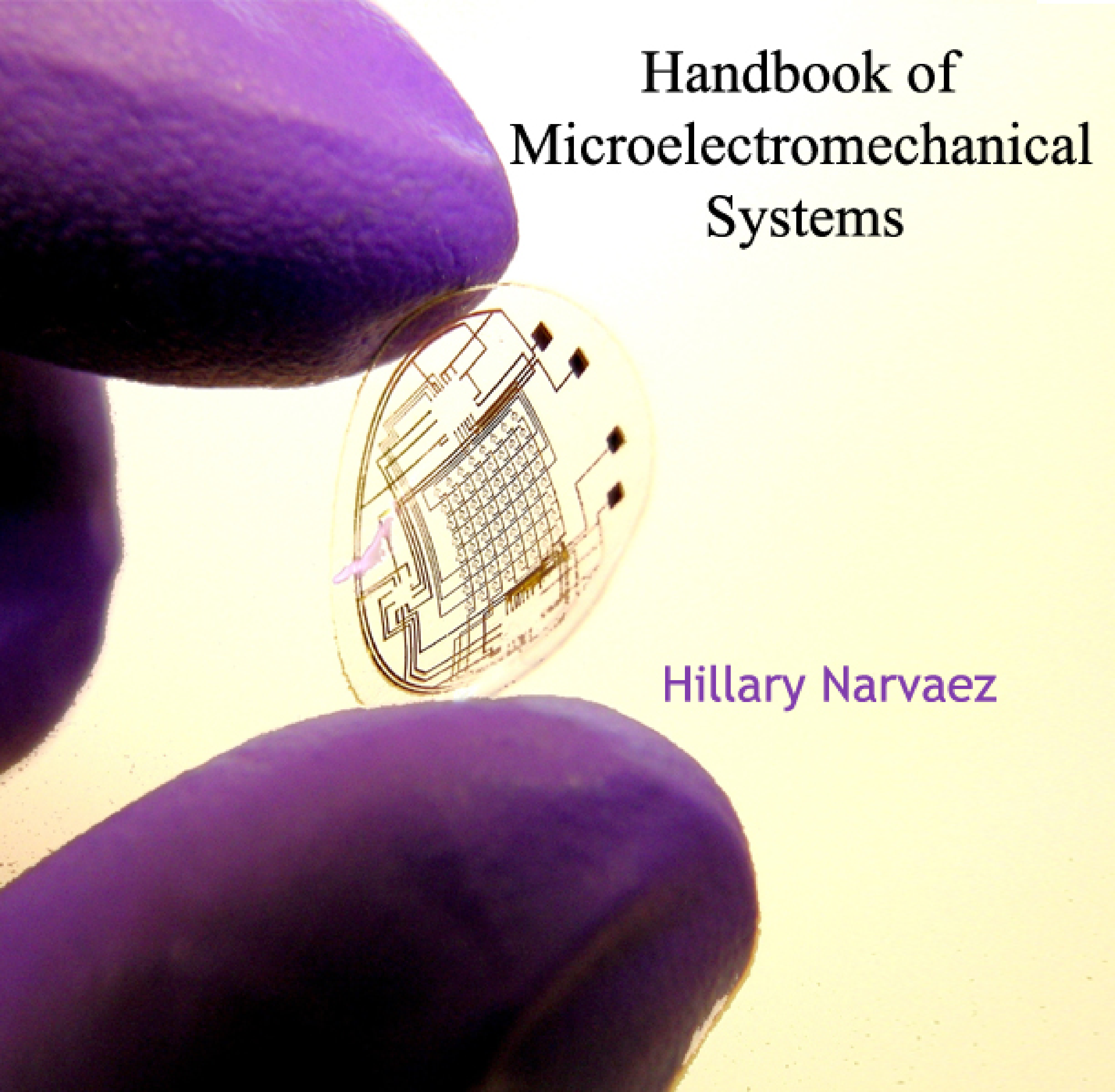


Handbook of Microelectromechanical Systems

Hillary Narvaez

First Edition, 2012

ISBN 978-81-323-3575-7

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

Table of Contents

Chapter- 1

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

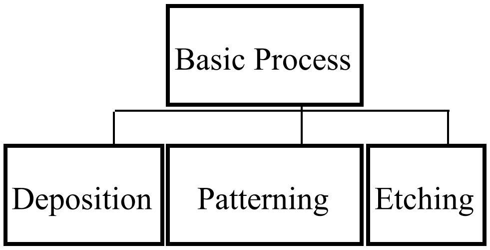
Metals can also be used to create MEMS elements. While metals do not have some of the advantages displayed by silicon in terms of mechanical properties, when used within their limitations, metals can exhibit very high degrees of reliability.

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 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

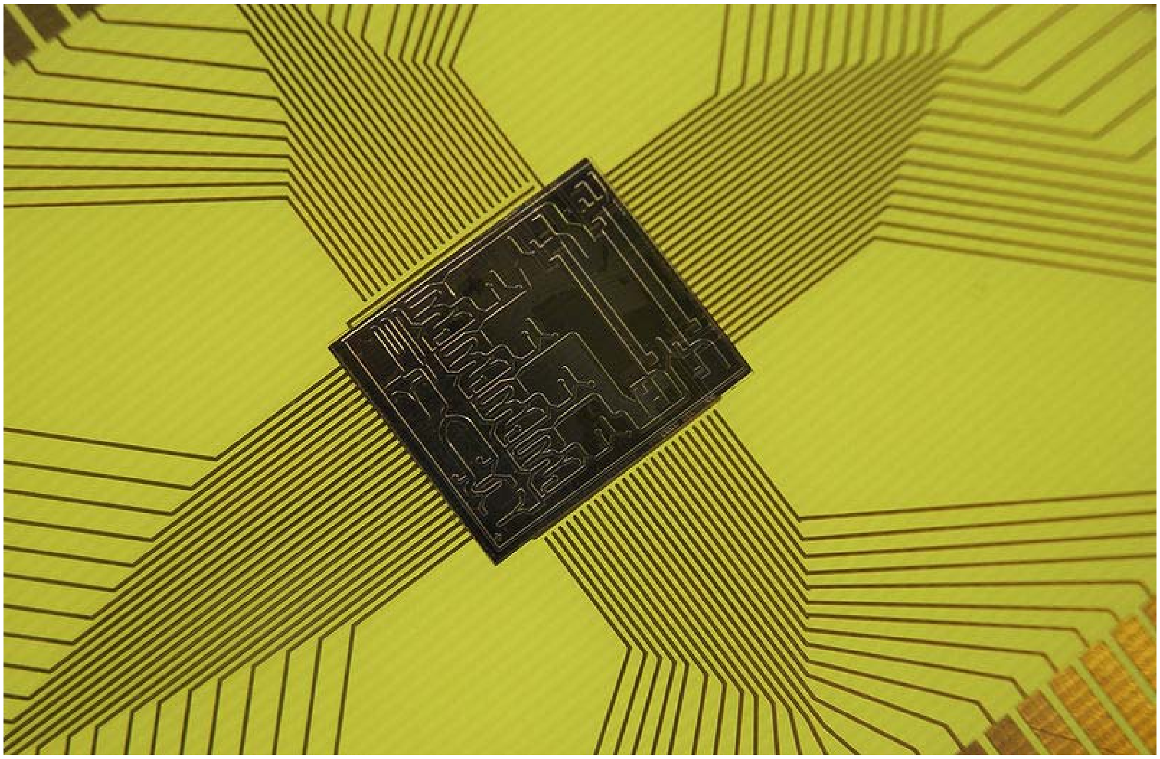
Surface micromachining uses layers deposited on the surface of a substrate as the structural materials, rather than using the substrate itself. Surface micromachining was created in the late 1980s to render micromachining of silicon more compatible with planar integrated circuit technology, with the 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- 2

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.

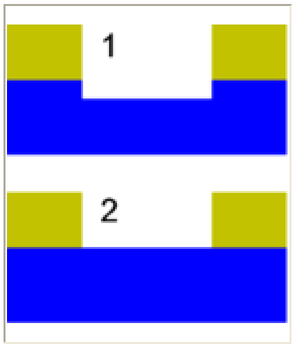
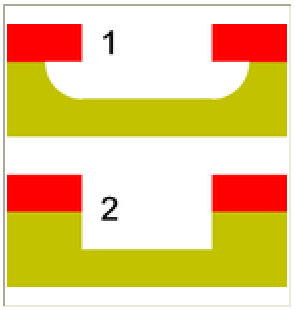
Etching is used in microfabrication to chemically remove layers from the surface of a wafer during manufacturing. Etching is a critically important process module, and every wafer undergoes many etching steps before it is complete.

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

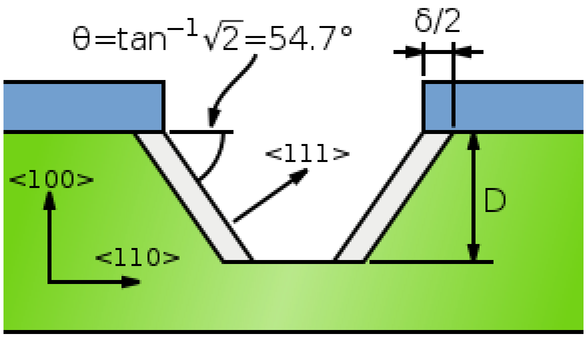
The first etching processes used liquid-phase ("wet") etchants. The wafer can be immersed in a bath of etchant, which must be agitated to achieve good process control. For instance, buffered hydrofluoric acid (BHF) is used commonly to etch silicon dioxide over a silicon substrate.

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.

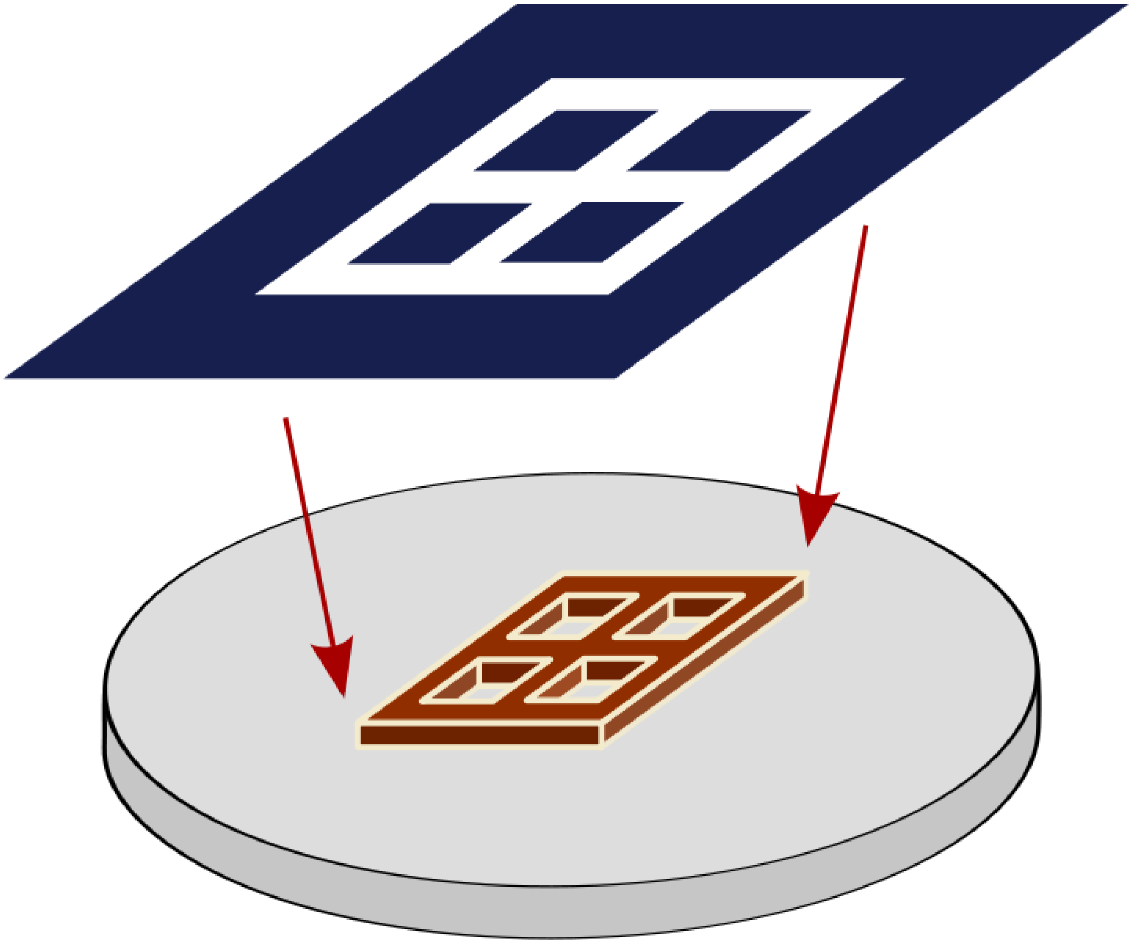
Contact printing is liable to damage both the mask and the wafer, and this was the primary reason it was abandoned for high volume production. Both contact and proximity lithography require the light intensity to be uniform across an entire wafer, and the mask to align precisely to features already on 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).

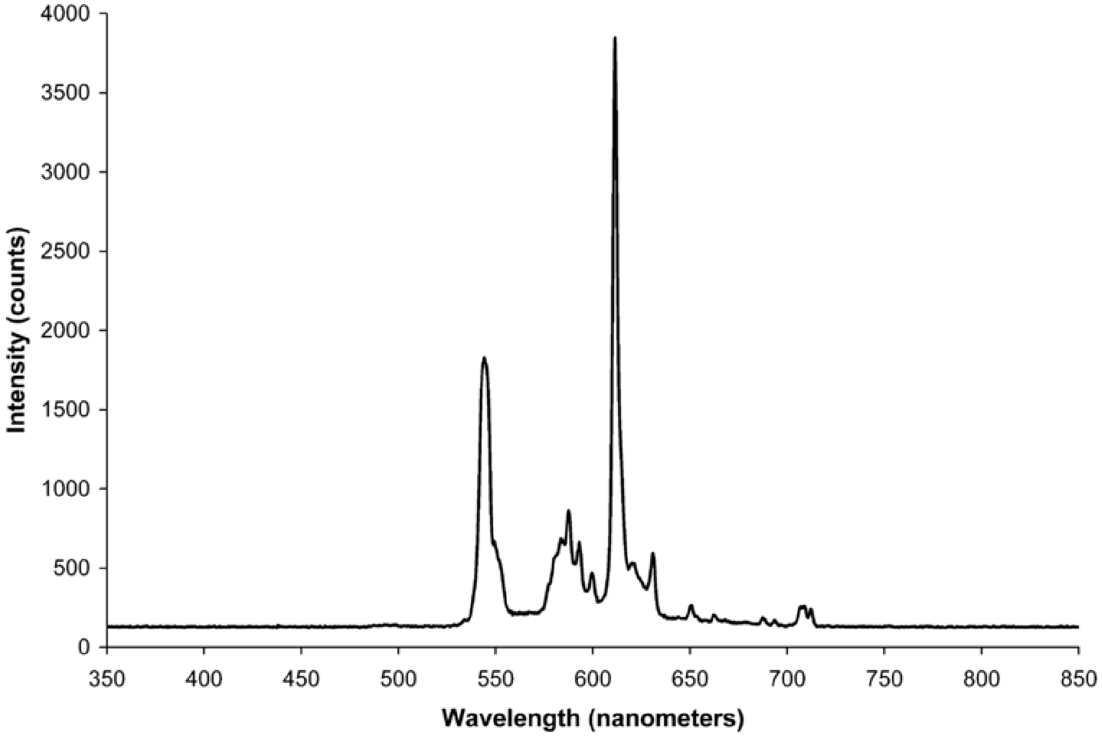
- Dai Nippon Printing Company
- Toppan Photomasks
- Photronics
- Hoya Japan
- Taiwan Mask Corporation
- Compugraphics Photomask Solutions

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.

λ is the wavelength of light used

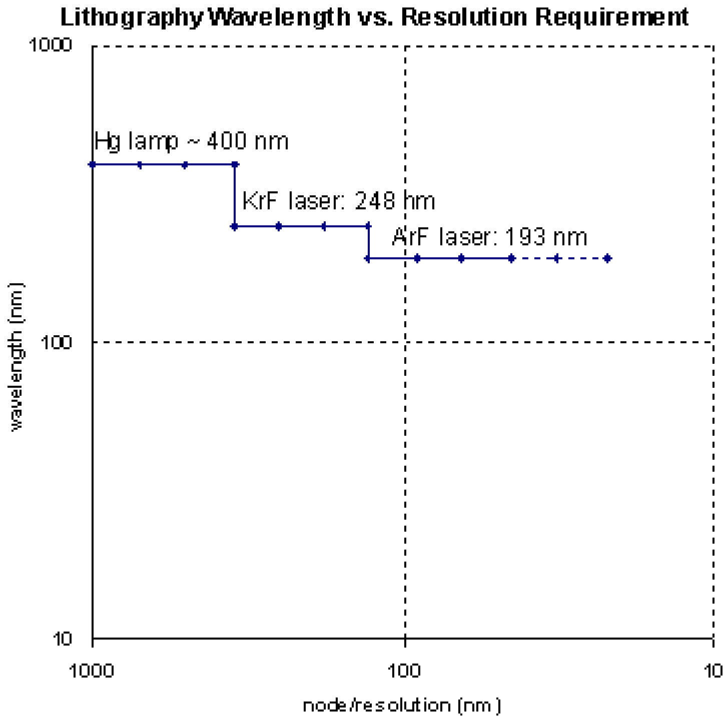
NA is the numerical aperture of the lens as seen from the wafer

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

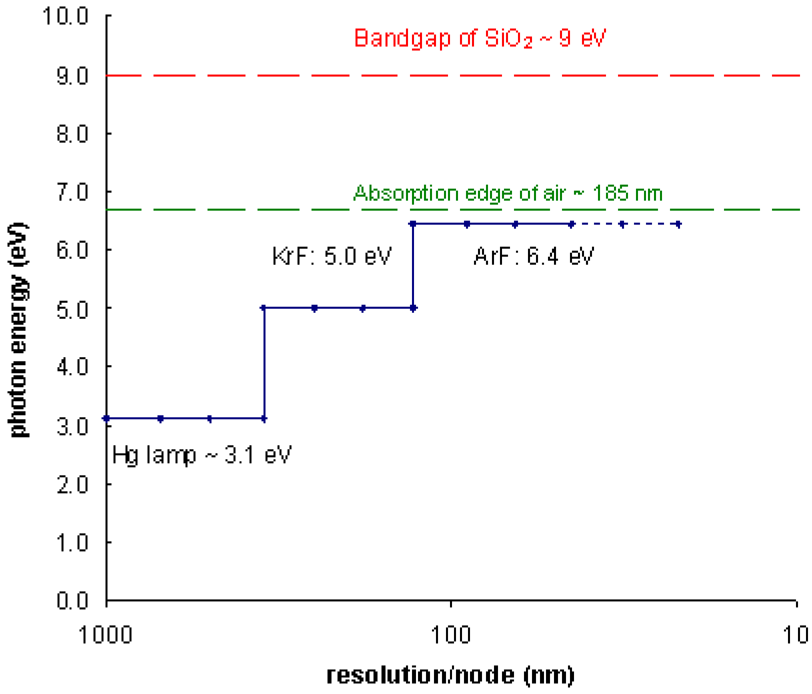


The evolution of lithography wavelength corresponding to different light sources. It is worth noting that the same light source may be used for several technology generations.

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- 3

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. ≥ 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

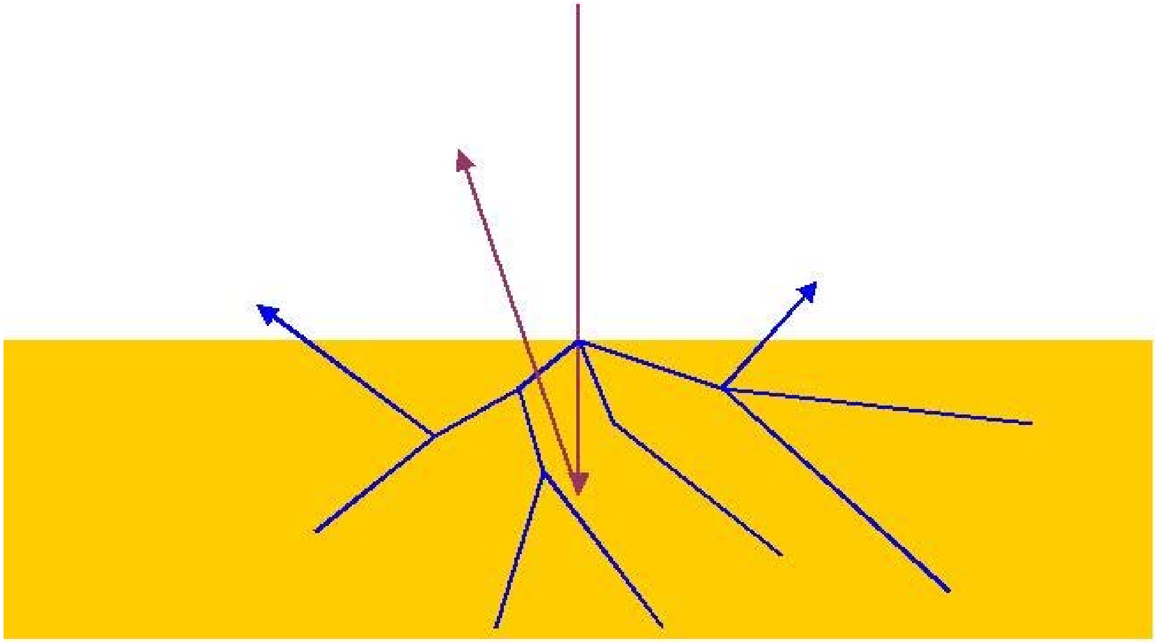
Despite the high resolution of electron-beam lithography, the generation of defects during electron-beam lithography is often not considered by users. Defects may be classified into two categories: data-related defects, and physical defects.

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.

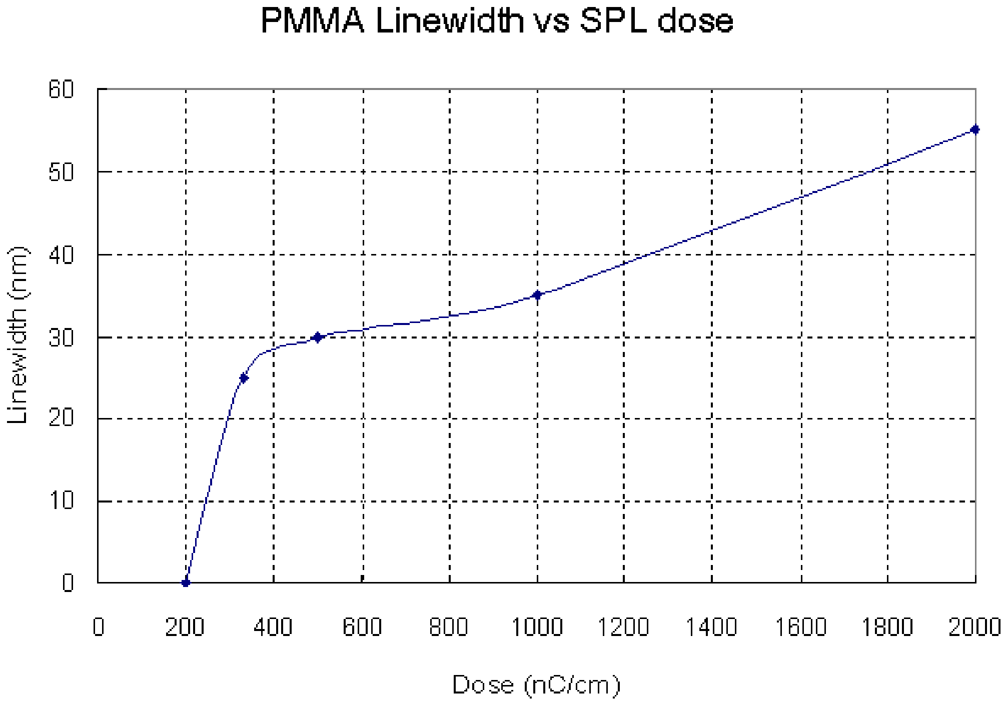
Electron beam resist performance

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- 4

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 .

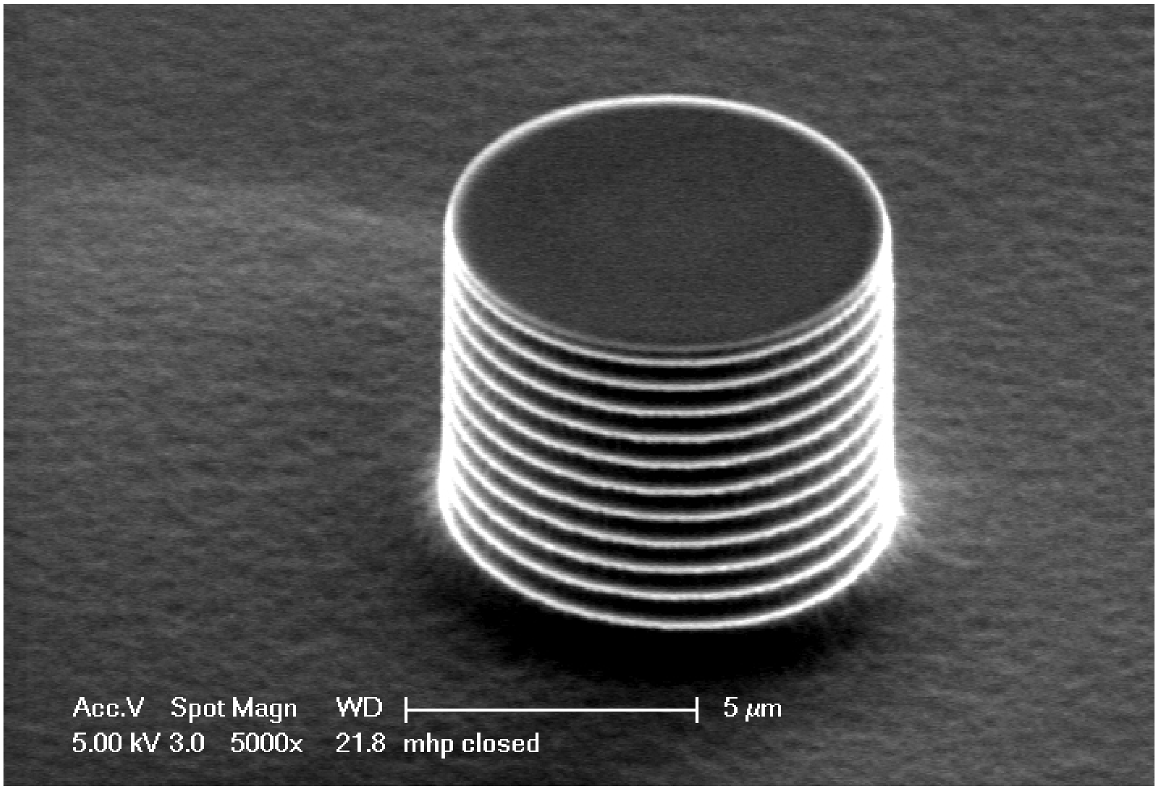
There are two main technologies for high-rate DRIE: cryogenic and Bosch, although the Bosch process is the only recognised production technique. Both Bosch and cryo processes can fabricate 90° (truly vertical) walls, but often the walls are slightly tapered, e.g. 88° or 92° ("retrograde").

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.

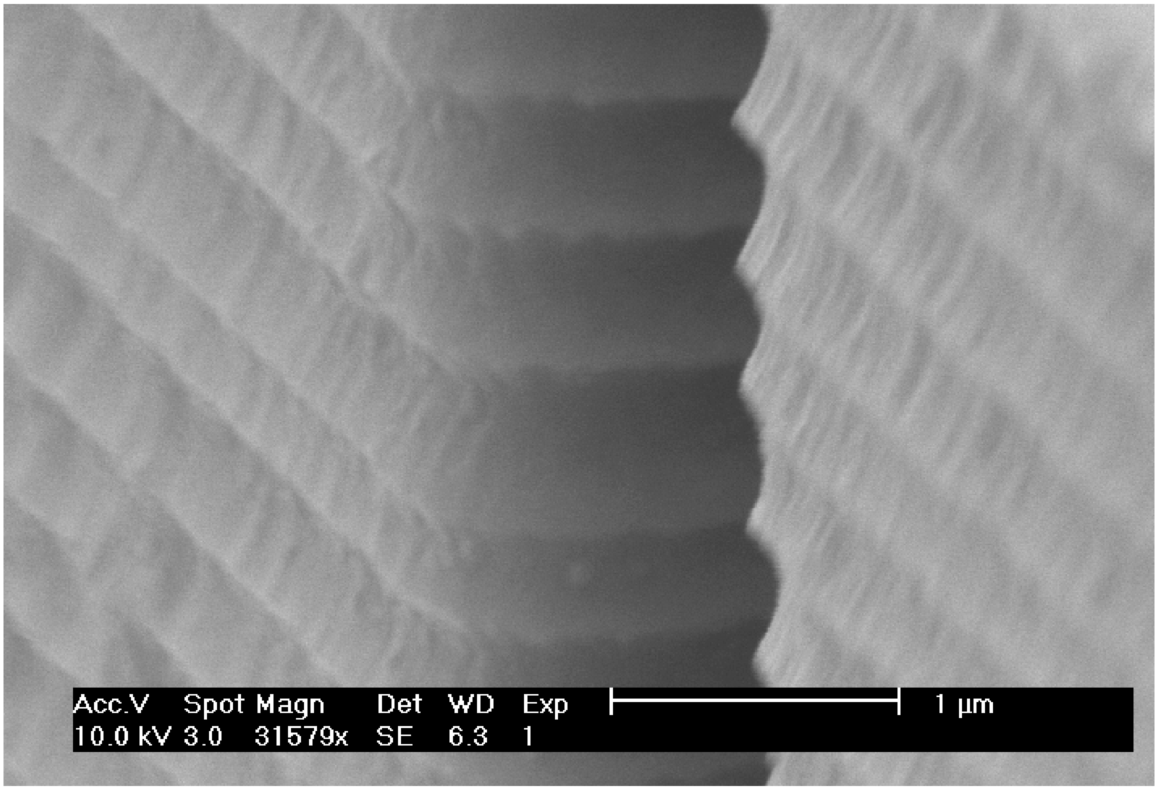
Bosch process



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.

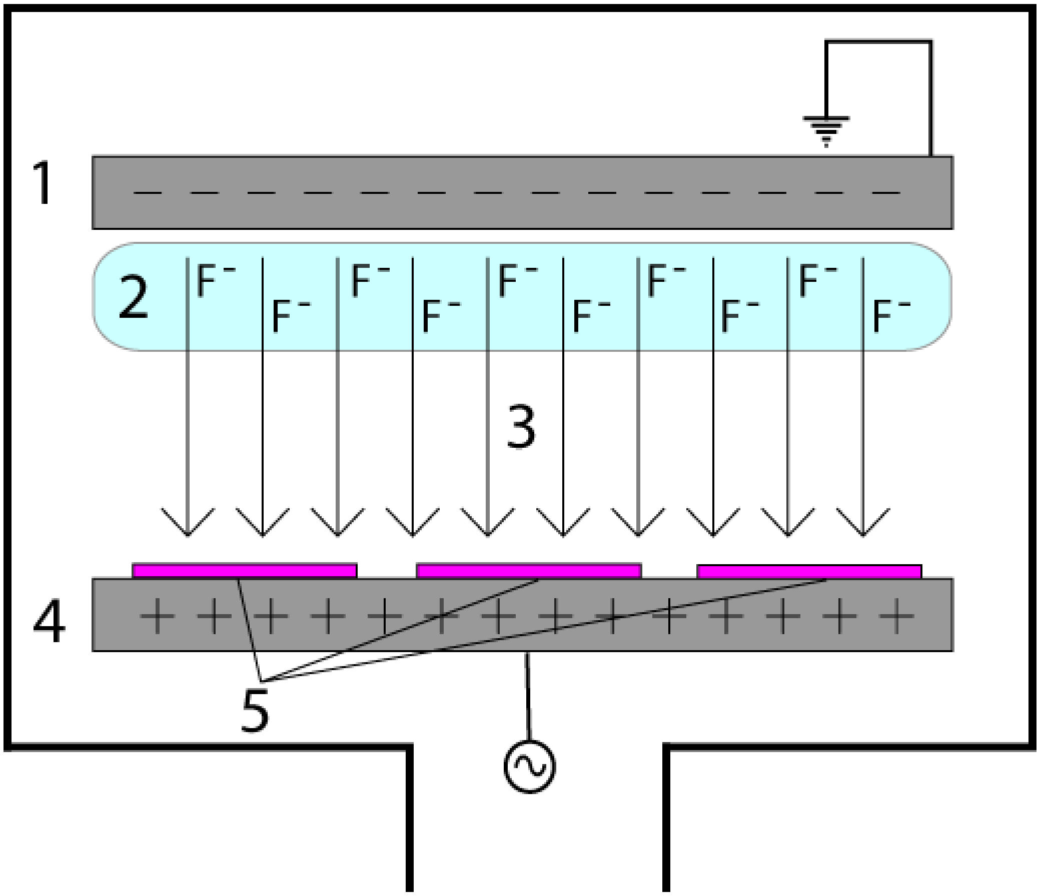
Equipment

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 sputtering process. But this approach has been found to be useful for transferring high-fidelity patterns on non-planar surfaces.

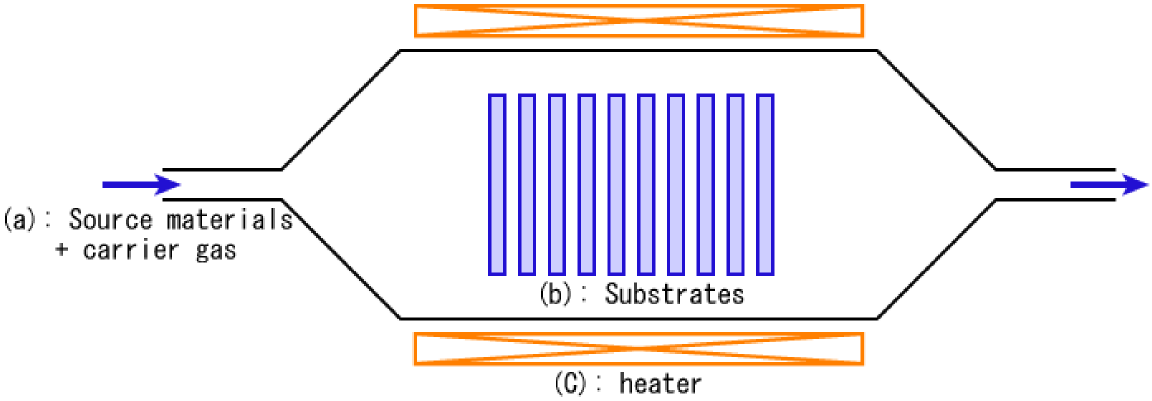
Chemical vapor deposition

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

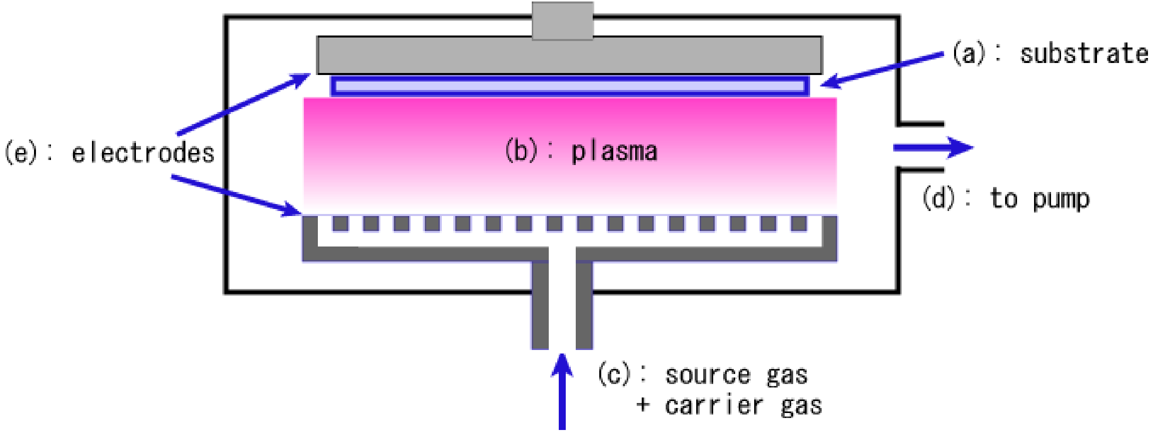
Chemical vapor deposition (CVD) is a chemical process used to produce high-purity, high-performance solid materials. The process is often used in the semiconductor industry to produce thin films. In a typical CVD process, the wafer (substrate) is exposed to one or more volatile precursors, which react 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:

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

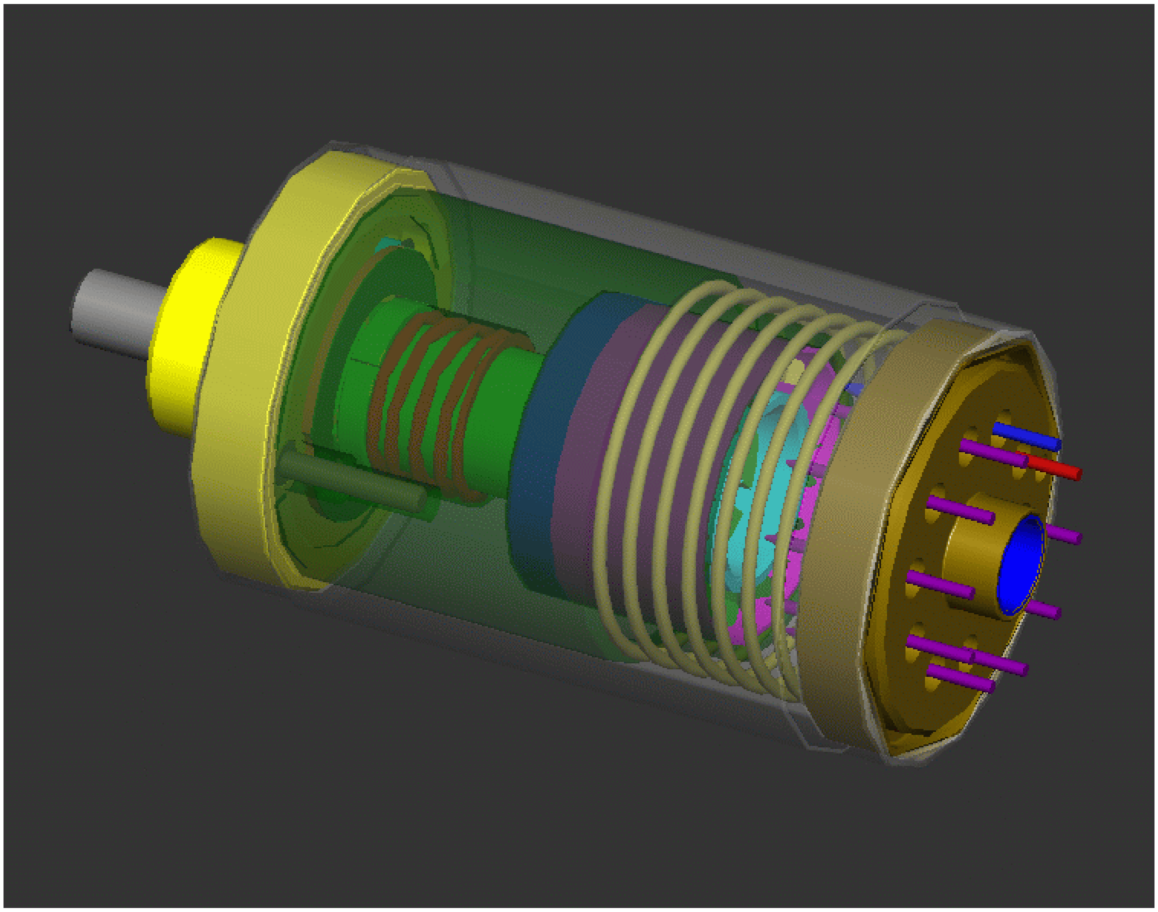
The usual source for tungsten is tungsten hexafluoride, which may be deposited in two ways:

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

Chapter- 5

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

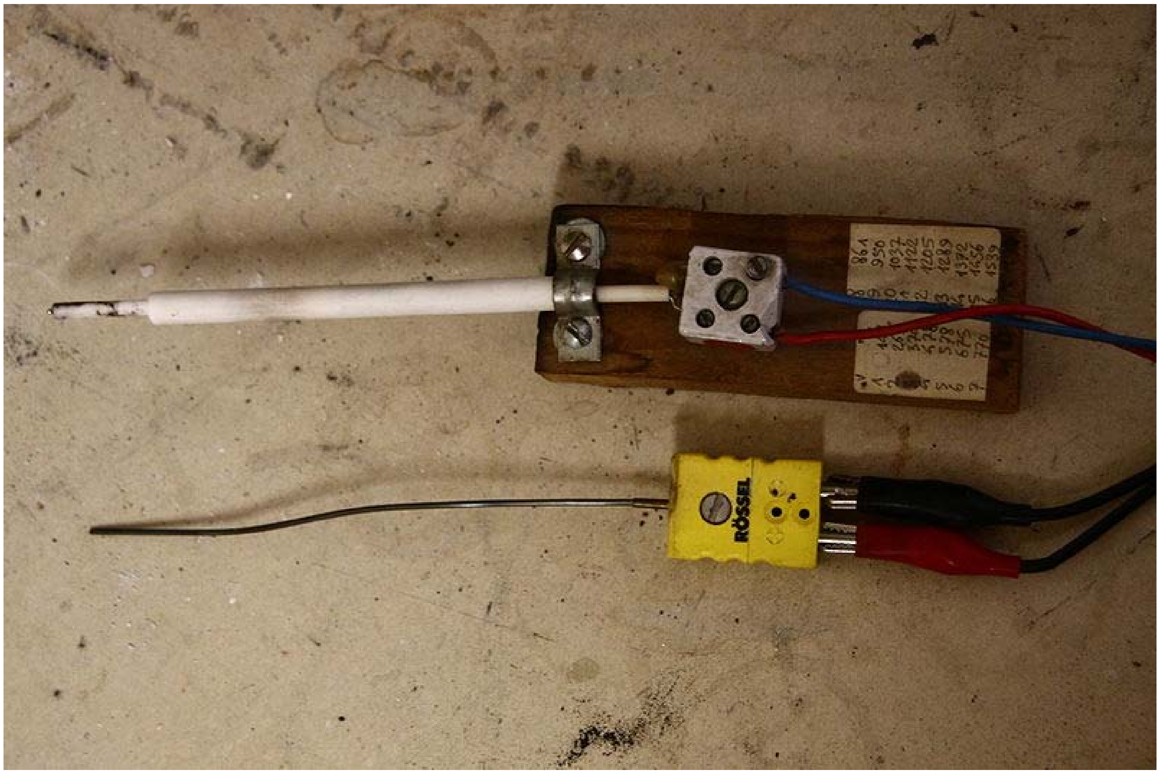
- Piezoelectric accelerometer
- Shear mode accelerometer
- Surface micromachined capacitive (MEMS)
- Thermal (submicrometre CMOS process)
- Bulk micromachined capacitive
- Bulk micromachined piezoelectric resistive
- Capacitive spring mass base
- Electromechanical servo (Servo Force Balance)
- Null-balance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

Chapter- 6

Other Applications of Microelectromechanical Systems

Sensor



Thermocouple sensor for high temperature measurement

A **sensor** is a device that measures a physical quantity and converts it into a signal which can be read by an observer or by an instrument. For example, a mercury-in-glass thermometer converts the measured temperature into expansion and contraction of a liquid which can be read on a calibrated glass tube. A thermocouple converts temperature to an output voltage which can be read by a voltmeter. For accuracy, most sensors are calibrated against known standards.

Use

Sensors are used in everyday objects such as touch-sensitive elevator buttons (tactile sensor) and lamps which dim or brighten by touching the base. There are also innumerable applications for sensors of which most people are never aware. Applications include cars, machines, aerospace, medicine, manufacturing and robotics.

A sensor is a device which receives and responds to a signal. A sensor's sensitivity indicates how much the sensor's output changes when the measured quantity changes. For instance, if the mercury in a thermometer moves 1 cm when the temperature changes by 1 °C, the sensitivity is 1 cm/°C (it is basically the slope Dy/Dx assuming a linear characteristic). Sensors that measure very small changes must have very high sensitivities. Sensors also have an impact on what they measure; for instance, a room temperature thermometer inserted into a hot cup of liquid cools the liquid while the liquid heats the thermometer. Sensors need to be designed to have a small effect on what is measured, making the sensor smaller often improves this and may introduce other advantages. Technological progress allows more and more sensors to be manufactured on a microscopic scale as microsensors using MEMS technology. In most cases, a microsensor reaches a significantly higher speed and sensitivity compared with macroscopic approaches.

Classification of measurement errors

A good sensor obeys the following rules:

- Is sensitive to the measured property
- Is insensitive to any other property likely to be encountered in its application
- Does not influence the measured property

Ideal sensors are designed to be linear or linear to some simple mathematical function of the measurement, typically logarithmic. The output signal of such a sensor is linearly proportional to the value or simple function of the measured property. The sensitivity is then defined as the ratio between output signal and measured property. For example, if a sensor measures temperature and has a voltage output, the sensitivity is a constant with the unit [V/K]; this sensor is linear because the ratio is constant at all points of measurement.

Sensor deviations

If the sensor is not ideal, several types of deviations can be observed:

- The sensitivity may in practice differ from the value specified. This is called a sensitivity error, but the sensor is still linear.
- Since the range of the output signal is always limited, the output signal will eventually reach a minimum or maximum when the measured property exceeds the limits. The full scale range defines the maximum and minimum values of the measured property.
- If the output signal is not zero when the measured property is zero, the sensor has an offset or bias. This is defined as the output of the sensor at zero input.
- If the sensitivity is not constant over the range of the sensor, this is called nonlinearity. Usually this is defined by the amount the output differs from ideal behavior over the full range of the sensor, often noted as a percentage of the full range.
- If the deviation is caused by a rapid change of the measured property over time, there is a dynamic error. Often, this behaviour is described with a bode plot showing sensitivity error and phase shift as function of the frequency of a periodic input signal.
- If the output signal slowly changes independent of the measured property, this is defined as drift (telecommunication).
- Long term drift usually indicates a slow degradation of sensor properties over a long period of time.
- Noise is a random deviation of the signal that varies in time.
- Hysteresis is an error caused by when the measured property reverses direction, but there is some finite lag in time for the sensor to respond, creating a different offset error in one direction than in the other.
- If the sensor has a digital output, the output is essentially an approximation of the measured property. The approximation error is also called digitization error.
- If the signal is monitored digitally, limitation of the sampling frequency also can cause a dynamic error, or if the variable or added noise noise changes periodically at a frequency near a multiple of the sampling rate may induce aliasing errors.
- The sensor may to some extent be sensitive to properties other than the property being measured. For example, most sensors are influenced by the temperature of their environment.

All these deviations can be classified as systematic errors or random errors. Systematic errors can sometimes be compensated for by means of some kind of calibration strategy. Noise is a random error that can be reduced by signal processing, such as filtering, usually at the expense of the dynamic behaviour of the sensor.

Resolution

The resolution of a sensor is the smallest change it can detect in the quantity that it is measuring. Often in a digital display, the least significant digit will fluctuate, indicating that changes of that magnitude are only just resolved. The resolution is related to the precision with which the measurement is made. For example, a scanning tunneling probe (a fine tip near a surface collects an electron tunnelling current) can resolve atoms and molecules.actuator is something that converts energy into motion

Types

Sensors in Nature

All living organisms contain biological sensors with functions similar to those of the mechanical devices described. Most of these are specialized cells that are sensitive to:

- Light, motion, temperature, magnetic fields, gravity, humidity, vibration, pressure, electrical fields, sound, and other physical aspects of the external environment
- Physical aspects of the internal environment, such as stretch, motion of the organism, and position of appendages (proprioception)
- Environmental molecules, including toxins, nutrients, and pheromones
- Estimation of biomolecules interaction and some kinetics parameters
- Internal metabolic milieu, such as glucose level, oxygen level, or osmolality
- Internal signal molecules, such as hormones, neurotransmitters, and cytokines
- Differences between proteins of the organism itself and of the environment or alien creatures.

Biosensor

A **biosensor** is an analytical device for the detection of an analyte that combines a biological component with a physicochemical detector component.

It consists of 3 parts:

- the *sensitive biological element* (biological material (eg. tissue, microorganisms, organelles, cell receptors, enzymes, antibodies, nucleic acids, etc), a biologically derived material or biomimic) The sensitive elements can be created by biological engineering.
- the *transducer* or the *detector element* (works in a physicochemical way; optical, piezoelectric, electrochemical, etc.) that transforms the signal resulting from the interaction of the analyte with the biological element into another signal (i.e., transducers) that can be more easily measured and quantified;
- associated electronics or signal processors that are primarily responsible for the display of the results in a user-friendly way.. This sometimes accounts for the most expensive part of the sensor device, however it is possible to generate a user friendly display that includes transducer and sensitive element.

A common example of a commercial biosensor is the blood glucose biosensor, which uses the enzyme glucose oxidase to break blood glucose down. In doing so it first oxidizes glucose and uses two electrons to reduce the FAD (a component of the enzyme) to FADH2. This in turn is oxidized by the electrode (accepting two electrons from the electrode) in a number of steps. The resulting current is a measure of the concentration of glucose. In this case, the electrode is the transducer and the enzyme is the biologically active component.

Recently, arrays of many different detector molecules have been applied in so called electronic nose devices, where the pattern of response from the detectors is used to fingerprint a substance.. Current commercial electronic noses, however, do not use biological elements.

A canary in a cage, as used by miners to warn of gas, could be considered a biosensor. Many of today's biosensor applications are similar, in that they use organisms which respond to toxic substances at a much lower concentrations than humans can detect to warn of the presence of the toxin. Such devices can be used in environmental monitoring, trace gas detection and in water treatment facilities.

Principles of Detection

Photometric

Many optical biosensors based on the phenomenon of surface plasmon resonance are evanescent wave techniques. This utilises a property of gold and other materials; specifically that a thin layer of gold on a high refractive index glass surface can absorb laser light, producing electron waves (surface plasmons) on the gold surface. This occurs only at a specific angle and wavelength of incident light and is highly dependent on the surface of the gold, such that binding of a target analyte to a receptor on the gold surface produces a measurable signal.

Surface plasmon resonance sensors operate using a sensor chip consisting of a plastic cassette supporting a glass plate, one side of which is coated with a microscopic layer of gold. This side contacts the optical detection apparatus of the instrument. The opposite side is then contacted with a microfluidic flow system. The contact with the flow system creates channels across which reagents can be passed in solution. This side of the glass sensor chip can be modified in a number of ways, to allow easy attachment of molecules of interest. Normally it is coated in carboxymethyl dextran or similar compound.

Light of a fixed wavelength is reflected off the gold side of the chip at the angle of total internal reflection, and detected inside the instrument. This induces the evanescent wave to penetrate through the glass plate and some distance into the liquid flowing over the surface.

The refractive index at the flow side of the chip surface has a direct influence on the behaviour of the light reflected off the gold side. Binding to the flow side of the chip has an effect on the refractive index and in this way biological interactions can be measured to a high degree of sensitivity with some sort of energy.

Other evanescent wave biosensors have been commercialised using waveguides where the propagation constant through the waveguide is changed by the absorption of molecules to the waveguide surface. One such example, Dual Polarisation Interferometry uses a buried waveguide as a reference against which the change in propagation constant is measured. Other configurations such as the Mach-Zehnder have reference arms lithographically defined on a substrate. Higher levels of integration can be achieved using resonator geometries where the resonant frequency of a ring resonator changes when molecules are absorbed.

Other optical biosensors are mainly based on changes in absorbance or fluorescence of an appropriate indicator compound and do not need a total internal reflection geometry. For example, a fully operational prototype device detecting casein in milk has been fabricated. The device is based on detecting changes in absorption of a gold layer. A widely used research tool, the micro-array, can also be considered a biosensor.

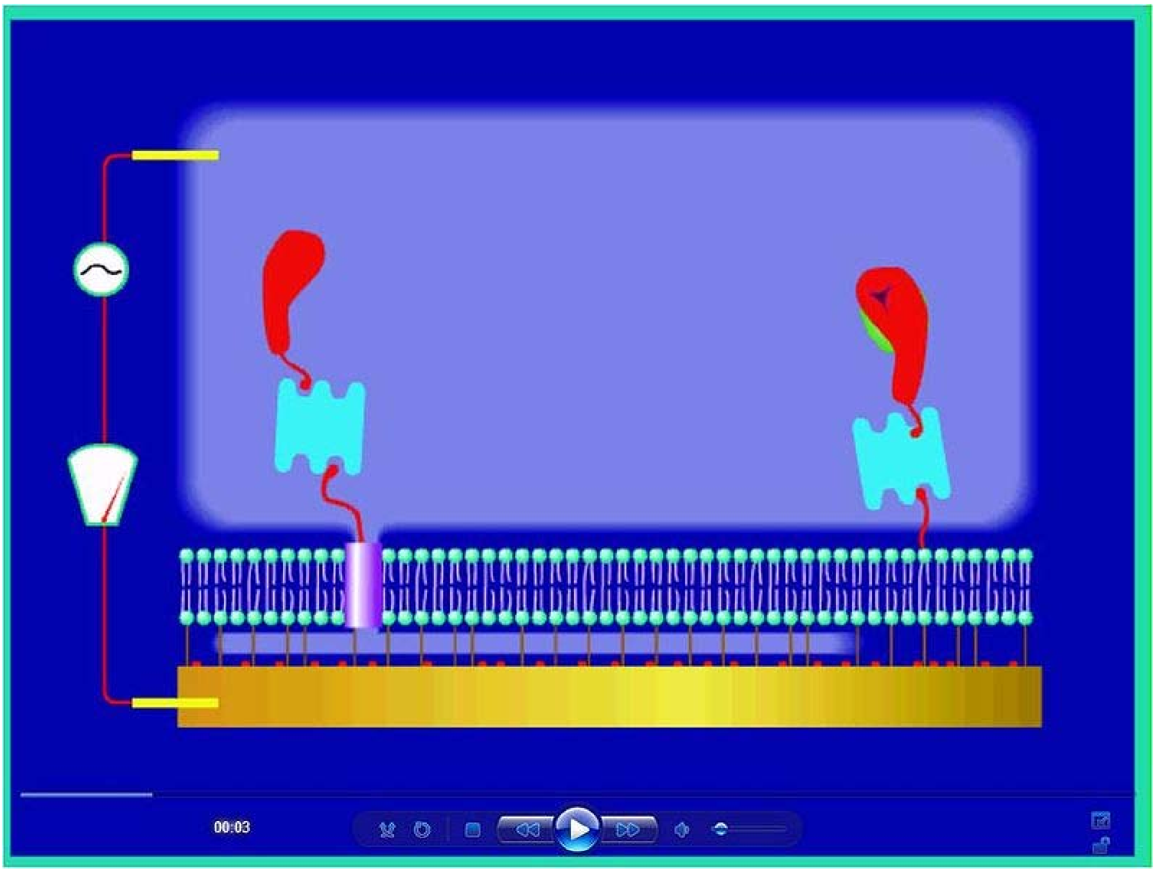
Biological biosensors often incorporate a genetically modified form of a native protein or enzyme. The protein is configured to detect a specific analyte and the ensuing signal is read by a detection instrument such as a fluorometer or luminometer. An example of a recently developed biosensor is one for detecting cytosolic concentration of the analyte cAMP (cyclic adenosine monophosphate), a second messenger involved in cellular signaling triggered by ligands interacting with receptors on the cell membrane. Similar systems have been created to study cellular responses to native ligands or xenobiotics (toxins or small molecule inhibitors). Such "assays" are commonly used in drug discovery development by pharmaceutical and biotechnology companies. Most cAMP assays in current use require lysis of the cells prior to measurement of cAMP. A live-cell biosensor for cAMP can be used in non-lysed cells with the additional advantage of multiple reads to study the kinetics of receptor response.

Electrochemical

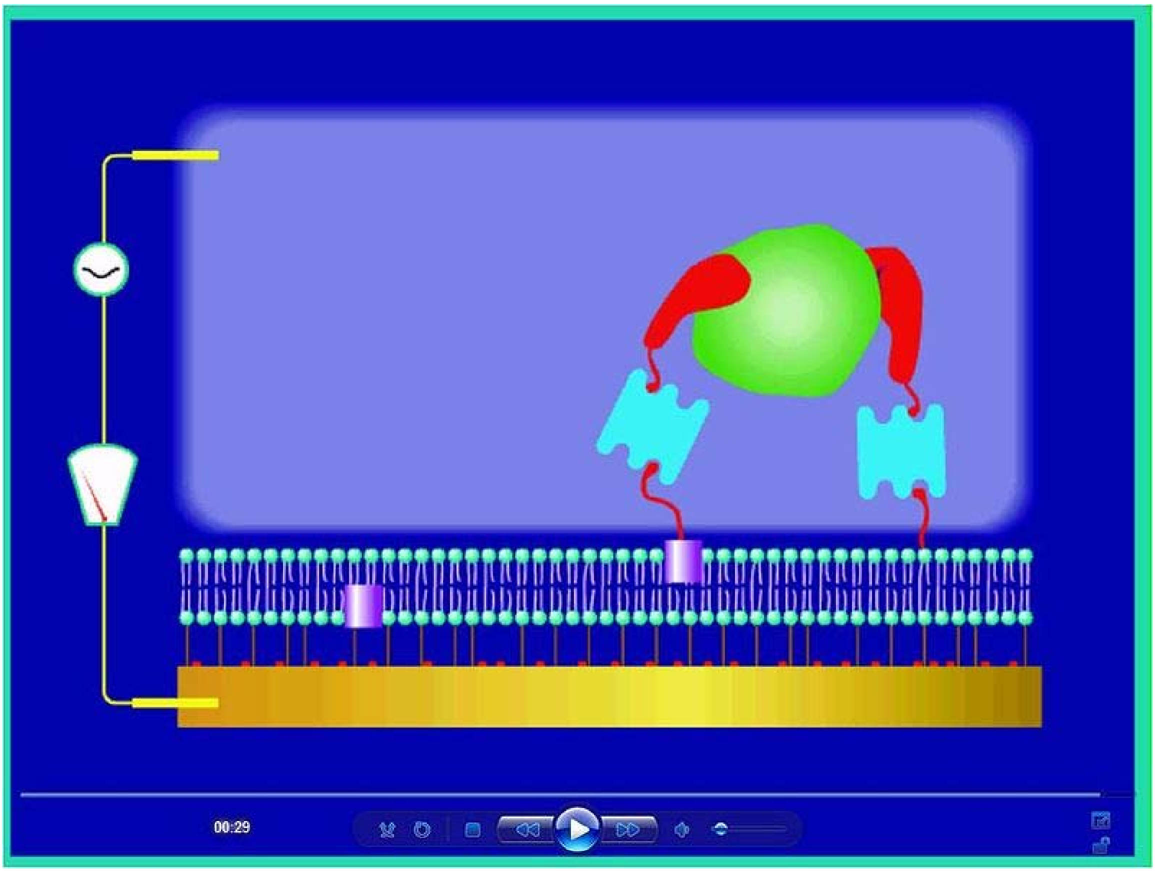
Electrochemical biosensors are normally based on enzymatic catalysis of a reaction that produces or consumes electrons (such enzymes are rightly called redox enzymes). The sensor substrate usually contains three electrodes; a reference electrode, a working electrode and a sink electrode. An auxiliary electrode (also known as a counter electrode) may also be present as an ion source. The target analyte is involved in the reaction that takes place on the active electrode surface, and the ions produced create a potential which is subtracted from that of the reference electrode to give a signal. We can either measure the current (rate of flow of electrons is now proportional to the analyte concentration) at a fixed potential or the potential can be measured at zero current (this gives a logarithmic response). Note that potential of the working or active electrode is space charge sensitive and this is often used. Further, the label-free and direct electrical detection of small peptides and proteins is possible by their intrinsic charges using biofunctionalized ion-sensitive field-effect transistors.

Another example, the potentiometric biosensor, works contrary to the current understanding of its ability. Such biosensors are screenprinted, conducting polymer coated, open circuit potential biosensors based on conjugated polymers immunoassays. They have only two electrodes and are extremely sensitive and robust. They enable the detection of analytes at levels previously only achievable by HPLC and LC/MS and without rigorous sample preparation. The signal is produced by electrochemical and physical changes in the conducting polymer layer due to changes occurring at the surface of the sensor. Such changes can be attributed to ionic strength, pH, hydration and redox reactions, the latter due to the enzyme label turning over a substrate().

Ion Channel Switch



ICS - channel open



ICS - channel closed

The use of ion channels has been shown to offer highly sensitive detection of target biological molecules.. By imbedding the ion channels in supported or tethered bilayer membranes (t-BLM) attached to a gold electrode, an electrical circuit is created. Capture molecules such as antibodies can be bound to the ion channel so that the binding of the target molecule controls the ion flow through the channel. This results in a measurable change in the electrical conduction which is proportional to the concentration of the target.

An Ion Channel Switch (ICS) biosensor can be created using gramicidin, a dimeric peptide channel, in a tethered bilayer membrane.. One peptide of gramicidin, with attached antibody, is mobile and one is fixed. Breaking the dimer stops the ionic current through the membrane. The magnitude of the change in electrical signal is greatly increased by separating the membrane from the metal surface using a hydrophilic spacer.

Quantitative detection of an extensive class of target species, including proteins, bacteria, drug and toxins has been demonstrated using different membrane and capture configurations.

Others

Piezoelectric sensors utilise crystals which undergo an elastic deformation when an electrical potential is applied to them. An alternating potential (A.C.) produces a standing wave in the crystal at a characteristic frequency. This frequency is highly dependent on the elastic properties of the crystal, such that if a crystal is coated with a biological recognition element the binding of a (large) target analyte to a receptor will produce a change in the resonance frequency, which gives a binding signal. In a mode that uses surface acoustic waves (SAW), the sensitivity is greatly increased. This is a specialised application of the Quartz crystal microbalance as a biosensor.

Thermometric and magnetic based biosensors are rare.

Applications

There are many potential applications of biosensors of various types. The main requirements for a biosensor approach to be valuable in terms of research and commercial applications are the identification of a target molecule, availability of a suitable biological recognition element, and the potential for disposable portable detection systems to be preferred to sensitive laboratory-based techniques in some situations. Some examples are given below:

- Glucose monitoring in diabetes patients ←**historical market driver**
- Other medical health related targets
- Environmental applications e.g. the detection of pesticides and river water contaminants
- Remote sensing of airborne bacteria e.g. in counter-bioterrorist activities
- Detection of pathogens
- Determining levels of toxic substances before and after bioremediation
- Detection and determining of organophosphate
- Routine analytical measurement of folic acid, biotin, vitamin B12 and pantothenic acid as an alternative to microbiological assay
- Determination of drug residues in food, such as antibiotics and growth promoters, particularly meat and honey.
- Drug discovery and evaluation of biological activity of new compounds.
- Protein engineering in biosensors
- Detection of toxic metabolites such as mycotoxins

Glucose monitoring

Commercially available gluocose monitors rely on amperometric sensing of glucose by means of glucose oxidase, which oxidises glucose producing hydrogen peroxide which is detected by the electrode. To overcome the limitation of ameperometric sensors, a flurry of research is present into novel sensing methods, such as fluorescent glucose biosensors.

Biosensors in food analysis

There are several applications of biosensors in food analysis. In food industry optic coated with antibodies are commonly used to detect pathogens and food toxins. The light system in these biosensors has been fluorescence, since this type of optical measurement can greatly amplify the signal.

A range of immuno- and ligand-binding assays for the detection and measurement of small molecules such as water-soluble vitamins and chemical contaminants (drug residues) such as sulfonamides and Beta-agonists have been developed for use on SPR based sensor systems, often adapted from existing ELISA or other immunological assay. These are in widespread use across the food industry.

Surface Attachment of the biological elements

An important part in a biosensor is to attach the biological elements (small molecules/protein/cells) to the surface of the sensor (be it metal, polymer or glass). The simplest way is to functionalize the surface in order to coat it with the biological elements. This can be done by polylysine, aminosilane, epoxysilane or nitrocellulose in the case of silicon chips/silica glass. Subsequently the bound biological agent may be for example fixed by Layer by layer depositation of alternatively charged polymer coatings Alternatively three dimensional lattices (hydrogel/xerogel) can be used to chemically or physically entrap these (where by chemically entraped it is meant that the biological element is kept in place by a strong bond, while physically they are kept in place being unable to pass through the pores of the gel matrix). The most commonly used hydrogel is sol-gel, a glassy silica generated by polymerization of silicate monomers (added as tetra alkyl orthosilicates, such as TMOS or TEOS) in the presence of the biological elements (along with other stabilizing polymers, such as PEG) in the case of physical entrapment. Another group of hydrogels, which set under conditions suitable for cells or protein, are acrylate hydrogel, which polymerize upon radical initiation. One type of radical initiator is a peroxide radical, typically generated by combining a persulfate with TEMED (Polyacrylamide gel are also commonly commonly used for protein electrophoresis), alternatively light can be used in combination with a photoinitiator, such as DMPA (2,2-dimethoxy-2-phenylacetophenone). Smart materials that mimic the biological components of a sensor can also be classified as biosensors using only the active or catalytic site or analogous configurations of a biomolecule.

In biomedicine and biotechnology, sensors which detect analytes thanks to a biological component, such as cells, protein, nucleic acid or biomimetic polymers, are called biosensors. Whereas a non-biological sensor, even organic (=carbon chemistry), for biological analytes is referred to as sensor or nanosensor (such a microcantilevers). This terminology applies for both in vitro and in vivo applications. The encapsulation of the biological component in biosensors, presents with a slightly different problem that ordinary sensors, this can either be done by means of a semipermeable barrier, such as a dialysis membrane or a hydrogel, a 3D polymer matrix, which either physically constrains the sensing macromolecule or chemically (macromolecule is bound to the scaffold).

Actuator

An **actuator** is a mechanical device for moving or controlling a mechanism or system. It is operated by a source of energy, usually in the form of an electric current, hydraulic fluid pressure or pneumatic pressure, and converts that energy into some kind of motion.

Examples and applications

- Mechanical actuators operate by conversion of rotary motion into linear motion, or vice versa. Conversion is commonly made via a few simple types of mechanism including:
  - **Screw**: Screw jack, ball screw and roller screw actuators all operate on the principle of the simple machine known as the screw. By rotating the actuator's nut, the screw shaft moves in a line. By moving the screw shaft, the nut rotates.
  - **Wheel and axle**: Hoist, winch, rack and pinion, chain drive, belt drive, rigid chain and rigid belt actuators operate on the principle of the wheel and axle. By rotating a wheel/axle (e.g. drum, gear, pulley or shaft) a linear member (e.g. cable, rack, chain or belt) moves. By moving the linear member, the wheel/axle rotates.
- In engineering, actuators are frequently used as mechanisms to introduce motion, or to clamp an object so as to prevent motion. In electronic engineering, actuators are a subdivision of transducers. They are devices which transform an input signal (mainly an electrical signal) into motion. Specific examples include: electrical motors, pneumatic actuators, hydraulic actuators, linear actuators, comb drive, piezoelectric actuators and amplified piezoelectric actuators, thermal bimorphs, micromirror devices and electroactive polymers.
- Motors are mostly used when circular motions are needed, but can also be used for linear applications by transforming circular to linear motion with a bolt and screw transducer. On the other hand, some actuators are intrinsically linear, such as piezoelectric actuators.

Thermodynamic efficiency

The efficiency of an actuator is a standard tool used to calculate or estimate the usefulness of any actuating mechanism. It is a dimensionless quantity which is generally lower than 1 expressing the energy conversion factor.

$$\varepsilon = \frac{\text{useful work}}{\text{spent energy}} = \frac{\text{output energy}}{\text{input energy}}$$

Interferometric modulator display

The Interferometric Modulator (iMoD) is a technology used in electronic displays that can create various colors through the interference of reflected light. The color is selected with an electrically switched light modulator comprising a microscopic cavity that is switched on and off using driver integrated circuits similar to those used to address LCD displays. An iMoD based reflective flat panel display can include hundreds of thousands of individually addressable iMoD elements. iMoD displays represent one of the foremost examples of a microelectromechanical systems (MEMS) based device.

In one state an iMoD subpixel reflects light at a specific wavelength and gives a pure, bright color at one intensity while in a second state it absorbs incident light and appears black to the viewer. The wing of butterflies employ the same phenomena. When not being addressed, an iMoD display consumes very little power. Unlike conventional back-lit LCD displays (and like other reflective display technologies), it is clearly visible in bright ambient light such as sunlight. iMoD prototypes are currently capable of 15 frames per second and are expected to reach 30, suitable for video playback.

The iMoD was invented by Mark W. Miles, a MEMS researcher and founder of Etalon, Inc., and (co-founder) of Iridigm Display Corporation. Qualcomm took over the development of this technology after its acquisition of Iridigm in 2004, and subsequently formed Qualcomm MEMS Technologies (QMT). Qualcomm has allowed commercialization of the technology under the trademarked name "Mirasol", and this energy-efficient, biomimetic technology sees application and use in mobile phones, etc.

Working principle

A pixel in an iMoD based display consists of one or more subpixels that are actually individual microscopic interferometric cavities similar in operation to Fabry-Pérot interferometers (etalons), and the scales in butterfly wings. While a simple etalon consists of two half-silvered mirrors, an iMoD comprises a reflective membrane which can move in relation to a semitransparent thin film stack. With an air gap defined within this cavity, the iMoD behaves like an optically resonant structure whose reflected color is determined by the size of the airgap. Application of a voltage to the iMoD creates electrostatic forces which bring the membrane into contact with the thin film stack. When this happens the behavior of the iMoD changes to that of an induced absorber. The consequence is that almost all incident light is absorbed and no colors are reflected. It is this binary operation that is the basis for the iMoD's application in reflective flat panel displays. Since the display utilizes light from ambient sources, the display's brightness increases in high ambient environments (i.e. sunlight). In contrast, a back-lit LCD display suffers from incident light.

For a practical RGB display, a single RGB pixel is built from several subpixels, because the brightness of a monochromatic pixel is not adjusted. A monochromatic array of subpixels represents different brightness levels for each color, and for each pixel, there are three such arrays: red, green and blue.

Commercial uses

iMoD displays are now available in the commercial marketplace. QMT's displays, using iMoD technology, are found in the Acoustic Research ARWH1 Stereo Bluetooth headset device, the Showcare Monitoring system (Korea), the Hisense C108, and mp3 applications from Freestyle Audio and Skullcandy. In the mobile phone marketplace, Taiwanese manufacturers Inventec and Cal-Comp have announced phones with Mirasol displays, and LG claims to be developing 'one or more' handsets using Mirasol technology. These products all have only 2-color (black plus one other) "bi-chromic" displays. None are full color.

Related technology

A similar technology to iMoD is time-multiplexed optical shutter (TMOS). A TMOS display uses an array of micromechanical devices (optical shutters) which, when activated, 'tap-out' light from a waveguide at the rear of the display. TMOS displays are efficient compared to LCD displays: Uni-Pixel notes that a TMOS display transmits 61% of the back-light, whereas a LCD display transmits only 5%.

Bio-MEMS

Bio-MEMS is an abbreviation of **biological microelectromechanical systems** and refers to a special class of microelectromechanical systems (MEMS) where biological matter is manipulated to analyze and measure its activity under any class of scientific study. This class of devices belongs to one of the areas of development based on microtechnology. Among the applications based in Bio-MEMS are: biological and biomedical analysis and measurements and micro total analysis systems (TAS). One of the more popular approaches to Bio-MEMS as of late has been through biomimetics.

Optical switch

In telecommunication, an **optical switch** is a switch that enables signals in optical fibers or integrated optical circuits (IOCs) to be selectively switched from one circuit to another.

The word is used on several levels. In commercial terms (such as "the telecom optical switch market size") it refers to any piece of circuit switching equipment between fibers. The majority of installed systems in this category actually use electronic switching between fiber transponders. Systems that perform this function by physically switching light are often referred to as "photonic" switches, independent of how the light itself is switched. Away from the world of telecom systems, an optical switch is the unit that actually switches light between fibers, and a photonic switch is one that does this by exploiting nonlinear material properties to steer light (i.e., to switch wavelengths or signals within a given fiber).

Hence a certain portion of the optical switch market is made up of photonic switches. These will contain within them an optical switch, which will, in a small number of cases, be a photonic switch.

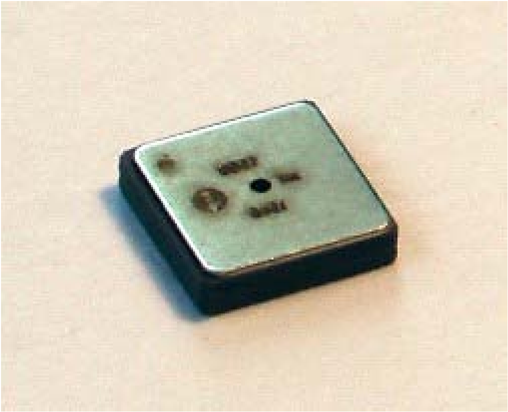
An optical switch may operate by mechanical means, such as physically shifting an optical fiber to drive one or more alternative fibers, or by electro-optic effects, magneto-optic effects, or other methods. Slow optical switches, such as those using moving fibers, may be used for alternate routing of an optical transmission path, such as routing around a fault. Fast optical switches, such as those using electro-optic or magneto-optic effects, may be used to perform logic operations; also included in this category are the semiconductor optical amplifiers, which are optoelectronic devices that can be used as optical switches and be integrated with discrete or integrated microelectronic circuits.

A search on "optical switch" yielded some 8000 patents, roughly categorized as follows:

- MEMS approaches involving arrays of micromirrors that can deflect an optical signal to the appropriate receiver (e.g., U.S. Patent 6,396,976);
- Piezoelelectric Beam Steering involving piezoelctric ceramics providing enhanced optical switching characteristics
- Inkjet methods involving the intersection of two waveguides so that light is deflected from one to the other when an inkjet-like bubble is created (e.g., Patent 6,212,308);
- Liquid crystals (e.g., Patent 4,948,229) that rotate polarized light either 0 degrees or 90 degrees depending on the applied electric field to support switching;
- Thermal methods (e.g., Patent 5,037,169) that vary the index of refraction in one leg of an interferometer to switch the signal on or off;
- Nonlinear methods (e.g., Patent 5,319,492) that vary the diffraction pattern in a medium by taking advantage of the material nonlinear properties to deflect light to the desired receiver;
- Acousto-optic methods that change the refractive index as a result of strain induced by an acoustic field to deflect light (e.g., Patent 6,922,498);
- Amplifiers and attenuators in output fibers that adjust the signal to the digital "0" power range (when the fiber is not switched to) or to the normal power range when it is (e.g., Patent 7,027,211).

Pressure sensor

Digital air pressure sensor



Compact digital barometric pressure sensor

A **pressure sensor** measures pressure, typically of gases or liquids. Pressure is an expression of the force required to stop a fluid from expanding, and is usually stated in terms of force per unit area. A pressure sensor usually acts as a transducer; it generates a signal as a function of the pressure imposed.

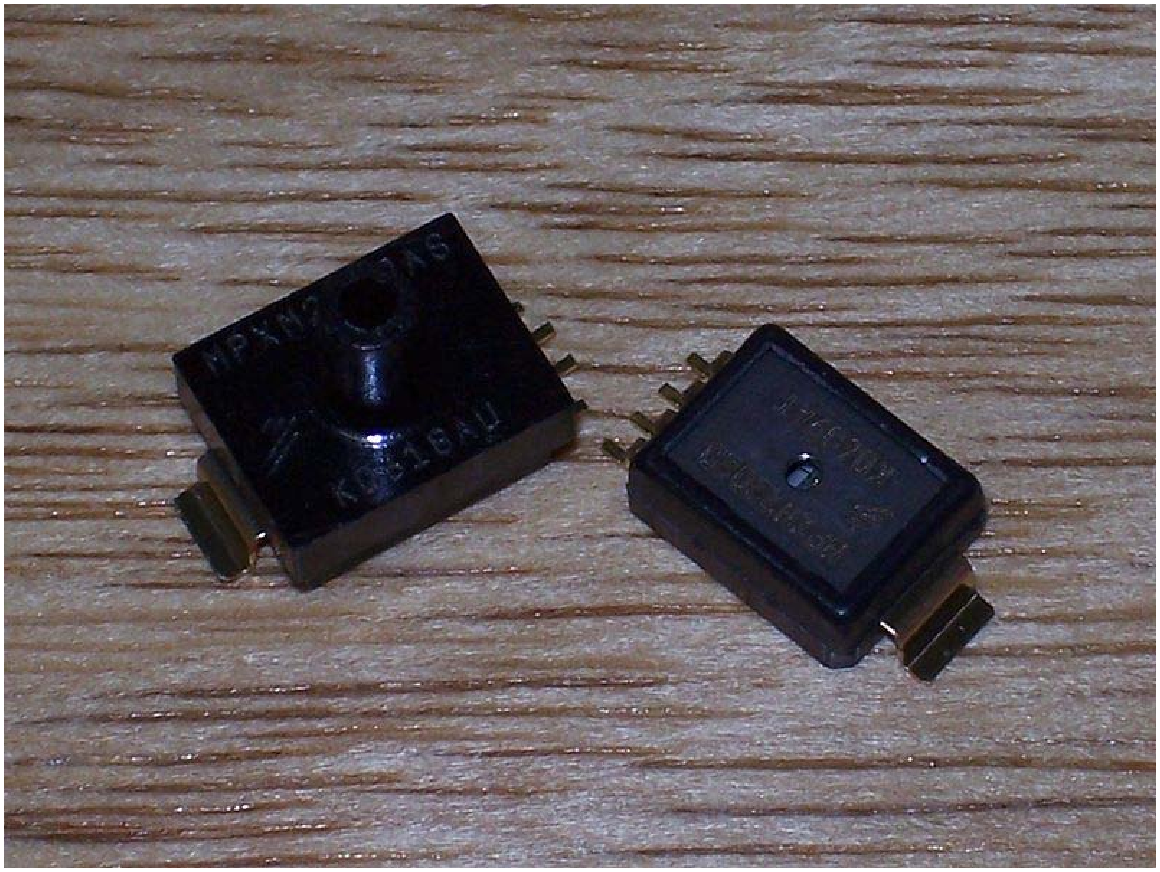
Pressure sensors are used for control and monitoring in thousands of everyday applications. Pressure sensors can also be used to indirectly measure other variables such as fluid/gas flow, speed, water level, and altitude. Pressure sensors can alternatively be called **pressure transducers**, **pressure transmitters**, **pressure senders**, **pressure indicators** and **piezometers**, **manometers**, among other names.

Pressure sensors can vary drastically in technology, design, performance, application suitability and cost. A conservative estimate would be that there may be over 50 technologies and at least 300 companies making pressure sensors worldwide.

There is also a category of pressure sensors that are designed to measure in a dynamic mode for capturing very high speed changes in pressure. Example applications for this type of sensor would be in the measuring of combustion pressure in an engine cylinder or in a gas turbine. These sensors are commonly manufactured out of piezoelectric materials such as quartz.

Some pressure sensors, such as those found in some traffic enforcement cameras, function in a binary (on/off) manner, i.e., when pressure is applied to a pressure sensor, the sensor acts to complete or break an electrical circuit. These types of sensors are also known as a pressure switch.

Types of pressure measurements



silicon piezoresistive pressure sensors

Pressure sensors can be classified in terms of pressure ranges they measure, temperature ranges of operation, and most importantly the type of pressure they measure. In terms of pressure type, pressure sensors can be divided into five categories:

- **Absolute pressure sensor**

This sensor measures the pressure relative to perfect vacuum pressure (0 PSI or no pressure). Atmospheric pressure, is 101.325 kPa (14.7 PSI) at sea level with reference to vacuum.

- **Gauge pressure sensor**

This sensor is used in different applications because it can be calibrated to measure the pressure relative to a given atmospheric pressure at a given location. A tire pressure gauge is an example of gauge pressure indication. When the tire pressure gauge reads 0 PSI, there is really 14.7 PSI (atmospheric pressure) in the tire.

- **Vacuum pressure sensor**

This sensor is used to measure pressure less than the atmospheric pressure at a given location. This has the potential to cause some confusion as industry may refer to a vacuum sensor as one which is referenced to either atmospheric pressure (i.e. measure Negative gauge pressure) or relative to absolute vacuum.

- **Differential pressure sensor**

This sensor measures the difference between two or more pressures introduced as inputs to the sensing unit, for example, measuring the pressure drop across an oil filter. Differential pressure is also used to measure flow or level in pressurized vessels.

- **Sealed pressure sensor**

This sensor is the same as the gauge pressure sensor except that it is previously calibrated by manufacturers to measure pressure relative to sea level pressure.

Pressure Sensing Technology

There are two basic categories of analog pressure sensors.

Force Collector Types These types of electronic pressure sensors generally use a force collector (such a diaphragm, piston, bourdon tube, or bellows) to measure strain (or deflection) due to applied force (pressure) over an area.

- ***Piezoresistive Strain Gage***

  Uses the piezoresistive effect of bonded or formed strain gages to detect strain due to applied pressure. Common technology types are Silicon (Monocrystalline), Polysilicon Thin Film, Bonded Metal Foil, Thick Film, and Sputtered Thin Film. Generally, the strain gauges are connected to form a Wheatstone bridge circuit to maximize the output of the sensor. This is the most commonly employed sensing technology for general purpose pressure measurement. Generally, these technologies are suited to measure absolute, gauge, vacuum, and differential pressures.

- ***Capacitive***

  Uses a diaphragm and pressure cavity to create a variable capacitor to detect strain due to applied pressure. Common technologies use metal, ceramic, and silicon diaphragms. Generally, these technologies are most applied to low pressures (Absolute, Differential and Gauge)

- ***Electromagnetic***

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

- ***Piezoelectric***

  Uses the piezoelectric effect in certain materials such as quartz to measure the strain upon the sensing mechanism due to pressure. This technology is commonly employed for the measurement of highly dynamic pressures.

- ***Optical***

  Uses the physical change of an optical fiber to detect strain due to applied pressure. A common example of this type utilizes Fiber Bragg Gratings. This technology is employed in challenging applications where the measurement may be highly remote, under high temperature, or may benefit from technologies inherently immune to electromagnetic interference.

- ***Potentiometric***

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

Other Types

These types of electronic pressure sensors use other properties (such as density) to infer pressure of a gas, or liquid.

- ***Resonant***

  Uses the changes in resonant frequency in a sensing mechanism to measure stress, or changes in gas density, caused by applied pressure. This technology may be used in conjunction with a force collector, such as those in the category above. Alternatively, resonant technology may be employed by expose the resonating element itself to the media, whereby the resonant frequency is dependent upon the density of the media. Sensors have been made out of vibrating wire, vibrating cylinders, quartz, and silicon MEMS. Generally, this technology is considered to provide very stable readings over time.

- ***Thermal***

Uses the changes in thermal conductivity of a gas due to density changes to measure pressure. A common example of this type is the Pirani gauge.

- ***Ionization***

  Measures the flow of charged gas particles (ions) which varies due to density changes to measure pressure. Common examples are the Hot and Cold Cathode gages.

- ***Others***

  There are numerous other ways to derive pressure from its density (speed of sound, mass, index of refraction) among others.

Applications

There are many applications for pressure sensors:

- **Pressure sensing**

This is the direct use of pressure sensors to measure pressure. This is useful in weather instrumentation, aircraft, cars, and any other machinery that has pressure functionality implemented.

- **Altitude sensing**

This is useful in aircraft, rockets, satellites, weather balloons, and many other applications. All these applications make use of the relationship between changes in pressure relative to the altitude. This relationship is governed by the following equation:

$$h = (1 - (P/P_{ref})^{0.190284}) \times 145366.45 ft$$

This equation is calibrated for an altimeter, up to 36,090 feet (11,000 m). Outside that range, an error will be introduced which can be calculated differently for each different pressure sensor. These error calculations will factor in the error introduced by the change in temperature as we go up.

Barometric pressure sensors can have an altitude resolution of less than 1 meter, which is significantly better than GPS systems (about 20 meters altitude resolution). In navigation applications altimeters are used to distinguish between stacked road levels for car navigation and floor levels in buildings for pedestrian navigation.

- **Flow sensing**

This is the use of pressure sensors in conjunction with the venturi effect to measure flow. Differential pressure is measured between two segments of a venturi tube that have a different aperture. The pressure difference between the two segments is directly proportional to the flow rate through the venturi tube. A low pressure sensor is almost always required as the pressure difference is relatively small.

- **Level / Depth sensing**

A pressure sensor may also be used to calculate the level of a fluid. This technique is commonly employed to measure the depth of a submerged body (such as a diver or submarine), or level of contents in a tank (such as in a water tower). For most practical purposes, fluid level is directly proportional to pressure. In the case of fresh water where the contents are under atmospheric pressure, 1psi = 27.7 inH20 / 1Pa = 9.81 mmH20. The basic equation for such a measurement is

$$P = p * g * h$$

Where P = Pressure, p = Density of the Fluid, g = Standard Gravity, h = Height of fluid column above pressure sensor

- **Leak Testing**

A pressure sensor may be used to sense the decay of pressure due to a system leak. This is commonly done by either comparison to a known leak using differential pressure, or by means of utilizing the pressure sensor to measure pressure change over time.

MEMS gyroscope

Inexpensive (as of 2010, approximately US$5 per part in quantity) vibrating structure gyroscopes manufactured with MEMS technology have become widely available. These are packaged similarly to other integrated circuits and may provide either analog or digital outputs. In many cases, a single part includes gyroscopic sensors for multiple axes. Some parts incorporate both a gyroscope and an accelerometer, in which case the output has six full degrees of freedom. InvenSense, STMicroelectronics, and Analog Devices are major manufacturers.

Internally, MEMS gyroscopes use lithographically constructed versions of one or more of the mechanisms outlined above (tuning forks, vibrating wheels, or resonant solids of various designs).

RF MEMS

The **RF MEMS** acronym stands for radio frequency microelectromechanical system, and refers to components of which moving sub-millimeter-sized parts provide RF functionality. RF functionality can be implemented using a variety of RF technologies. Besides RF MEMS technology, ferrite, ferroelectric, GaAs, GaN, InP, RF CMOS, SiC, and SiGe technology are available to the RF designer. Each of the RF technologies offers a distinct trade-off between cost, frequency, gain, large scale integration, lifetime, linearity, noise figure, packaging, power consumption, power handling, reliability, repeatability, ruggedness, size, supply voltage, switching time and weight.

Components

There are various types of RF MEMS components, such as RF MEMS resonators and self-sustained oscillators with low phase noise , RF MEMS tunable inductors, and RF MEMS switches, switched capacitors and varactors.

Switches, switched capacitors and varactors

RF MEMS switches, switched capacitors and varactors, which can replace field effect transistor (FET) switches and PIN diodes, are classified by actuation method (electrostatic, electrothermal, magnetic, piezoelectric), by axis of deflection (laterally, vertically), by circuit configuration (series, shunt), by clamp configuration (cantilever, fixed-fixed beam), or by contact interface (capacitive, ohmic) . Electrostatically-actuated RF MEMS components offer low insertion loss and high isolation, high linearity, high power handling and high Q factor, do not consume power, but require a high supply voltage and hermetic wafer level packaging (WLP) (anodic or glas frit wafer bonding) or single chip packaging (SCP) (thin film capping, liquid crystal polymer (LCP) or low temperature co-fired ceramic (LTCC) packaging).

RF MEMS switches were pioneered by Hughes Research Laboratories, Malibu, CA , Raytheon, Dallas, TX , and Rockwell Science, Thousand Oaks, CA , during the nineties. The component shown in Fig. 1, is a center-pulled capacitive fixed-fixed beam RF MEMS switch, developed and patented by Raytheon in 1994. A capacitive fixed-fixed beam RF MEMS switch is in essence a micro-machined capacitor with a moving top electrode - i.e. the beam.

From an electromechanical perspective, the components behave like a mass-spring system, actuated by an electrostatic force. The spring constant is a function of the dimensions of the beam, of the Young's modulus, of the residual stress and of the Poisson ratio of its material. The electrostatic force is a function of the capacitance and the bias voltage. Knowledge of spring constant and mass allows for calculation of the pull-in voltage, which is the bias voltage necessary to pull-in the beam, and of the switching time.

From an RF perspective, the components behave like a series RLC circuit with negligible resistance and inductance. The up- and down-state capacitance are in the order of 50 fF and 1.2 pF, which are functional values for millimeter-wave circuit design. Switches typically have a capacitance ratio of 30 or higher, while switched capacitors and varactors have a capacitance ratio of about 1.2 to 10. The loaded Q factor is between 20 and 50 in the X-, Ku- and Ka-band.

RF MEMS switched capacitors are capacitive fixed-fixed beam switches with a low capacitance ratio. RF MEMS varactors are capacitive fixed-fixed beam switches which are biased below pull-in voltage. Other examples of RF MEMS switches are ohmic cantilever switches, and capacitive single pole N throw (SPNT) switches based on the axial gap wobble motor .

Microfabrication

An RF MEMS fabrication process allows for integration of SiCr or TaN thin film resistors (TFR), metal-air-metal (MAM) capacitors, metal-insulator-metal (MIM) capacitors, and RF MEMS components. An RF MEMS fabrication process can be realized on a variety of wafers: fused silica (quartz), borosilicate glass, LCP, sapphire, and passivated silicon and III-V compound semiconducting wafers. As shown in Fig. 2, RF MEMS components can be fabricated in class 100 clean rooms using 6 to 8 optical lithography steps with a 5 µm contact alignment error, whereas state-of-the-art monolithic microwave integrated circuit (MMIC) and radio frequency integrated circuit (RFIC) fabrication processes require 13 to 25 lithography steps. The essential microfabrication steps are:

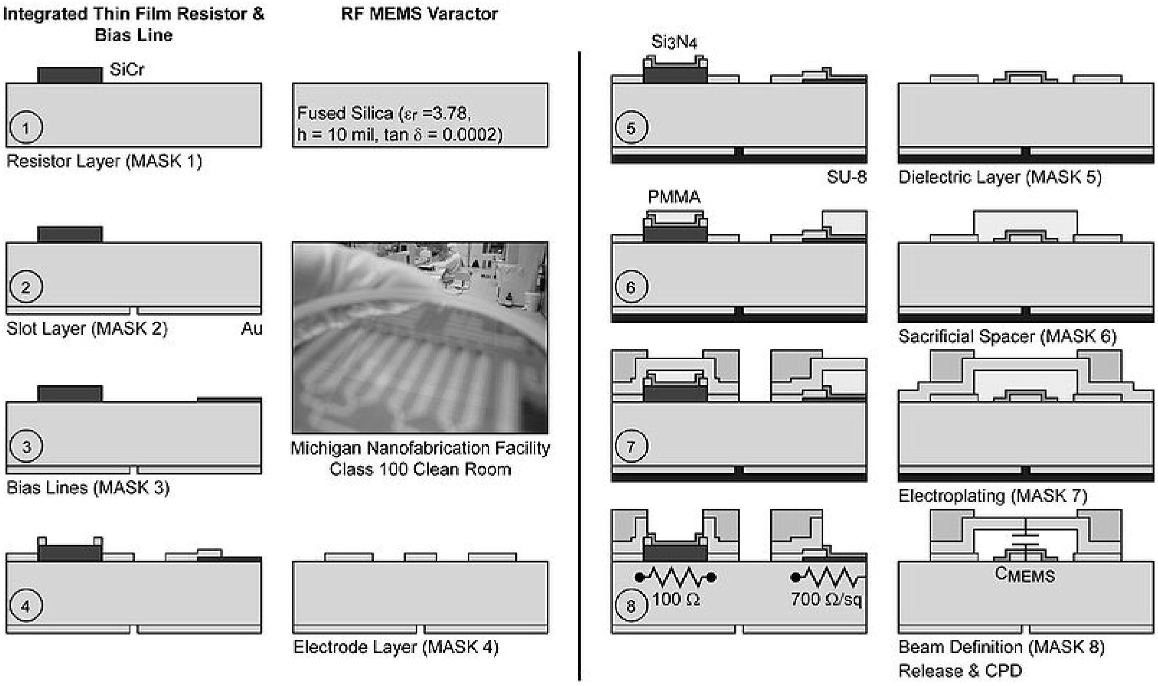


Fig. 2: RF MEMS fabrication process

- Deposition of the bias lines (Fig. 2, step 3)
- Deposition of the electrode layer (Fig. 2, step 4)
- Deposition of the dielectric layer (Fig. 2, step 5)
- Deposition of the sacrificial spacer (Fig. 2, step 6)
- Deposition of seed layer and subsequent electroplating (Fig. 2, step 7)

- Beam definition, release and critical point drying (Fig. 2, step 8)

RF MEMS fabrication processes, unlike barium strontium titanate (BST) or lead zirconate titanate (PZT) ferroelectric and MMIC fabrication processes, do not require electron beam lithography, molecular beam epitaxy (MBE), or metal organic chemical vapor deposition (MOCVD). With the exception of the removal of the sacrificial spacer, the fabrication steps are compatible with a CMOS fabrication process.

Applications

Applications of RF MEMS resonators and switches include oscillators and routing networks. RF MEMS components are also applied in radar sensors (passive electronically scanned (sub)arrays and T/R modules) and software-defined radio (reconfigurable antennas, tunable band-pass filters).

Antennas

Polarization and radiation pattern reconfigurability, and frequency tunability, are usually achieved by incorporation of lumped components based on III-V semiconductor technology, such as single pole single throw (SPST) switches or varactor diodes. However, these components can be readily replaced by RF MEMS switches and varactors in order to take advantage of the low insertion loss and high Q factor offered by RF MEMS technology. In addition, RF MEMS components can be integrated monolithically on low-loss dielectric substrates, such as borosilicate glass, fused silica or LCP, whereas III-V semiconducting substrates are generally lossy and have a high dielectric constant. A low loss tangent and low dielectric constant are of importance for the efficiency and the bandwidth of the antenna.

The prior art includes an RF MEMS frequency tunable fractal antenna for the 0.1–6 GHz frequency range , and the actual integration of RF-MEMS on a self-similar Sierpinski gasket antenna to increase its number of resonant frequencies, extending its range to 8 GHz, 14 GHz and 25 GHz ,, an RF MEMS radiation pattern reconfigurable spiral antenna for 6 and 10 GHz , an RF MEMS radiation pattern reconfigurable spiral antenna for the 6–7 GHz frequency band based on packaged Radant MEMS SPST-RMSW100 switches , an RF MEMS multiband Sierpinski fractal antenna, again with integrated RF MEMS switches, functioning at different bands from 2.4 to 18 GHz , and a 2-bit Ka-band RF MEMS frequency tunable slot antenna .

Filters

RF bandpass filters are used to increase out-of-band rejection, if the antenna fails to provide sufficient selectivity. Out-of-band rejection eases the dynamic range requirement of low noise amplifier LNA and mixer in the light of interference. Off-chip RF bandpass filters based on lumped bulk acoustic wave (BAW), ceramic, surface acoustic wave (SAW), quartz crystal, and thin film bulk acoustic resonator (FBAR) resonators have superseded distributed RF bandpass filters based on transmission line resonators, printed on substrates with low loss tangent, or based on waveguide cavities. RF MEMS resonators offer the potential of on-chip integration of high-Q resonators and low-loss bandpass filters. The Q factor of RF MEMS resonators is in the order of 1000-1000 .

Tunable RF bandpass filters offer a significant size reduction over switched RF bandpass filter banks. They can be implemented using III-V semiconducting varactors, BST or PZT ferroelectric and RF MEMS switches, switched capacitors and varactors, and yttrium iron garnet (YIG) ferrites. RF MEMS technology offers the tunable filter designer a compelling trade-off between insertion loss, linearity, power consumption, power handling, size, and switching time.

Phase shifters

RF MEMS phase shifters have enabled wide-angle passive electronically scanned arrays, such as lenses, reflect arrays, subarrays and switched beamforming networks, with high effective isotropically radiated power (EIRP), also referred to as the power-aperture product, and high G_r/T. EIRP is the product of the transmit gain, G_t, and the transmit power, P_t. G_r/T is the quotient of the receive gain and the antenna noise temperature. A high EIRP and G_r/T are a prerequisite for long-range detection. The EIRP and G_r/T are a function of the number of antenna elements per subarray and of the maximum scanning angle. The number of antenna elements per subarray should be chosen to optimize the EIRP or the EIRP x G_r/T product, as shown in Fig. 3 and Fig. 4.

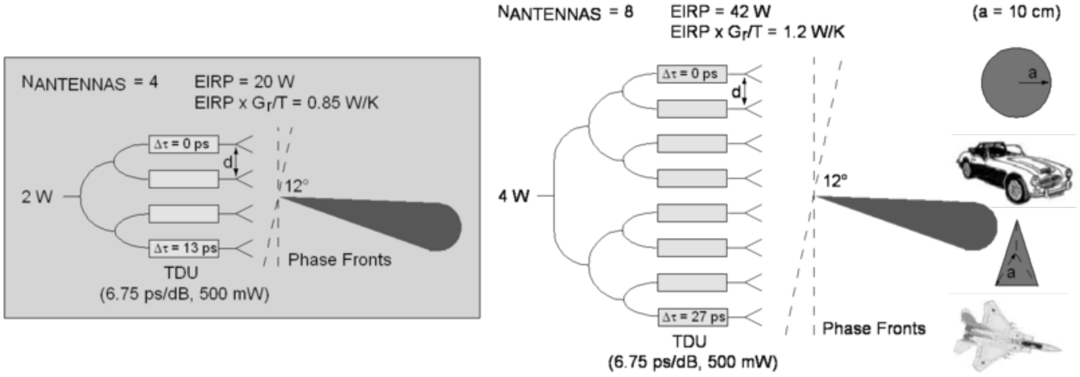


Fig. 3: Radar sensor sensitivity: EIRP x G_r/T

Passive subarrays based on RF MEMS phase shifters may be used to lower the amount of T/R modules in an active electronically scanned array. The statement is illustrated with examples in Fig. 3: assume a one-by-eight passive subarray is used for transmit as well as receive, with following characteristics: f = 38 GHz, $G_r = G_t = 10$ dBi, BW = 2 GHz, $P_t = 4$ W. The low loss (6.75 ps/dB) and good power handling (500 mW) of the RF MEMS phase shifters allow an EIRP of 40 W and a G_r/T of 0.036 1/K. The number of antenna elements per subarray should be chosen in order to optimize the EIRP or the EIRP x G_r/T product, as shown in Fig. 3 and Fig. 4. The radar range equation can be used to calculate the maximum range for which targets can be detected with 10 dB of SNR at the input of the receiver.

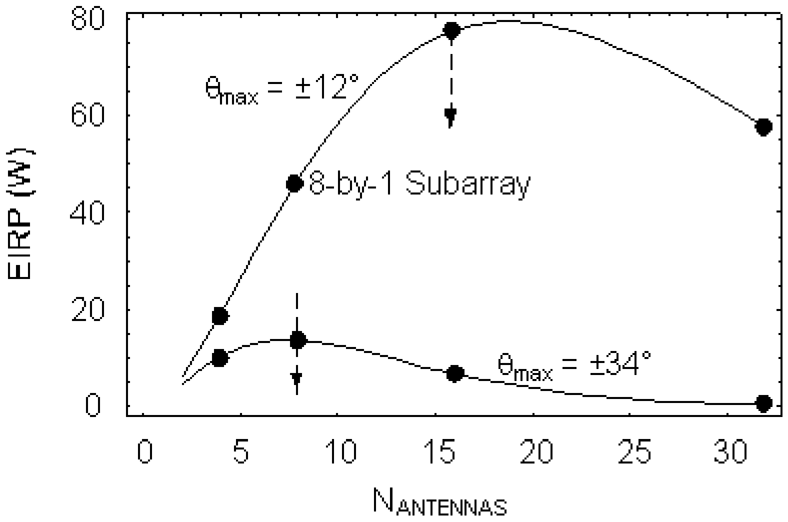


Fig. 4: EIRP versus number of antenna elements in a passive subarray.

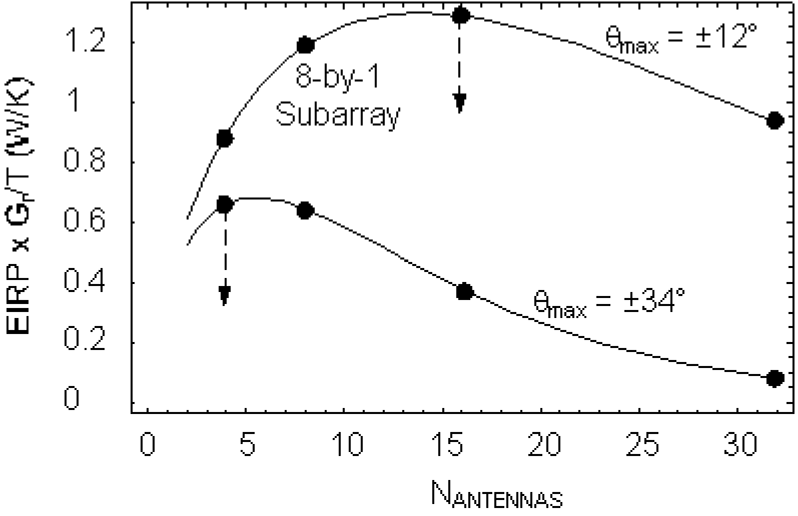


Fig. 5: EIRP x G_r/T versus number of antenna elements in a passive subarray.

$$R = \sqrt[4]{\frac{\lambda^2\,\mathrm{EIRP}\,G_R/T\,\sigma}{64\,\pi^3\,k_B\,\mathrm{BW}\,\mathrm{SNR}}}$$

in which k_B is the Boltzmann constant, λ is the free-space wavelength, and σ is the RCS of the target. Range values are tabulated in Table 1 for following targets: a sphere with a radius, a, of 10 cm ($\sigma = \pi a^2$), a dihedral corner reflector with facet size, a, of 10 cm ($\sigma = 12 a^4/\lambda^2$), the rear of a car ($\sigma = 20 m^2$) and for a contemporary non-evasive fighter jet ($\sigma = 400 m^2$). A Ka-band hybrid ESA capable of detecting a car 100 m in front and engaging a fighter jet at 10 km can be realized using 2.5 and 422 passive subarrays (and T/R modules), respectively.

Table 1: Maximum Detectable Range
(SNR = 10 dB)

	RCS (m^2)	Range (m)
Sphere	0.0314	10
Rear of Car	20	51
Dihedral Corner Reflector	60.9	67
Fighter Jet	400	107

The usage of true-time-delay TTD phase shifters instead of RF MEMS phase shifters allows ultra-wideband (UWB) radar waveforms with associated high range resolution, and avoids beam squinting or frequency scanning. TTD phase shifters are designed using the switched-line principle or the distributed loaded-line principle . Switched-line TTD phase shifters are superior to distributed loaded-line TTD phase shifters in terms of time delay per decibel noise figure (NF), especially at frequencies up to X-band, but are inherently digital and require low-loss and high-isolation SPNT switches. Distributed loaded-line TTD phase shifters, however, can be realized analogously or digitally, and in smaller form factors, which is important at the subarray level. Analog phase shifters are biased through a single bias line, whereas multibit digital phase shifters require a parallel bus along with complex routing schemes at the subarray level. In addition, usage of an analog bias voltage avoids large phase quantization errors, which deteriorate the EIRP and beam-pointing accuracy, and elevate the sidelobe level of an electronically scanned array.

The prior art in passive electronically scanned arrays, shown in Fig. 6, includes an X-band continuous transverse stub (CTS) array fed by a line source synthesized by sixteen 5-bit reflect-type RF MEMS phase shifters based on ohmic cantilever RF MEMS switches, an X-band 2-D lens consisting of parallel-plate waveguides and featuring 25,000 ohmic cantilever RF MEMS switches , and a W-band switched beamforming network based on an RF MEMS SP4T switch and a Rotman lens focal plane scanner .

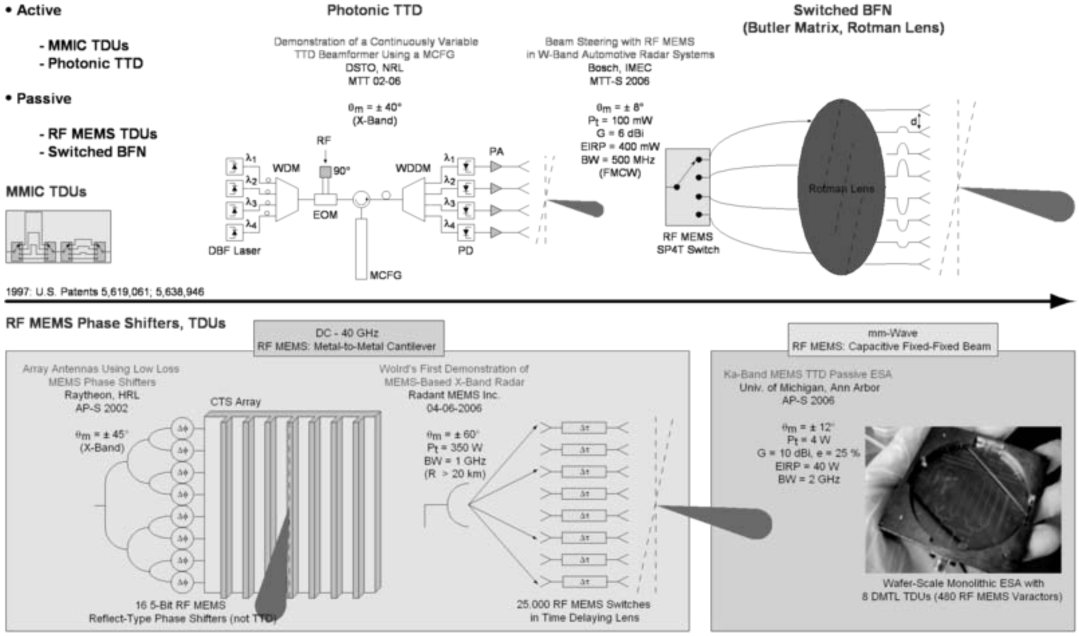


Fig. 6: RF MEMS components in a passive electronically scanned array

T/R modules

Within a T/R module, as shown in Fig. 7, RF MEMS limiters, tunable matching networks and TTD phase shifters can be used to protect the LNA, load-pull the power amplifier (PA) and time delay the RF signal, respectively. Whether RF MEMS T/R switches - i.e. single pole double throw (SPDT) switches, can be used depends on the duty cycle and the pulse repetition frequency (PRF) of the pulse-Doppler radar waveform. To date, RF MEMS duplexers can only be used in low PRF and medium PRF radar waveforms for long-range detection, which use pulse compression and therefore have a duty cycle in the order of microseconds.

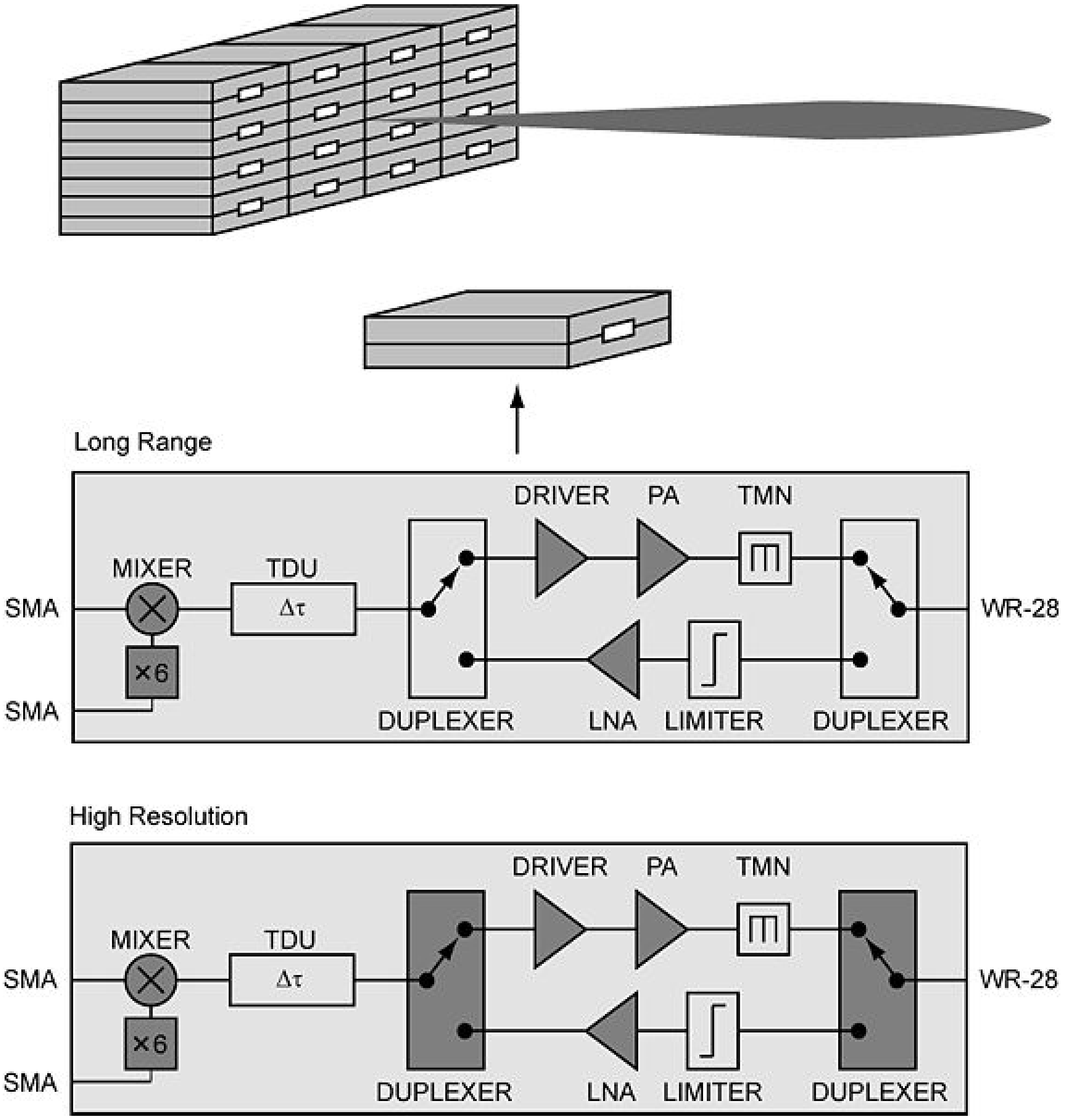


Fig. 7: RF MEMS components in a T/R module.