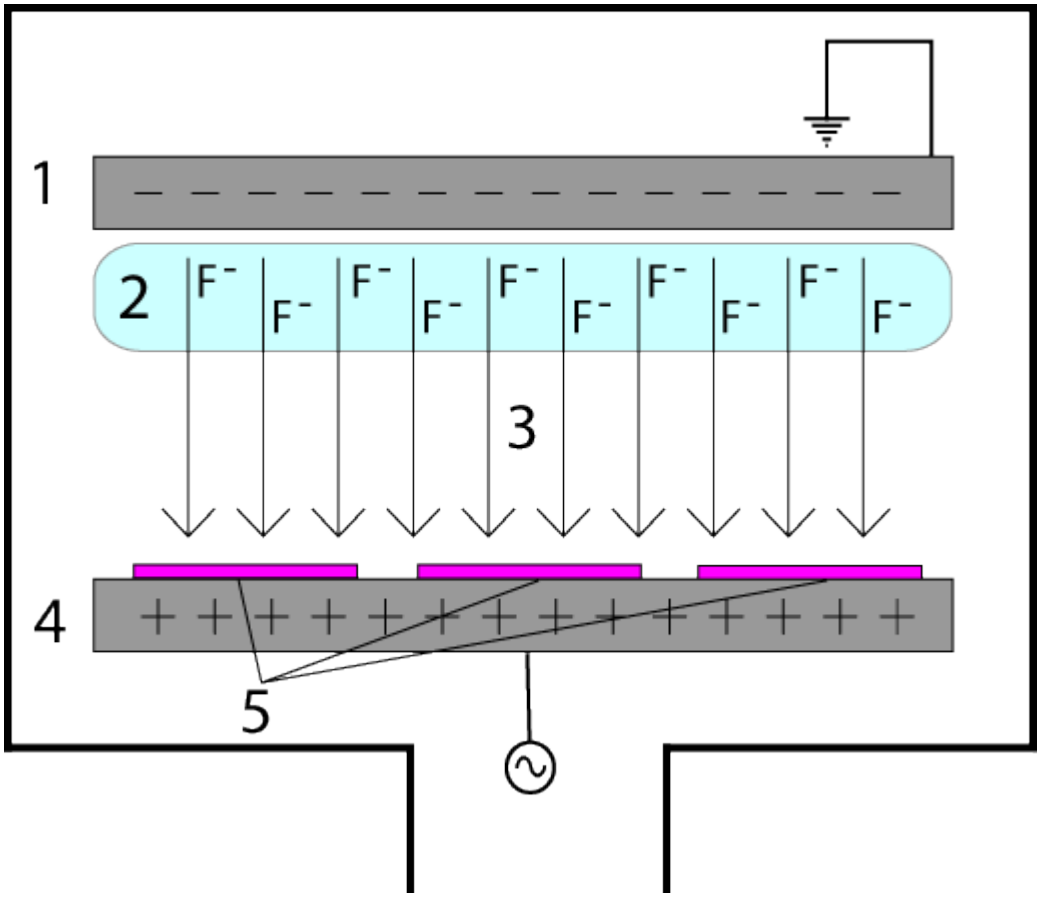
A typical (parallel plate) RIE system consists of a cylindrical vacuum chamber, with a wafer platter situated in the bottom portion of the chamber. The wafer platter is electrically isolated from the rest of the chamber, which is usually grounded. Gas enters through small inlets in the top of the chamber, and exits to the vacuum pump system through the bottom. The types and amount of gas used vary depending upon the etch process; for instance, sulfur hexafluoride is commonly used for etching silicon. Gas pressure is typically maintained in a range between a few millitorr and a few hundred millitorr by adjusting gas flow rates and/or adjusting an exhaust orifice.

Other types of RIE systems exist, including inductively coupled plasma (ICP) RIE. In this type of system, the plasma is generated with an RF powered magnetic field. Very high plasma densities can be achieved, though etch profiles tend to be more isotropic.

A combination of parallel plate and inductively coupled plasma RIE is possible. In this system, the ICP is employed as a high density source of ions which increases the etch

rate, whereas a separate RF bias is applied to the substrate (silicon wafer) to create directional electric fields near the substrate to achieve more anisotropic etch profiles.

Method of operation



A diagram of a common RIE setup. An RIE consists of two electrodes (1 and 4) that create an electric field (3) meant to accelerate ions (2) toward the surface of the samples (5).

Plasma is initiated in the system by applying a strong RF (radio frequency) electromagnetic field to the wafer platter. The field is typically set to a frequency of 13.56 megahertz, applied at a few hundred watts. The oscillating electric field ionizes the gas molecules by stripping them of electrons, creating a plasma.

In each cycle of the field, the electrons are electrically accelerated up and down in the chamber, sometimes striking both the upper wall of the chamber and the wafer platter. At the same time, the much more massive ions move relatively little in response to the RF

electric field. When electrons are absorbed into the chamber walls they are simply fed out to ground and do not alter the electronic state of the system. However, electrons absorbed into the wafer platter cause the platter to build up charge due to its DC isolation. This charge build up develops a large negative voltage on the platter, typically around a few hundred volts. The plasma itself develops a slightly positive charge due to the higher concentration of positive ions compared to free electrons.

Because of the large voltage difference, positive ions tend to drift toward the wafer platter, where they collide with the samples to be etched. The ions react chemically with the materials on the surface of the samples, but can also knock off (sputter) some material by transferring some of their kinetic energy. Due to the mostly vertical delivery of reactive ions, reactive ion etching can produce very anisotropic etch profiles, which contrast with the typically isotropic profiles of wet chemical etching.

Etch conditions in an RIE system depend strongly on the many process parameters, such as pressure, gas flows, and RF power. A modified version of RIE is deep reactive-ion etching, used to excavate deep features.

Ion beam lithography

By analogy to E-beam lithography, focused ion beam lithography scans an ion beam across a surface to form a pattern. The ion beam may be used for directly sputtering the surface, or may induce chemical reactions in the exposed top layer (resist). In the case of direct sputtering, redeposition is a common occurrence, which will affect the final surface profile. In the case of resist exposure, the slowing down of secondary electrons is the basis for forming the final image, which therefore limit the resolution of acceptably printed features to about 30 nm, even though the beam spot size can be smaller. The ion beam may also either be focused through a (stencil) mask or directly on the surface without a mask. The masked approach causes erosion of the openings by the ion beam through the inherent sputterning process. But this approach has been found to be useful for transferring high-fidelity patterns on non-planar surfaces.

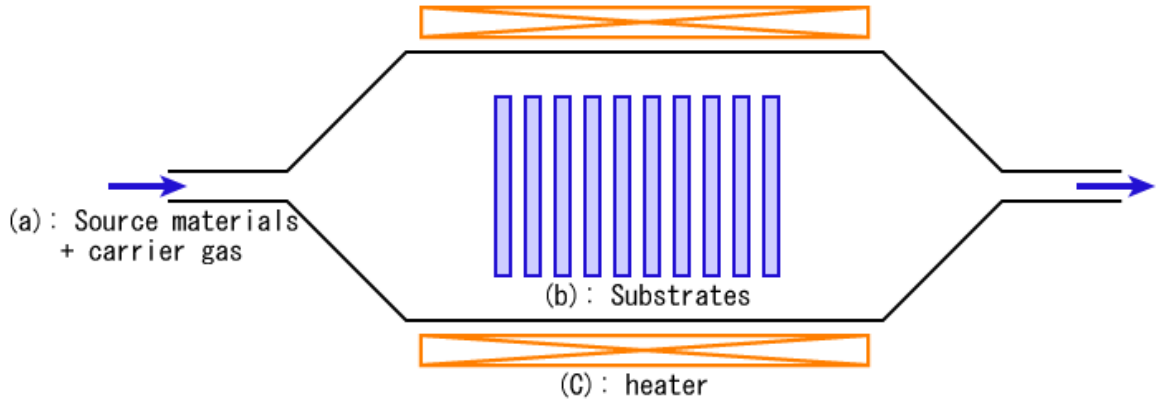
Chemical vapor deposition

DC plasma (violet) enhances the growth of carbon nanotubes in this laboratory-scale PECVD apparatus.

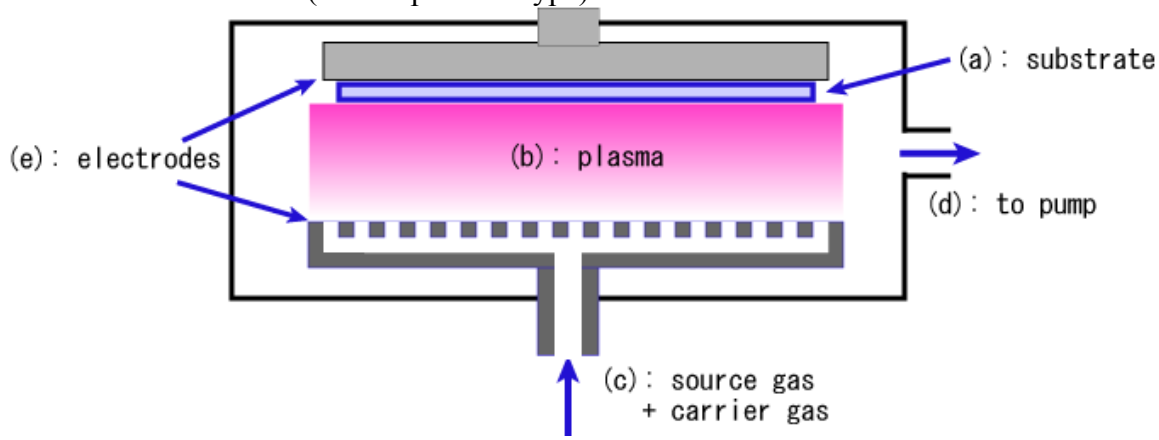
Chemical vapor deposition (CVD) is a chemical process used to produce high-purity, high-performance solid materials. The process is often used in the semiconductor industry to produce thin films. In a typical CVD process, the wafer (substrate) is exposed to one or more volatile precursors, which react and/or decompose on the substrate surface to produce the desired deposit. Frequently, volatile by-products are also produced, which are removed by gas flow through the reaction chamber.

Microfabrication processes widely use CVD to deposit materials in various forms, including: monocrystalline, polycrystalline, amorphous, and epitaxial. These materials include: silicon, carbon fiber, carbon nanofibers, filaments, carbon nanotubes, SiO_2, silicon-germanium, tungsten, silicon carbide, silicon nitride, silicon oxynitride, titanium nitride, and various high-k dielectrics. The CVD process is also used to produce synthetic diamonds.

Types of chemical vapor deposition



Hot-wall thermal CVD (batch operation type)



Plasma assisted CVD

A number of forms of CVD are in wide use and are frequently referenced in the literature. These processes differ in the means by which chemical reactions are initiated (e.g., activation process) and process conditions.

- Classified by operating pressure
  - *Atmospheric pressure CVD* (APCVD) - CVD processes at atmospheric pressure.
  - *Low-pressure CVD* (LPCVD) - CVD processes at subatmospheric pressures. Reduced pressures tend to reduce unwanted gas-phase reactions

and improve film uniformity across the wafer. Most modern CVD processes are either LPCVD or UHVCVD.
    - *Ultrahigh vacuum CVD* (UHVCVD) - CVD processes at a very low pressure, typically below 10^{-6} Pa (~10^{-8} torr). Note that in other fields, a lower division between high and ultra-high vacuum is common, often 10^{-7} Pa.
- Classified by physical characteristics of vapor
    - *Aerosol assisted CVD* (AACVD) - A CVD process in which the precursors are transported to the substrate by means of a liquid/gas aerosol, which can be generated ultrasonically. This technique is suitable for use with non-volatile precursors.
    - *Direct liquid injection CVD* (DLICVD) - A CVD process in which the precursors are in liquid form (liquid or solid dissolved in a convenient solvent). Liquid solutions are injected in a vaporization chamber towards injectors (typically car injectors). Then the precursor vapors are transported to the substrate as in classical CVD process. This technique is suitable for use on liquid or solid precursors. High growth rates can be reached using this technique.
- Plasma methods
    - *Microwave plasma-assisted CVD* (MPCVD)
    - *Plasma-Enhanced CVD* (PECVD) - CVD processes that utilize plasma to enhance chemical reaction rates of the precursors. PECVD processing allows deposition at lower temperatures, which is often critical in the manufacture of semiconductors.
    - *Remote plasma-enhanced CVD* (RPECVD) - Similar to PECVD except that the wafer substrate is not directly in the plasma discharge region. Removing the wafer from the plasma region allows processing temperatures down to room temperature.
- *Atomic layer CVD* (ALCVD) – Deposits successive layers of different substances to produce layered, crystalline films.
- *Combustion Chemical Vapor Deposition* (CCVD) - nGimat's proprietary Combustion Chemical Vapor Deposition process is an open-atmosphere, flame-based technique for depositing high-quality thin films and nanomaterials.
- *Hot wire CVD* (HWCVD) - also known as catalytic CVD (Cat-CVD) or hot filament CVD (HFCVD). Uses a hot filament to chemically decompose the source gases.
- *Metalorganic chemical vapor deposition* (MOCVD) - CVD processes based on metalorganic precursors.
- *Hybrid Physical-Chemical Vapor Deposition* (HPCVD) - Vapor deposition processes that involve both chemical decomposition of precursor gas and vaporization of a solid source.
- *Rapid thermal CVD* (RTCVD) - CVD processes that use heating lamps or other methods to rapidly heat the wafer substrate. Heating only the substrate rather than the gas or chamber walls helps reduce unwanted gas phase reactions that can lead to particle formation.
- *Vapor phase epitaxy* (VPE)

Substances commonly deposited for ICs

Here we, discusses the CVD processes often used for integrated circuits (ICs). Particular materials are deposited best under particular conditions.

Polysilicon

Polycrystalline silicon is deposited from silane (SiH_4), using the following reaction:

$$SiH_4 \rightarrow Si + 2\ H_2$$

This reaction is usually performed in LPCVD systems, with either pure silane feedstock, or a solution of silane with 70-80% nitrogen. Temperatures between 600 and 650 °C and pressures between 25 and 150 Pa yield a growth rate between 10 and 20 nm per minute. An alternative process uses a hydrogen-based solution. The hydrogen reduces the growth rate, but the temperature is raised to 850 or even 1050 °C to compensate.

Polysilicon may be grown directly with doping, if gases such as phosphine, arsine or diborane are added to the CVD chamber. Diborane increases the growth rate, but arsine and phosphine decrease it.

Silicon dioxide

Silicon dioxide (usually called simply "oxide" in the semiconductor industry) may be deposited by several different processes. Common source gases include silane and oxygen, dichlorosilane ($SiCl_2H_2$) and nitrous oxide (N_2O), or tetraethylorthosilicate (TEOS; $Si(OC_2H_5)_4$). The reactions are as follows:

$$SiH_4 + O_2 \rightarrow SiO_2 + 2\ H_2$$
$$SiCl_2H_2 + 2\ N_2O \rightarrow SiO_2 + 2\ N_2 + 2\ HCl$$
$$Si(OC_2H_5)_4 \rightarrow SiO_2 + \text{byproducts}$$

The choice of source gas depends on the thermal stability of the substrate; for instance, aluminium is sensitive to high temperature. Silane deposits between 300 and 500 °C, dichlorosilane at around 900 °C, and TEOS between 650 and 750 °C, resulting in a layer of *low- temperature oxide* (LTO). However, silane produces a lower-quality oxide than the other methods (lower dielectric strength, for instance), and it deposits nonconformally. Any of these reactions may be used in LPCVD, but the silane reaction is also done in APCVD. CVD oxide invariably has lower quality than thermal oxide, but thermal oxidation can only be used in the earliest stages of IC manufacturing.

Oxide may also be grown with impurities (alloying or "doping"). This may have two purposes. During further process steps that occur at high temperature, the impurities may diffuse from the oxide into adjacent layers (most notably silicon) and dope them. Oxides containing 5–15% impurities by mass are often used for this purpose. In addition, silicon dioxide alloyed with phosphorus pentoxide ("P-glass") can be used to smooth out uneven

surfaces. P-glass softens and reflows at temperatures above 1000 °C. This process requires a phosphorus concentration of at least 6%, but concentrations above 8% can corrode aluminium. Phosphorus is deposited from phosphine gas and oxygen:

$$4\ PH_3 + 5\ O_2 \rightarrow 2\ P_2O_5 + 6\ H_2$$

Glasses containing both boron and phosphorus (borophosphosilicate glass, BPSG) undergo viscous flow at lower temperatures; around 850 °C is achievable with glasses containing around 5 weight % of both constituents, but stability in air can be difficult to achieve. Phosphorus oxide in high concentrations interacts with ambient moisture to produce phosphoric acid. Crystals of BPO_4 can also precipitate from the flowing glass on cooling; these crystals are not readily etched in the standard reactive plasmas used to pattern oxides, and will result in circuit defects in integrated circuit manufacturing.

Besides these intentional impurities, CVD oxide may contain byproducts of the deposition process. TEOS produces a relatively pure oxide, whereas silane introduces hydrogen impurities, and dichlorosilane introduces chlorine.

Lower temperature deposition of silicon dioxide and doped glasses from TEOS using ozone rather than oxygen has also been explored (350 to 500 °C). Ozone glasses have excellent conformality but tend to be hygroscopic – that is, they absorb water from the air due to the incorporation of silanol (Si-OH) in the glass. Infrared spectroscopy and mechanical strain as a function of temperature are valuable diagnostic tools for diagnosing such problems.

Silicon nitride

Silicon nitride is often used as an insulator and chemical barrier in manufacturing ICs. The following two reactions deposit nitride from the gas phase:

$$3\ SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12\ H_2$$
$$3\ SiCl_2H_2 + 4\ NH_3 \rightarrow Si_3N_4 + 6\ HCl + 6\ H_2$$

Silicon nitride deposited by LPCVD contains up to 8% hydrogen. It also experiences strong tensile stress, which may crack films thicker than 200 nm. However, it has higher resistivity and dielectric strength than most insulators commonly available in microfabrication (10^{16} Ω·cm and 10 MV/cm, respectively).

Another two reactions may be used in plasma to deposit SiNH:

$$2\ SiH_4 + N_2 \rightarrow 2\ SiNH + 3\ H_2$$
$$SiH_4 + NH_3 \rightarrow SiNH + 3\ H_2$$

These films have much less tensile stress, but worse electrical properties (resistivity 10^6 to 10^{15} Ω·cm, and dielectric strength 1 to 5 MV/cm).

Metals

Some metals (notably aluminium and copper) are seldom or never deposited by CVD. As of 2010, a commercially, cost effective, viable CVD process for copper did not exist, though copper formate, copper(hfac)2, Cu(II) ethyl acetoacetate, and other precursors have been used. Copper deposition of the metal has been done mostly by electroplating, in order to reduce the cost. Aluminum can be deposited from tri-isobutyl aluminium (TIBAL), tri ethyl/methyl aluminum (TEA,TMA), or dimethylaluminum hydride (DMAH), but physical vapor deposition methods are usually preferred.

However, CVD processes for molybdenum, tantalum, titanium, nickel, and tungsten are widely used. These metals can form useful silicides when deposited onto silicon. Mo, Ta and Ti are deposited by LPCVD, from their pentachlorides. Nickel, molybdenum, and tungsten can be deposited at low temperatures from their carbonyl precursors. In general, for an arbitrary metal *M*, the reaction is as follows:

$$2\ MCl_5 + 5\ H_2 \rightarrow 2\ M + 10\ HCl$$

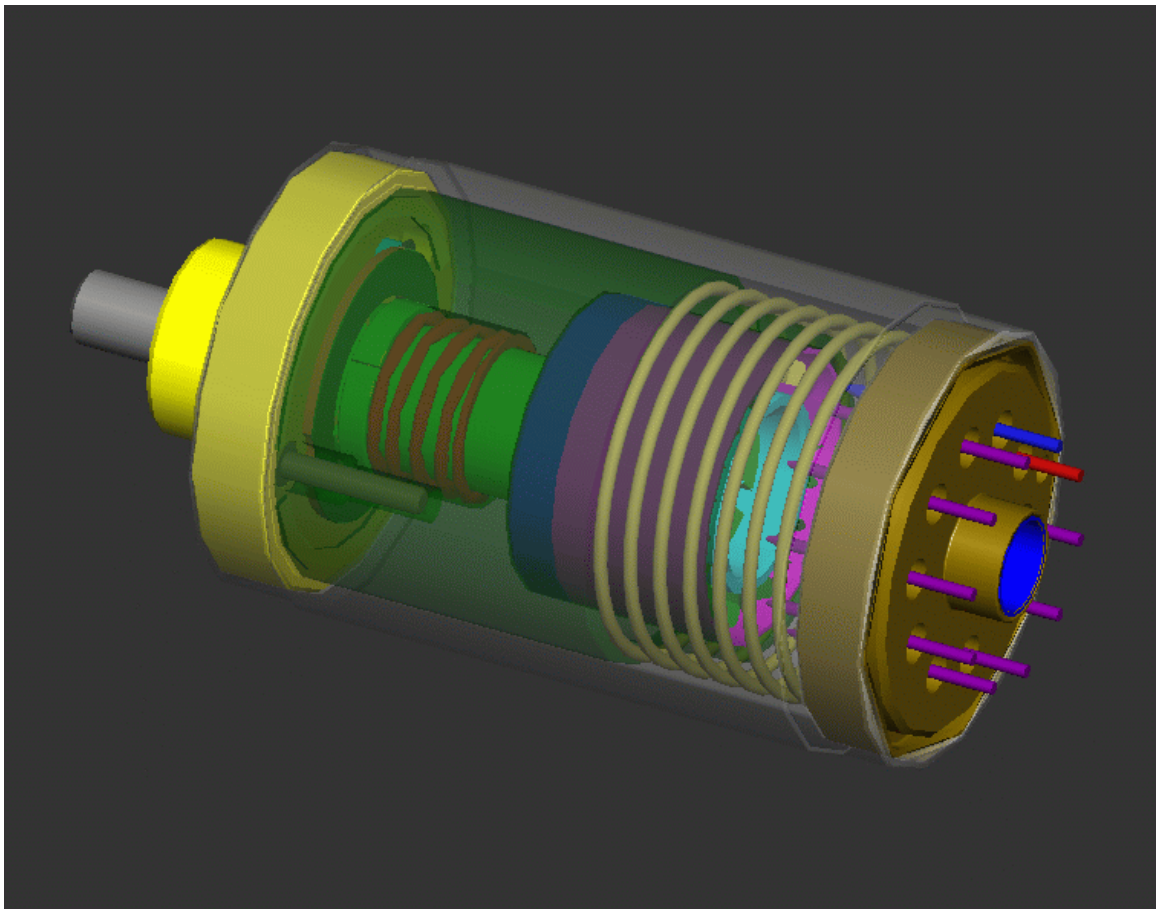
The usual source for tungsten is tungsten hexafluoride, which may be deposited in two ways:

$$WF_6 \rightarrow W + 3\ F_2$$
$$WF_6 + 3\ H_2 \rightarrow W + 6\ HF$$

Chapter- 12

Applications of Microelectromechanical Systems

Accelerometer



A depiction of an accelerometer designed at Sandia National Laboratories.

An **accelerometer** is a device that measures proper acceleration.

Single- and multi-axis models are available to detect magnitude and direction of the acceleration as a vector quantity, and can be used to sense orientation, acceleration, vibration shock, and falling. Micromachined accelerometers are increasingly present in portable electronic devices and video game controllers, to detect the position of the device or provide for game input.

Physical principles

An accelerometer measures proper acceleration, which is the acceleration it experiences relative to freefall and is the acceleration felt by people and objects. Put another way, at any point in spacetime the equivalence principle guarantees the existence of a local inertial frame, and an accelerometer measures the acceleration relative to that frame. Such accelerations are popularly measured in terms of g-force.

An accelerometer at rest relative to the Earth's surface will indicate approximately 1 g *upwards*, because any point on the Earth's surface is accelerating upwards relative to the local inertial frame (the frame of a freely falling object near the surface). To obtain the acceleration due to motion with respect to the Earth, this "gravity offset" must be subtracted and corrections for effects caused by the Earth's rotation relative to the inertial frame.

The reason for the appearance of a gravitational offset is Einstein's equivalence principle, which states that the effects of gravity on an object are indistinguishable from acceleration. When held fixed in a gravitational field by, for example, applying a ground reaction force or an equivalent upward thrust, the reference frame for an accelerometer (its own casing) accelerates upwards with respect to a free-falling reference frame. The effects of this acceleration are indistinguishable from any other acceleration experienced by the instrument, so that an accelerometer cannot detect the difference between sitting in a rocket on the launch pad, and being in the same rocket in deep space while it uses its engines to accelerate at 1 g. For similar reasons, an accelerometer will read *zero* during any type of free fall. This includes use in a coasting spaceship in deep space far from any mass, a spaceship orbiting the Earth, an airplane in a parabolic "zero-g" arc, or any free-fall in vacuum. Another example is free-fall at a sufficiently high altitude that atmospheric effects can be neglected.

However this does not include a (non-free) fall in which air resistance produces drag forces that reduce the acceleration, until constant terminal velocity is reached. At terminal velocity the accelerometer will indicate 1 g acceleration upwards. For the same reason a skydiver, upon reaching terminal velocity, does not feel as though he or she were in "free-fall", but rather experiences a feeling similar to being supported (at 1 g) on a "bed" of uprushing air.

Acceleration is quantified in the SI unit metres per second per second (m/s^2), in the cgs unit gal (Gal), or popularly in terms of g-force (*g*).

For the practical purpose of finding the acceleration of objects with respect to the Earth, such as for use in an inertial navigation system, a knowledge of local gravity is required. This can be obtained either by calibrating the device at rest, or from a known model of gravity at the approximate current position.

Structure

Conceptually, an accelerometer behaves as a damped mass on a spring. When the accelerometer experiences an acceleration, the mass is displaced to the point that the spring is able to accelerate the mass at the same rate as the casing. The displacement is then measured to give the acceleration.

In commercial devices, piezoelectric, piezoresistive and capacitive components are commonly used to convert the mechanical motion into an electrical signal. Piezoelectric accelerometers rely on piezoceramics (e.g. lead zirconate titanate) or single crystals (e.g. quartz, tourmaline). They are unmatched in terms of their upper frequency range, low packaged weight and high temperature range. Piezoresistive accelerometers are preferred in high shock applications. Capacitive accelerometers typically use a silicon micro-machined sensing element. Their performance is superior in the low frequency range and they can be operated in servo mode to achieve high stability and linearity.

Modern accelerometers are often small *micro electro-mechanical systems* (MEMS), and are indeed the simplest MEMS devices possible, consisting of little more than a cantilever beam with a proof mass (also known as seismic mass). Damping results from the residual gas sealed in the device. As long as the Q-factor is not too low, damping does not result in a lower sensitivity.

Under the influence of external accelerations the proof mass deflects from its neutral position. This deflection is measured in an analog or digital manner. Most commonly, the capacitance between a set of fixed beams and a set of beams attached to the proof mass is measured. This method is simple, reliable, and inexpensive. Integrating piezoresistors in the springs to detect spring deformation, and thus deflection, is a good alternative, although a few more process steps are needed during the fabrication sequence. For very high sensitivities quantum tunneling is also used; this requires a dedicated process making it very expensive. Optical measurement has been demonstrated on laboratory scale.

Another, far less common, type of MEMS-based accelerometer contains a small heater at the bottom of a very small dome, which heats the air inside the dome to cause it to rise. A thermocouple on the dome determines where the heated air reaches the dome and the deflection off the center is a measure of the acceleration applied to the sensor.

Most micromechanical accelerometers operate *in-plane*, that is, they are designed to be sensitive only to a direction in the plane of the die. By integrating two devices perpendicularly on a single die a two-axis accelerometer can be made. By adding an additional *out-of-plane* device three axes can be measured. Such a combination always

has a much lower misalignment error than three discrete models combined after packaging.

Micromechanical accelerometers are available in a wide variety of measuring ranges, reaching up to thousands of *g*′s. The designer must make a compromise between sensitivity and the maximum acceleration that can be measured.

Applications

Engineering

Accelerometers can be used to measure vehicle acceleration. They allow for performance evaluation of both the engine/drive train and the braking systems. Useful numbers like 0-60 mph, 60-0 mph and 1/4 mile times can all be found using accelerometers.

Accelerometers can be used to measure vibration on cars, machines, buildings, process control systems and safety installations. They can also be used to measure seismic activity, inclination, machine vibration, dynamic distance and speed with or without the influence of gravity. Applications for accelerometers that measure gravity, wherein an accelerometer is specifically configured for use in gravimetry, are called gravimeters.

Notebook computers equipped with accelerometers can contribute to the *Quake-Catcher Network* (QCN), a BOINC project aimed at scientific research of earthquakes.

Biology

Accelerometers are also increasingly used in the Biological Sciences. High frequency recordings of bi-axial or tri-axial acceleration (>10 Hz) allows the discrimination of behavioral patterns while animals are out of sight. Furthermore, recordings of acceleration allow researchers to quantify the rate at which an animal is expending energy in the wild, by either determination of limb-stroke frequency or measures such as Overall Dynamic Body Acceleration Such approaches have mostly been adopted by marine scientists due to an inability to study animals in the wild using visual observations, however an increasing number of terrestrial biologists are adopting similar approaches. This device can be connected to an amplifier to amplify the signal.

Industry - Machinery Health Monitoring

Accelerometers are also used for machinery health monitoring of rotating equipment such as pumps, fans, rollers, compressors, and cooling towers,. Vibration monitoring programs are proven to save money, reduce downtime, and improve safety in plants worldwide by detecting conditions such as shaft misalignment, rotor imbalance, gear failure or bearing fault which can lead to costly repairs. Accelerometer vibration data allows the user to monitor machines and detect these faults before the rotating equipment fails. Vibration monitoring programs are utilized in industries such as automotive manufacturing, machine tool applications, pharmaceutical production, power generation and power

plants, pulp and paper, food and beverage production, water and wastewater, hydropower, petrochemical and steel manufacturing.

Building and structural monitoring

Accelerometers are used to measure the motion and vibration of a structure that is exposed to dynamic loads. Dynamic loads originate from a variety of sources including:

- Human activities - walking, running, dancing or skipping
- Working machines - inside a building or in the surrounding area
- Construction work - driving piles, demolition, drilling and excavating
- Moving loads on bridges
- Vehicle collisions
- Impact loads - falling debris
- Concussion loads - internal and external explosions
- Collapse of structural elements
- Wind loads and wind gusts
- Air blast pressure
- Loss of support because of ground failure
- Earthquakes and aftershocks

Measuring and recording how a structure responds to these inputs is critical for assessing the safety and viability of a structure. This type of monitoring is called Dynamic Monitoring.

Medical applications

Zoll's AED Plus uses CPR-D•padz which contain an accelerometer to measure the depth of CPR chest compressions.

Within the last several years, Nike, Polar and other companies have produced and marketed sports watches for runners that include footpods, containing accelerometers to help determine the speed and distance for the runner wearing the unit.

In Belgium, accelerometer-based step counters are promoted by the government to encourage people to walk a few thousand steps each day.

Herman Digital Trainer uses accelerometers to measure strike force in physical training.

Navigation

An **Inertial Navigation System** (INS) is a navigation aid that uses a computer and motion sensors (accelerometers) to continuously calculate via dead reckoning the position, orientation, and velocity (direction and speed of movement) of a moving object without the need for external references. Other terms used to refer to inertial navigation

systems or closely related devices include **inertial guidance system**, **inertial reference platform**, and many other variations.

An accelerometer alone is unsuitable to determine changes in altitude over distances where the vertical decrease of gravity is significant, such as for aircraft and rockets. In the presence of a gravitational gradient, the calibration and data reduction process is numerically unstable.

Transport

Accelerometers are used to detect apogee in both professional and in amateur rocketry.

Accelerometers are also being used in Intelligent Compaction rollers. Accelerometers are used alongside gyroscopes in inertial guidance systems.

One of the most common uses for MEMS accelerometers is in airbag deployment systems for modern automobiles. In this case the accelerometers are used to detect the rapid negative acceleration of the vehicle to determine when a collision has occurred and the severity of the collision. Another common automotive use is in electronic stability control systems, which use a lateral accelerometer to measure cornering forces. The widespread use of accelerometers in the automotive industry has pushed their cost down dramatically. Another automotive application is the monitoring of noise, vibration and harshness (NVH), conditions that cause discomfort for drivers and passengers and may also be indicators of mechanical faults.

Tilting trains use accelerometers and gyroscopes to calculate the required tilt.

Vulcanology

Modern electronic accelerometers are used in remote sensing devices intended for the monitoring of active volcanos to detect the motion of magma

Consumer electronics

Accelerometers are increasingly being incorporated into personal electronic devices.

Motion input

Some smartphones, digital audio players and personal digital assistants contain accelerometers for user interface control; often the accelerometer is used to present landscape or portrait views of the device's screen, based on the way the device is being held.

Smartphones can download an Automatic Collision Notification (ACN) app such as My-911, similar to the Onstar AACN service, Ford Link's 911 Assist, Toyota's Safety

Connect, Lexus Link, or BMW Assist. The phone's accelerometer detects crash-strength G-forces and automatically calls for assistance unless manually cancelled.

Nintendo's Wii video game console uses a controller called a Wii Remote that contains a three-axis accelerometer and was designed primarily for motion input. Users also have the option of buying an additional motion-sensitive attachment, the Nunchuk, so that motion input could be recorded from both of the user's hands independently.

The Sony PlayStation 3 uses the DualShock 3 remote which uses a six-axis accelerometer that can be used to make steering more realistic in racing games, such as Motorstorm and Burnout Paradise.

The Nokia 5500 sport features a 3D accelerometer that can be accessed from software. It is used for step recognition (counting) in a sport application, and for tap gesture recognition in the user interface. Tap gestures can be used for controlling the music player and the sport application, for example to change to next song by tapping through clothing when the device is in a pocket. Other uses for accelerometer in Nokia phones include Pedometer functionality in Nokia Sports Tracker. Some other devices provide the tilt sensing feature with a cheaper component, which is not a true accelerometer.

Sleep phase alarm clocks use accelerometric sensors to detect movement of a sleeper, so that it can wake the person when he/she is not in REM phase, therefore awakes more easily.

Orientation sensing

A number of 21st century devices use accelerometers to align the screen depending on the direction the device is held, i.e. switching between portrait and landscape modes. Such devices include many tablet PCs and some smartphones and digital cameras.

For example, Apple uses an LIS302DL accelerometer in the iPhone, iPod Touch and the 4th&5th generation iPod Nano allowing the device to know when it is tilted on its side. Third-party developers have expanded its use with fanciful applications such as electronic bobbleheads. The BlackBerry Storm phone was also an early user of this orientation sensing feature.

The Nokia N95 and Nokia N82 have accelerometers embedded inside them. It was primarily used as a tilt sensor for tagging the orientation to photos taken with the built-in camera, later thanks to a firmware update it became available to other applications.

As of January 2009, almost all new mobile phones and digital cameras such as Canon PowerShot and Ixus range contain at least a tilt sensor (sometimes an accelerometer) for the purpose of auto image rotation, motion-sensitive mini-games, and to correct shake when taking photographs.

Image stabilization

Camcorders use accelerometers for image stabilization. Still cameras use accelerometers for anti-blur capturing. The camera holds off snapping the CCD "shutter" when the camera is moving. When the camera is still (if only for a millisecond, as could be the case for vibration), the CCD is "snapped". An example application which has used such technology is the Glogger VS2, a phone application which runs on Symbian OS based phone with accelerometer such as Nokia N96. Some digital cameras, contain accelerometers to determine the orientation of the photo being taken and also for rotating the current picture when viewing.

Device integrity

Many laptops feature an accelerometer, such as Lenovo's (formerly IBM's) Active Protection System, Apple's Sudden Motion Sensor and HP's 3D DriveGuard, which is used to detect drops. If a drop is detected, the heads of the hard disk are parked to avoid data loss and possible head or disk damage by the ensuing shock.

Gravimetry

A **gravimeter** or gravitometer, is an instrument used in gravimetry for measuring the local gravitational field. A gravimeter is a type of accelerometer, except that accelerometers are susceptible to all vibrations including noise, that cause oscillatory accelerations. This is counteracted in the gravimeter by integral vibration isolation and signal processing. Though the essential principle of design is the same as in accelerometers, gravimeters are typically designed to be much more sensitive than accelerometers in order to measure very tiny changes within the Earth's gravity, of 1 *g*. In contrast, other accelerometers are often designed to measure 1000 *g* or more, and many perform multi-axial measurements. The constraints on temporal resolution are usually less for gravimeters, so that resolution can be increased by processing the output with a longer "time constant".

Types of accelerometer

- Piezoelectric accelerometer
- Shear mode accelerometer
- Surface micromachined capacitive (MEMS)
- Thermal (submicrometre CMOS process)
- Bulk micromachined capacitive
- Bulk micromachined piezoelectric resistive
- Capacitive spring mass base
- Electromechanical servo (Servo Force Balance)
- Null-balance
- Strain gauge
- Resonance
- Magnetic induction

- Optical
- Surface acoustic wave (SAW)
- Laser accelerometer
- DC response
- High temperature
- Low frequency
- High gravity
- Triaxial
- Modally tuned impact hammers
- Seat pad accelerometers
- Pendulating integrating gyroscopic accelerometer