A (not so) short introduction to MEMS

Franck CHOLLET, Haobing LIU

FEMTO-ST Institute, Université de Franche-Comté, France
Please note that this work is published under a:

[Creative Commons Attribution-NonCommercial 2.5]

You are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution.

- Noncommercial.
  You may not use this work for commercial purposes.

- For any reuse or distribution, you must make clear to others the license terms of this work.

- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code [http://creativecommons.org/licenses/by-nc/2.5/legalcode].
# Contents

## 1 Why MEMS?
  1.1 What is MEMS and comparison with microelectronics . . . . . 5
  1.2 Why MEMS technology . . . . . . . . . . . . . . . . . . . . 6
    1.2.1 Advantages offered . . . . . . . . . . . . . . . . . . . 6
    1.2.2 Diverse products and markets . . . . . . . . . . . . . 7
    1.2.3 Economy of MEMS manufacturing and applications . . . . 9
  1.3 Major drivers for MEMS technology . . . . . . . . . . . . . . 11
  1.4 Mutual benefits between MEMS and microelectronics . . . . . 13

## 2 Introduction to MEMS design
  2.1 Physical scaling laws . . . . . . . . . . . . . . . . . . . . . 15
  2.2 The principles of design and reliability . . . . . . . . . . . . . 18
  2.3 MEMS design tools . . . . . . . . . . . . . . . . . . . . . . . 21

## 3 How MEMS are made
  3.1 Overview of MEMS fabrication process . . . . . . . . . . . . . . 23
  3.2 The MEMS materials . . . . . . . . . . . . . . . . . . . . . . 27
  3.3 Bulk micromachining, wet and dry etching . . . . . . . . . . . 29
    3.3.1 Isotropic and anisotropic wet etching . . . . . . . . . . 29
    3.3.2 Dry etching . . . . . . . . . . . . . . . . . . . . . . . . 36
    3.3.3 Wafer bonding . . . . . . . . . . . . . . . . . . . . . . 37
  3.4 Surface micromachining and thin-films . . . . . . . . . . . . . . 39
    3.4.1 Thin-film fabrication . . . . . . . . . . . . . . . . . . . 40
    3.4.2 Design limitation . . . . . . . . . . . . . . . . . . . . . 56
    3.4.3 Microstructure release . . . . . . . . . . . . . . . . . . 58
  3.5 DRIE micromachining . . . . . . . . . . . . . . . . . . . . . . 60
  3.6 Other microfabrication techniques . . . . . . . . . . . . . . . 64
    3.6.1 Micro-molding and LIGA . . . . . . . . . . . . . . . . . 64
    3.6.2 Polymer MEMS . . . . . . . . . . . . . . . . . . . . . . . 64
Chapter 1

Why MEMS?

1.1 What is MEMS and comparison with microelectronics

Micro Electro Mechanical Systems or MEMS is a term coined around 1989 by Prof. R. Howe [1] and others to describe an emerging research field, where mechanical elements, like cantilevers or membranes, had been manufactured at a scale more akin to microelectronics circuit than to lathe machining. But MEMS is not the only term used to describe this field and from its multicultural origin it is also known as Micromachines, a term often used in Japan, or more broadly as Microsystem Technology (MST), in Europe. However, if the etymology of the word is more or less well known, the dictionaries are still mum about an exact definition. Actually, what could link inkjet printer head, video projector DLP system, disposable bio-analysis chip and airbag crash sensor - yes, they are all MEMS, but what is MEMS?

It appears that these devices share the presence of features below 100 µm that are not machined using standard machining but using other techniques globally called micro-fabrication technology. Of course, this simple definition would also include microelectronics, but there is a characteristic that electronic circuits do not share with MEMS. While electronic circuits are inherently solid and compact structures, MEMS have holes, cavity, channels, cantilevers, membranes, etc, and, in some way, imitate ‘mechanical’ parts. This has a direct impact on their manufacturing process. Actually, even when MEMS are based on silicon, microelectronics process needs to be adapted to cater for thicker layer deposition, deeper etching and to introduce special steps to free the mechanical structures. Then, many more MEMS are not based on silicon and can be manufactured in polymer, in glass, in quartz or even in metals...
Thus, if similarities between MEMS and microelectronics exist, they now clearly are two distinct fields. Actually, MEMS needs a completely different set of mind, where next to electronics, mechanical and material knowledge plays a fundamental role.

1.2 Why MEMS technology

1.2.1 Advantages offered

The development of a MEMS component has a cost that should not be underestimated, but the technology has the possibility to bring unique benefits. The reasons that prompt the use of MEMS technology can be classified broadly in three classes:

miniaturization of existing devices For example the production of silicon based gyroscope which reduced existing devices weighting several kg and with a volume of 1000 cm$^3$ to a chip of a few grams contained in a 0.5 cm$^3$ package.

using physical principles that do not work at larger scale A typical example is given by the biochips where electrical field are use to pump the reactant around the chip. This so called electro-osmotic effect based on the existence of a drag force in the fluid works only in channels with dimension of a fraction of one mm, that is, at micro-scale.

developing tools for operation in the micro-world In 1986 H. Rohrer and G. Binnig at IBM were awarded the Nobel price in physics for their work on scanning tunnelling microscope. This work heralded the development of a new class of microscopes (atomic force microscope, scanning near-field optical microscope...) that shares the presence of micromachined sharp micro-tips with radius below 50 nm. This microtool was used to position atoms in complex arrangement, writing Chinese character or helping verify some prediction of quantum mechanics. Another example of this class of MEMS devices at a slightly larger scale would be the development of micro-grippers to handle cells for analysis.

By far miniaturization is often the most important driver behind MEMS development. The common perception is that miniaturization reduces cost, by decreasing material consumption and allowing batch fabrication, but an important collateral benefit is also in the increase of applicability. Actually, reduced mass and size allow placing the MEMS in places where a traditional system won’t have been able to fit. Finally, these two effects concur to
increase the total market of the miniaturized device compared to its costlier and bulkier ancestor. A typical example is brought by the accelerometer developed as a replacement for traditional airbag triggering sensor and that is now used in many appliances, as in digital cameras to help stabilize the image or even in the contact-less game controller integrated inside the latest handphones.

However often miniaturization alone cannot justify the development of new MEMS. After all if the bulky component is small enough, reliable enough, and particularly cheap then there is probably no reason to miniaturize it. Micro-fabrication process cost cannot usually compete with metal sheet punching or other conventional mass production methods.

But MEMS technology allows something different, at the same time you make the component smaller you can make it better. The airbag crash sensor gives us a good example of the added value that can be brought by developing a MEMS device. Some non-MEMS crash sensors are based on a metal ball retained by a rolling spring or a magnetic field. The ball moves in response to a rapid car deceleration and shorts two contacts inside the sensor. A simple and cheap method, but the ball can be blocked or contact may have been contaminated and when your start your engine, there is no easy way to tell if the sensor will work or not. MEMS devices can have a built-in self-test feature, where a micro-actuator will simulate the effect of deceleration and allow checking the integrity of the system every time you start the engine.

Another advantage that MEMS can bring relates with the system integration. Instead of having a series of external components (sensor, inductor...) connected by wire or soldered to a printed circuit board, the MEMS on silicon can be integrated directly with the electronics. Whether it is on the same chip or in the same package it results in increased reliability and decreased assembly cost, opening new application opportunities.

As we see, MEMS technology not only makes the things smaller but often makes them better.

### 1.2.2 Diverse products and markets

The previous difficulty we had to define MEMS stems from the vast number of products that fall under the MEMS umbrella. The MEMS component currently on the market can be broadly divided in six categories (Table [1.1]), where next to the well-known pressure and inertia sensors produced by different manufacturer like Motorola, Analog Devices, SensoNor or Delphi we have many other products. The micro-fluidic application are best known for the inkjet printer head popularized by Hewlett Packard, but they also include the burgeoning bioMEMS market with micro analysis system like the
capillary electrophoresis system from Agilent or the DNA chips.
Optical MEMS includes the component for the fiber optic telecommunication like the switch based on a moving mirror produced by Sercalo. They also include the optical switch matrix that is now waiting for the recovery of the telecommunication industry. This component consists of 100s of micro-mirror that can redirect the light from one input fiber to one output fiber, when the fibers are arranged either along a line (proposed by the now defunct Optical Micro Machines) or in a 2D configuration (Lambda router from Lucent). Moreover MOEMS deals with the now rather successful optical projection system that is competing with the LCD projector. The MEMS products are based either on an array of torsional micro-mirror in the Texas Instruments Digital Light Processor (DLP) system or on an array of controllable grating as in the Grating Light Valve (GLV) from Silicon Light Machines.
RF MEMS is also emerging as viable MEMS market. Next to passive components like high-Q inductors produced on the IC surface to replace the hybridized component as proposed by MEMSCAP we find RF switches and soon micromechanical filters.
But the list does not end here and we can find micromachined relays (MMR) produced for example by Omron, HDD read/write head and actuator or even toys, like the autonomous micro-robot EMRoS produced by EPSON.

<table>
<thead>
<tr>
<th>Product category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure sensor</td>
<td>Manifold pressure (MAP), tire pressure, blood pressure..</td>
</tr>
<tr>
<td>Inertia sensor</td>
<td>Accelerometer, gyroscope, crash sensor...</td>
</tr>
<tr>
<td>Microfluidics /</td>
<td>Inkjet printer nozzle, micro-bio-analysis systems, DNA chips...</td>
</tr>
<tr>
<td>bioMEMS</td>
<td></td>
</tr>
<tr>
<td>Optical MEMS /</td>
<td>Micro-mirror array for projection (DLP), micro-grating array for projection (GLV), optical fiber switch, adaptive optics...</td>
</tr>
<tr>
<td>MOEMS</td>
<td></td>
</tr>
<tr>
<td>RF MEMS</td>
<td>High Q-inductor, switches, antenna, filter..</td>
</tr>
<tr>
<td>Others</td>
<td>Relays, microphone, data storage, toys...</td>
</tr>
</tbody>
</table>

Table 1.1: MEMS products example

In 2002 these products represented a market of about 3.2B$, with roughly
one third in inkjet printer nozzle, one third in pressure sensor and the rest split between inertia sensors, RF MEMS, optical MEMS, projection display chip and bioMEMS [2]. Of course the MEMS market overall value is still small compared to the 180B$ IC industry - but there are two aspects that still make it very interesting:

- it is expected to grow at an annual rate of 18% for the foreseeable future, much higher than any projection for IC industry;

- MEMS chips have a large leveraging effect, and in the average a MEMS based systems will have 8 times more value than the MEMS chip price (e.g., a DLP projector is about 10 times the price of a MEMS DLP chip).

This last point has created very large difference between market studies, whether they reported market for components alone or for systems. The number cited above are in the average of other studies and represent the market for the MEMS components alone.

### 1.2.3 Economy of MEMS manufacturing and applications

However large the number of opportunities is, it should not make companies believe that they can invest in any of these fields randomly. For example, although the RF MEMS market seems to be growing fueled by the appetite for smaller wireless communication devices, it seems to grow mostly through internal growth. Actually the IC foundries are developing their own technology for producing, for example, high-Q inductors, and it seems that an external provider will have a very limited chance to penetrate the market.

Thus, market opportunities should be analyzed in detail to eliminate the false perception of a large market, taking into consideration the targeted customer inertia to change and the possibility that the targeted customer himself develop MEMS based solution. In that aspect, sensors seems an easy target being simple enough to allow full development within small business unit and having a large base of customers - in the other hand, an optical switch matrix is riskier because its value is null without the system that is built by a limited number of customers, which, most probably, also have the capabilities to develop in-house the MEMS component...

Some MEMS products already achieve high volume and benefit greatly from the batch fabrication technique. For example more than 100 millions MEMS accelerometers are sold every year in the world - and with newer use coming,
this number is still growing fast. But large numbers in an open market in-
variably means also fierce competition and ultimately reduced prices. Long
are gone the days where a MEMS accelerometer could be sold 10$ a piece - it is now less than 2$ and still dropping. Currently, the next target is a
3-axis accelerometer in a single package for about 4$, so that it can really
enter the toys industry. Note that there may be a few exceptions to this rule.
Actually, if the number of unit sold is also very large, the situation with the
inkjet printer nozzle is very different. Canon and Hewlett Packard devel-
oped a completely new product, the inkjet printer, which was better than
earlier dot matrix printer, creating a captive market for its MEMS based sys-
tem. This has allowed HP to repeatedly top the list of MEMS manufacturer
with sales in excess of 600M$. This enviable success is unfortunately most
probably difficult to emulate.

But these cases should not hide the fact that MEMS markets are es-
sentially niche markets. Few product will reach the million unit/year mark
and in 2006 among the more than 300 companies producing MEMS only 18
had sales above 100m$/year. Thus great care should be taken in balanc-
ing the research and development effort, because the difficulty of developing
new MEMS from scratch can be daunting and the return low. For exam-
ple, although Texas Instruments is now reaping the fruit of its Digital Light
Processor selling between 1996 and 2004 more than 4 millions chips for a
value now exceeding 200m$/year, the development of the technology by L.
Hornbeck took more than 10 years [3]. Few startup companies will ever have
this opportunity.

Actually it is not clear for a company what is the best approach for en-
tering the MEMS business, and we observe a large variety of business model
with no clear winner. For many years in microelectronics industry the abun-
dance of independent foundries and packaging companies has made fabless
approach a viable business model. However it is an approach only favored by
a handful of MEMS companies, and it seems for good reasons. A good insight
of the polymorphic MEMS business can be gained by studying the company
MemsTech, now a holding listed on the Kuala Lumpur Mesdaq (Malaysia)
and having office in Detroit, Kuala Lumpur and Singapore.
Singapore is actually where everything started in the mid-90’s for MemsTech
with the desire from an international company (EG&G) to enter the MEMS
sensor market. They found a suitable partner in Singapore at the Institute
of Microelectronics (IME), a research institute with vast experience in IC
technology.
This type of cooperation has been a frequent business model for MNC willing
to enter MEMS market, by starting with ex-house R&D contract develop-
ment of a component. EG&G and IME designed an accelerometer, patenting
along the way new fabrication process and developing a cheap plastic packaging process. Finally the R&D went well enough and the complete clean room used for the development was spun-off and used for the production of the accelerometer.

Here, we have another typical startup model, where IP developed in research institute and university ends up building a company. This approach is very typical of MEMS development, with a majority of the existing MEMS companies having been spun-off from a public research institute or a university. A few years down the road the fab continuously produced accelerometer and changed hands to another MNC before being bought back in 2001 by its management. During that period MemsTech was nothing else but a component manufacturer providing off-the-shelf accelerometer, just like what Motorola, Texas Instrument and others are doing.

But after the buyout, MemsTech needed to diversify its business and started proposing fabrication services. It then split in two entities: the fab, now called Sensfab, and the packaging and testing unit, Senzpak. Three years later, the company had increased its ‘off-the-shelf’ product offering, proposing accelerometer, pressure sensor, microphones and one IR camera developed in cooperation with local and overseas university.

This is again a typical behavior of small MEMS companies where growth is fueled by cooperation with external research institutions. Still at the same time MemsTech proposes wafer fabrication, packaging and testing services to external companies. This model where products and services are mixed is another typical MEMS business model, also followed by Silicon Microstructures in the USA, Colybris in Switzerland, MEMSCAP in France and some other. Finally, in June 2004 MemsTech went public on the Mesdaq market in Kuala Lumpur.

The main reason why the company could survive its entire series of avatar, is most probably because it had never overgrown its market and had the wisdom to remain a small company, with staff around 100 persons. Now, with a good product portfolio and a solid base of investor it is probably time for expansion.

1.3 Major drivers for MEMS technology

From the heyday of MEMS research at the end of the 1960s, started by the discovery of silicon large piezoresisitive effect by C. Smith[4] and the demonstration of anisotropic etching of silicon by J. Price[5] that paved the way to the first pressure sensor, one main driver for MEMS development has been the automotive industry. It is really amazing to see how many MEMS sensor
a modern car can use! From the first oil pressure sensors, car manufacturer quickly added manifold and tire pressure sensors, then crash sensors, one, then two and now up to five accelerometers. Recently the gyroscopes made their apparition for anti-skidding system and also for navigation unit - the list seems without end.

Miniatized pressure sensors were also quick to find their ways in medical equipment for blood pressure test. Since then biomedical application have drained a lot of attention from MEMS developer, and DNA chip or micro-analysis system are the latest successes in the list. Because you usually sell medical equipment to doctors and not to patients, the biomedical market has many features making it perfect for MEMS: a niche market with large added value.

Actually cheap and small MEMS sensors have many applications. Digital cameras have been starting using accelerometer to stabilize image, or to automatically find image orientation. Accelerometers are also being used in new contactless game controller or mouse.

These two later products are just a small part of the MEMS-based system that the computer industry is using to interface the arid beauty of digits with our human senses. The inkjet printer, DLP based projector, head-up display with MEMS scanner are all MEMS based computer output interfaces. Additionally, computer mass storage uses a copious amount of MEMS, for example, the hard-disk drive nowadays based on micromachined GMR head and dual stage MEMS micro-actuator. Of course in that last field more innovations are in the labs, and most of them use MEMS as the central reading/writing element.

The optical telecommunication industry has fueled the biggest MEMS R&D effort so far, when at the turn of the millennium, 10 s of companies started developing optical MEMS switch and similar components. We all know too well that the astounding 2D-switch matrix developed by Optical Micro Machines (OMM) and the 3D-matrix developed in just over 18 months at Lucent are now bed tale stories. However within a few years they placed optical MEMS as a serious contender for the future extension of the optical network, waiting for the next market rebound. Wireless telecommunications are also using more and more MEMS components. MEMS are slowly sipping into handphone replacing discrete elements one by one, RF switch, microphone, filters - until the dream of a 1 mm$^3$ handphone becomes true (with vocal recognition for numbering of course!). The latest craze seems to be in using accelerometers (again) inside handphone to convert them into game controller, the ubiquitous handphone becoming even more versatile.

Large displays are another consumer product that may prove to become a large market for MEMS. Actually, if plasma and LCD TV seems to become
more and more accepted, their price is still very high and recently vendors start offering large display based on MEMS projector at about half the price of their flat panel cousin. Projector based system can be very small and yet provide large size image. Actually, for the crown of the largest size the DLP projecting system from TI is a clear winner as evidenced by the digital cinema theaters that are burgeoning all over the globe. For home theater the jury is still debating - but MEMS will probably get a good share at it and DLP projector and similar technologies won’t be limited to PowerPoint presentation.

Finally, it is in the space that MEMS are finding an ultimate challenge and already some MEMS sensors have been used in satellite. The development of micro (less than 100kg) and nano (about 10kg) satellites is bringing the mass and volume advantage of MEMS to good use and some project are considering swarms of nano-satellite each replete with micromachined systems.

1.4 Mutual benefits between MEMS and microelectronics

The synergies between MEMS development and microelectronics are many. Actually MEMS clearly has its roots in microelectronics, as H. Nathanson at Westinghouse reported in 1967 the “resonant gate transistor” [6], which is now considered to be the first MEMS. This device used the resonant properties of a cantilevered beam acting as the gate of a field-effect transistor to provide electronic filtering with high-Q. But even long after this pioneering work, the emphasis on MEMS based on silicon was clearly a result of the vast knowledge on silicon material and on silicon based microfabrication gained by decades of research in microelectronics. Even quite recently the SOI technology developed for ICs has found a new life with MEMS. But the benefit is not unilateral and the MEMS technology has indirectly paid back this help by nurturing new electronic product. MEMS brought muscle and sight to the electronic brain, enabling a brand new class of embedded system that could sense, think and act while remaining small enough to be placed everywhere.

As a more direct benefit, MEMS can also help keep older microelectronics fab running. Actually MEMS devices most of the times have minimum features size of a several µm, allowing the use of older generation IC fabrication equipment that otherwise will have just been dumped. It is even possible to convert a complete plant and Analog Devices has redeveloped an older BiC-
MOS fabrication unit to successfully produce their renowned smart MEMS accelerometer. Moreover, as we have seen, MEMS component often have small market and although batch fabrication is a must, a large part of the MEMS production is still done using 100 mm (4”) and 150 mm (6”) wafers - and could use 5-6 years old IC production equipment. But this does not mean that equipment manufacturer cannot benefit from MEMS. Actually MEMS fabrication has specific needs (deeper etch, double side alignment, wafer bonding, thicker layer...) with a market large enough to support new product line. For example, firms like STS and Alcatel-Adixen producing MEMS deep RIE or EVGroup and Suss for their wafer bonder and double side mask aligner have clearly understood how to adapt their know-how to the MEMS fabrication market.
Chapter 2

Introduction to MEMS design

2.1 Physical scaling laws

The large decrease in size during miniaturization, that in some case can reach 1 or 2 orders of magnitude, has a tremendous impact on the behavior of micro-object when compared to their larger size cousin. We are already aware of some of the most visible implications of miniaturization. Actually nobody will be surprised to see a crumb stick to the rubbed surface of a plastic rod, whereas the whole bread loaf is not. Everybody will tell that it works with the crumb and not with the whole loaf because the crumb is lighter. Actually it is a bit more complicated than that.

The force that is attracting the crumb is the electrostatic force, which is proportional to the amount of charge on the surface of the crumb, which in turn is proportional to its surface. Thus when we shrink the size and go from the loaf to the crumb, we not only decrease the volume and thus the mass but we also decrease the surface and thus the electrostatic force. However, because the surface varies as the square of the dimension and the volume as the cube, this decrease in the force is relatively much smaller than the drop experienced by the mass. Thus finally not only the crumb mass is smaller, but, what is more important, the force acting on it becomes proportionally larger - making the crumb really fly!

To get a better understanding, we can refer to Figure 2.1 and consider a cube whose side goes from a length of 10 to a length of 1. The surface of the bigger cube is $6 \times 10 \times 10 = 600$ whereas its volume is $10 \times 10 \times 10 = 1000$. But now what happen to the scaled down cube? Its surface is $6 \times 1 \times 1 = 6$ and has been divided by 100 but its volume is $1 \times 1 \times 1 = 1$ and has been divided by 1000. Thus the volume/surface ratio has also shrunk by a factor of 10, making the surface effect proportionally 10 times larger with the smaller
cube than with the bigger one. This decrease of volume/surface ratio has profound implications for the design of MEMS. Actually it means that at a certain level of miniaturization, the surface effect will start to be dominant over the volume effects. For example, friction force (proportional to surface) will become larger than inertia (proportional to mass hence to volume), heat dissipation will become quicker and heat storage reduced: energy storage will become less attractive than energy coupling... This last example is well illustrated by one of the few ever built micromachines, the EMRoS micro-robot from Epson. The EMRoS (Epson Micro Robot System) is not powered with a battery (which stores energy proportional to its volume and becomes less interesting at small scale) but with solar cells whose output is clearly proportional to surface.

Then of course we can dwell into a more elaborate analysis of nature laws and try to see apart from geometrical factor what happens when we shrink the scale? Following an analysis pioneered by W. Trimmer [7], we may describe the way physical quantities vary with scale as a power of an arbitrary scale variable, $s$. We have just seen that volume scale as $s^3$, surface as $s^2$ and the volume/surface ratio as $s^1$. In the same vein we may have a look at different forces and see how they scale down (Table 2.1).

<table>
<thead>
<tr>
<th>Force</th>
<th>Scaling law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface tension</td>
<td>$s^1$</td>
</tr>
<tr>
<td>Electrostatic, Pressure, Muscle</td>
<td>$s^2$</td>
</tr>
<tr>
<td>Magnetic</td>
<td>$s^3$</td>
</tr>
<tr>
<td>Gravitational</td>
<td>$s^4$</td>
</tr>
</tbody>
</table>

Table 2.1: Scaling of nature forces.
From this table it appears that some forces that are insignificant at large scale becomes predominant at smaller scale. For example we see that gravity, which scales as $s^4$ (that is decrease by a factor 10,000 when the scale is shrunk by 10) will be relatively weak at micro-scale. However a more favorable force will be the tension force, which decrease as $s^1$ making it an important (and often annoying for non-fluidic application) force at micro-scale. The table also reveals that the electrostatic force will become more interesting than the magnetic force as the scale goes down. Of course this simple description is more qualitative than quantitative. Actually if we know that as the size shrinks the electrostatic force will finally exceed the magnetic force, a more detailed analysis is needed to find if it is at a size of 100 $\mu$m, 1 $\mu$m or 10 nm. In that particular case it has been shown that the prediction becomes true when the dimensions reach a few $\mu$m, right in the scale of MEMS devices. This has actually been the driver behind the design of the first electrostatic motors by R. Howe and R. Muller [8].

A more surprising consequence of miniaturization is that, contrary to what we would think at first, the relative manufacturing accuracy is sharply decreasing. This was first formalized by M. Madou [10] and it is indeed interesting to see that the relative accuracy of a MEMS device is at a few % not much better than standard masonry. Actually, if it is true that the absolute accuracy of MEMS patterning can reach 1 $\mu$m, the MEMS size is in the 10 $\mu$m-100 $\mu$m, meaning a relative patterning accuracy of 1%-10% or even less. We are here very far from single point diamond turning or the manufacturing of large telescope mirror that can both reach a relative accuracy of 0.0001%. So, ok, we have a low relative accuracy, but what does that mean in practice? Let’s take as an example the stiffness of a cantilever beam. From solid mechanics the stiffness, $k$, depends on the beam cross-section shape and for a rectangular cross-section it is proportional to

$$k = \frac{E hw^3}{4 L^3}, \quad (2.1)$$

where $E$ is the elasticity modulus, $h$ is the beam thickness, $w$ its width and $L$ its length. For a nominal beam width of 2 $\mu$m with an absolute fabrication accuracy of $\pm 0.2 \mu$m the relative accuracy is $\pm 10\%$. The stiffness for bending along the width direction varies as a power of 3 of the width and will thus have a relative accuracy of $\pm 30\%$. For a stiffness nominal value of 1 N/m, it means that the expected value can be anywhere between 0.7 N/m and 1.3 N/m - this is almost a variation by a factor of two! Our design needs to be tolerant to such variation, or the yield will be very low. In this particular case, one could improve the relative accuracy by taking advantage of the relatively constant absolute fabrication accuracy and increase the beam width nominal value to
4 \mu m - of course it also means doubling its length if one wants to keep the same spring constant.

### 2.2 The principles of design and reliability

Since the first days of pressure sensor development, MEMS designers have had to face the complexity of designing MEMS. Actually if IC design relies on an almost complete separation between fabrication process, circuit layout design and packaging, the most successful MEMS have been obtained by developing these three aspects simultaneously (Figure 2.2).

Actually MEMS fabrication process is so much intertwined with the device operation that MEMS design often involve a good deal of process development. If it is true that some standard processes are proposed by a few foundries (e.g., SOI process, and 3 layer surface micromachining by Memscap, epitaxy with buried interconnect by Bosch...), there is in MEMS nothing as ubiquitous as the CMOS process.

The success of the device often depends on physics, material property and the choice of fabrication techniques. Actually some industry observers are even claiming that in MEMS the rule is “One Product, One Process” - and many ways to achieve the same goal. Actually we are aware of at least five completely different processes that are currently used to fabricate commercial MEMS accelerometer with about the same characteristics and price - and for at least two companies the accelerometer is their only MEMS product.

And what about packaging then, the traditional back-end process? In MEMS it can account for more than 50% of the final product price and obviously should not be ignored. Actually the designer has to consider the packaging aspect too, and there are horror stories murmured in the industry where products had to be completely redeveloped after trials for packaging went...
unsuccessful. The main issues solved by MEMS packaging are less related with heat dissipation than with stress, hermetic encapsulation and often chip alignment and positioning. If chip orientation for IC is usually not a concern, it becomes one for single-axis MEMS accelerometer where the chip has to be aligned precisely with respect to the package. This may imply the use of alignment mark, on the MEMS and in the package. In other case the chip may need to be aligned with external access port. Actually MEMS sensors often need an access hole in the package to bring air or a liquid in contact with the sensing chip, complicating substantially the packaging. One of the innovative approaches to this problem has been to use a first level packaging during the fabrication process to shield the sensitive parts, finally linking the back-end with the front-end. Even for MEMS that do not need access to the environment, packaging can be a complex issue because of stress.

MEMS often use stress sensitive structure to measure the deformation of a member and the additional stress introduced during packaging could affect this function. Motorola solved this problem with its line of pressure sensor by providing calibration of the device after packaging - then any packaging induced drift will be automatically zeroed.

This kind of solution highlights the need to practice design for testing. In the case of Motorola this resulted in adding a few more pins in the package linked to test point to independently tweak variable gain amplifier. This cannot be an afterthought, but need to be taken into consideration early.

How will you test your device? At wafer level, chip level or after packaging? MEMS require here again much different answers than ICs. Understandably it will be difficult to find all the competence needed to solve these problems in one single designer, and good MEMS design will be teamwork with brainstorming sessions, trying to find the best overall solution. MEMS design cannot simply resume to a sequence of optimized answer for each of the individual process, device and packaging tasks - success will only come from a global answer to the complete system problem.

An early misconception about MEMS accelerometer was that these small parts with suspension that were only a few µm wide would be incredibly fragile and break with the first shock. Of course it wasn’t the case, first because silicon is a wonderful mechanical material tougher than steel and then because the shrinking dimension implied a really insignificant mass, and thus very little inertia forces. But sometime people can be stubborn and seldom really understand the predictive nature of the law of physics, preferring to trust their (too) common sense. Analog Devices was facing the hard task to convince the army that their MEMS based accelerometer could be used in military system, but it quickly appeared that it had to be a more direct proof than some equations on a white board. They decided to equip a mortar
shell with an accelerometer and a telemetry system, and then fired the shell. During flight, the accelerometer measured a periodic signal, that was later traced back to the natural wobbling of the shell. Then the shell hit his target and exploded. Of course the telemetry system went mum and the sensor was destroyed. However, the 'fragile' sensing part was still found in the debris... and it wasn’t broken.

In another example, the DLP chip from Texas Instruments has mirrors supported by torsion hinge 1 µm wide and 60 nm thick that clearly seems very prone to failure. TI engineers knew it wasn’t a problem because at this size the slippage between material grains occurring during cyclic deformation is quickly relieved on the hinge surface, and never build-up, avoiding catastrophic failure. But, again, they had to prove their design right in a more direct way. TI submitted the mirrors of many chips through 3 trillions ($10^{12}$) cycles, far more that what is expected from normal operation... and again not a single of the 100 millions tested hinges failed.

Of course, some designs will be intrinsically more reliable than other and following a taxonomy introduced by P. McWhorter, at Sandia National Laboratory [11], MEMS can be divided in four classes, with potentially increasing reliability problems.

<table>
<thead>
<tr>
<th>Class</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>No moving parts</td>
<td>Moving parts, no rubbing and impacting parts</td>
<td>Moving parts, impacting surfaces</td>
<td>Moving parts, impacting and rubbing surfaces</td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td>Accelerometer, Pressure sensor, High-Q inductor, Inkjet nozzle...</td>
<td>Gyroscopes, Resonator, Filter...</td>
<td>TI DLP, Relay, Valve, Pump...</td>
<td>Optical switch, scanner, locking system</td>
</tr>
</tbody>
</table>

Table 2.2: Taxonomy for evaluating MEMS devices reliability

By looking at this table it becomes clearer why developing the Texas Instruments DLP took many more years than developing accelerometer - the reliability of the final device was an issue and for example, mirrors had originally a tendency to stick to the substrate during operation. TI had to
go through a series of major improvements in the material and in the design to increase the reliability of their first design.

2.3 MEMS design tools

As we have seen miniaturization science is not always intuitive. What may be true at large scale may become wrong at smaller scale. This translates into an immediate difficulty to design new MEMS structure following some guts feeling. Our intuition may be completely wrong and will need to be backed up by accurate modeling. However simulation of MEMS can become incredibly complex and S. Senturia describes a multi-tiered approach that is more manageable [12] as shown in Figure 2.3.

Some simulation tools like Intellisuite by Intellisense or Coventorware by Coventor have been specifically devised for MEMS. They allow accurate modeling using meshing method (FEM, BEM) to solve the partial differential equation that describe a device in different physical domains. Moreover, they try to give a complete view of the MEMS design, which, as we said before, is material and process dependent, and thus they give access to material and process libraries. In this way it is possible to build quickly 3D model of MEMS from the mask layout using simulated process. However MEMS process simulation is still in its infancy and the process simulator is used as a simple tool to build quickly the simulation model from purely geometrical consideration, but cannot yet be used to optimize the fabrication process. One exception will be the simulation of anisotropic etching of silicon

Figure 2.3: MEMS multi-tiered simulation (adapted from [12] and expanded).
and some processes modeled for IC development (oxidation, resist development...) where the existing TCAD tools (SUPREM, etc) can be used. Complete MEMS devices are generally far too complex to be modeled entirely ab initio, and generally reduced models have to be used. For example, behavioral simulation is used by MEMSPPro from MemsCap where ANSYS is used to generate the reduced model, which then is run in circuit-analysis software like Spice. Sugar from C. Pister’s group at UC Berkeley is also based on lumped analysis of behavioral model, but the decomposition of the structure in simpler element is left to the designer. Still, although the actual tendency is to use numerical modeling extensively, it is our opinion that no good device modeling can be devised without a first analytic model based on algebraic equation. Developing a reduced order model based on some analytic expression help our intuition regains some of its power. For example, seeing that the stiffness varies as the beam width to the cube makes it clearer how we should shrink this beam: if the width is divided by a bit more than two, the stiffness is already ten times smaller. This kind of insight is invaluable. The analytic model devised need of course to be verified with a few examples using numerical simulation.

Finally the system level simulation is often not in the hand of the MEMS designer, but here block diagram and lumped model can be used, with only a limited set of key state variable. This model may then include the electronics and the MEMS device will be represented by one or more blocks, reusing the equation derived for the behavioral model.
Chapter 3

How MEMS are made

3.1 Overview of MEMS fabrication process

Micro-fabrication is the set of technologies used to manufacture structures with micrometric features. This task can unfortunately not rely on the traditional fabrication techniques such as milling, drilling, turning, forging and casting because of the scale. The fabrication techniques had thus to come from another source. As MEMS devices have about the same feature size as integrated circuits, MEMS fabrication technology quickly took inspiration from microelectronics. Techniques like photolithography, thin film deposition by chemical vapor deposition (CVD) or physical vapor deposition (PVD), thin film growth by oxidation and epitaxy, doping by ion implantation or diffusion, wet etching, dry etching, etc have all been adopted by the MEMS technologists. Standard book on microelectronics describe in details these techniques but, as MEMS and IC fabrication goals are different, some of these techniques have evolved as they were applied to MEMS and we will detail here their new capabilities. Moreover, MEMS has spurred many unique fabrication techniques that we will also describe in our panorama of MEMS fabrication introducing bulk micromachining, surface micromachining, LIGA, etc [18].

In general, MEMS fabrication tries to use batch process to benefit from the same economy of scale that is so successful in reducing the cost of ICs. As such, a typical fabrication process starts with a wafer (silicon, polymer, glass...) that may play an active role in the final device or may only be a substrate on which the MEMS is built. This wafer is processed with a succession of processes (Table 3.1) that add, modify or remove materials along precise patterns. The patterns (or the layout) is decided by the designer depending on the desired function but, for most materials, it is difficult to directly
Additive process | Modifying process | Subtractive process
---|---|---
Evaporation | Oxydation | Wet etching
Sputtering | Doping | Dry etching
CVD | Annealing | Sacrificial etching
Spin-coating | UV exposure | Development
... | ... | ...

Table 3.1: Process classification.

deposit or modify them locally. In fact there are few processes equivalent to turning or milling in micromachining. Focused Ion Beam (FIB), where a beam of high energy ion can be scanned to remove most materials and deposit some, can perform even down to nanoscale but the sequential processing approach it requires (as opposed to batch processing) is not cost effective for production. Thus the problem of patterning a material is generally split in two distinct steps: first, deposition and patterning of a surrogate layer that can be easily modified locally and then transfer of the pattern to the material of interest.

![Photoresist coating](photoresist_coating.png) → ![Mask aligning](mask_aligning.png) → ![UV-exposure](uv_exposure.png)

Figure 3.1: Photo-patterning in positive and negative photoresist.

In the most common process called photo-patterning, the surrogate layer used is a special solution (called a photoresist) which contains a polymer sensitive to UV-photon action (Figure 3.1). The liquid photoresist is first coated on the substrate as a thin-film and the solvent is then evaporated by baking on a hotplate leaving a solid layer of polymer with very uniform
thickness. The substrate is then brought to the mask aligner tool, where the patterning process will take place. The photoresist film is exposed to UV radiation through a mask which has been precisely aligned with the substrate. The mask has clear and opaque regions according to the desired pattern, the clear regions allowing the photoresist to be exposed to UV radiation and modifying it locally. The exposure creates a latent image of the mask feature in the surrogate layer. The contrast of this image may be enhanced by heat, which accelerates the chemical reaction initiated by the UV-exposure. To finish the process this latent image is revealed in a special chemical solution, the developer. Actually the exposure changes the solubility of the photoresist in the developer and the exact change of solubility depends on the type of photoresist used originally: for so-called positive photoresist the exposed region becomes more soluble in the developer, while for negative photoresist the reverse happens and the exposed region becomes insoluble. After development, the surrogate layer patterned over the whole surface of the wafer can be used for pattern transfer.

They are actually two main techniques that can be used to transfer the pattern: etching and lift-off (Figure 3.2). With etching the patterned layer allows protecting locally the underlying material. The material in the unprotected regions is then attacked physically or chemically before we finally remove the protective layer. For lift-off, the material is deposited on top of the patterned layer. Complete dissolution of this layer (used actually as a

![Figure 3.2: Pattern transfer by etching and lift-off.](image-url)
sacrificial layer) simultaneously washes away the deposited material it was supporting, leaving the deposited material only in the open regions of the pattern.

Combination of photo-patterning and etching is known as photolithography and is nowadays the most common techniques for micro-fabrication, lying at the roots of the IC revolution. An important step in this process is the fabrication of the photolithographic mask, which can clearly not itself use photolithography with a mask! Its fabrication is actually based on etching of a chromium film deposited on a transparent substrate. The pattern is obtained inside a pattern generator system where the photoresist film is exposed point by point by a beam (electron or laser) which is slowly scanned over the whole mask surface while its intensity is modulated by a computer control system. This process reaches a very high resolution (electron beam can be focused in spot as small as 10nm) but it is slow. This remains an acceptable trade-off for mask production, as the photolithographic step requires to produce a unique mask which is then repeatedly used to expose 1000’s of wafers.

Recently we have seen the emergence of new patterning techniques that try to reduce the cost attached to photo-patterning at small scale where there is a need to use deep-UV source with complex optics using immersion lens and systems under vacuum. The most promising ones are based on imprinting, where the desired pattern on a stamp is pressed against a protective resin film (hot-embossing, nano-imprint lithography, UV-NIL...). The resulting patterned layer can then be used with lithography or lift-off for pattern transfer.

This succession of patterning/patter transfer is repeated a certain number of times on different layers. Usually only a few layers are patterned and stacked together, but the production of more complex MEMS may use a stack of about 10 different layers - still a far cry from the IC process that could necessitate more than 30 different patterning steps.

MEMS production (Figure 3.3) does not end with wafer fabrication, the so-called front-end process, and there are, like in Electronics, back-end processes... with a twist. For example, when ‘mechanical’ parts (mobile elements, channels, cavities...) exist in the device, the processed wafer passes through a special step called ‘Release’ to free these parts. This step may happen before or after the wafer is diced in individual chips, just before the more traditional assembly, packaging and final tests.
3.2 The MEMS materials

The choice of a good material for MEMS application is no more based like in microelectronics on carrier mobility, but on more mechanical aspect: small or controllable internal stress, low processing temperature, compatibility with other materials, possibility to obtain thick layer, patterning possibilities... In addition, depending on the field of application, the material often needs to have extra properties. RF MEMS will want to be based on material with small loss tangent (for example high resistivity silicon), optical MEMS may need a transparent substrate, BioMEMS will need bio-compatibility, if not for the substrate, for a coating adhering well to the substrate, sensor application will need a material showing piezoresistance or piezoelectricity, etc.

Actually, because the issue of material contamination is much less important in MEMS than in IC fabrication, the MEMS designer often tries to use the material presenting the best properties for his unique application.

Still, from its microelectronics’ root MEMS has retained the predominant use of silicon and its compounds, silicon (di)oxide (SiO$_2$) and silicon nitride (Si$_x$N$_y$). But actually, it was not purely coincidental because silicon is, as K. Petersen explained in a famous paper [19], an excellent mechanical material. Actually, silicon is almost as strong but lighter than steel, has large critical stress and no elasticity limit at room temperature as it is a perfect crystal ensuring that it will recover from large strain. Unfortunately it is brittle and this may pose problem in handling wafer, but it is rarely a source of failure for MEMS components. For sensing application silicon has a large piezoresistive coefficient, and for optical MEMS it is transparent at the common telecommunication wavelengths.

In addition silicon has a stable oxide easy to grow at elevated temperature that is transparent and thermally and electrically insulating. Actually this oxide has the smallest coefficient of thermal expansion of all known materi-
als. Those properties are often put to good use during MEMS fabrication, where oxide support will be used to thermally insulate a pixel of an thermal camera for example.

Recently, a new substrate based on silicon and coming from IC industry has made its entry in the MEMS material list: the SOI (Silicon On Insulator) wafer. This substrate is composed of a thick silicon layer of several hundred $\mu$m (the handle), a thin layer of oxide of 1 or 2 $\mu$m and on top another silicon layer, the device layer. The thickness of this last layer is what differentiates the IC and the MEMS SOI wafers: in the first case it will reach at most a few $\mu$m where in the later case, the thickness can reach 100 $\mu$m or more. The high functionality of this material has allowed producing complete devices with very simple process, like the optical switch produced by Sercalo fabricated with its mobile mirror, actuator and fiber alignment feature with one single process step!

Another interesting compound is silicon nitride ($\text{Si}_x\text{N}_y$), which is stronger than silicon and can be deposited in thin layer with an excellent control of stress to produce 1 $\mu$m thick membrane of several cm$^2$. In general stoichiometric nitride film ($\text{Si}_3\text{N}_4$) will show tensile stress, but increasing the Si content will invariably ends in obtaining a compressive stress. A good control of stress is also obtained during deposition of poly-crystalline silicon. During LPCVD deposition, increasing the temperature from 560$^\circ$C to 620$^\circ$C lowers the as-deposited stress, changing the compressive stress usually present in polysilicon films to tensile stress [20]. A subsequent high temperature (>950$^\circ$C) anneal result in layer with very low stress, making Poly-Si the material of choice for building multi-layered structure on silicon surface. For example the Sandia National Lab’s ‘Summit V’ process stacks five layer of poly-silicon allowing an unparalleled freedom of design for complex MEMS structure. Closing the list of silicon compound we can add a newcomer, silicon carbide $\text{SiC}$. $\text{SiC}$ has unique thermal properties (albeit not yet on par with diamond) and has been used in high temperature sensor.

But silicon and its derivative are not the only choice for MEMS, many other materials are also used because they posses some unique properties. For example, other semiconductors like InP have also been micromachined mainly to take advantage of their photonics capabilities and serve as tunable laser source. Quartz crystal has strong piezoelectric effect that has been put into use to build resonant sensors like gyroscope or mass sensors. Biocompatibility will actually force the use of a limited list of already tested and approved material, or suggest the use of durable coating.

Glass is only second to silicon in its use in MEMS fabrication because it can easily form tight bond with silicon and also because it can be used to
obtain bio-compatible channels for BioMEMS.

Polymers are also often used for BioMEMS fabrication where they can be tailored to provide biodegradability or bioabsorbability. The versatility of polymers makes them interesting for other MEMS application, and for example the reflow appearing at moderate temperature has been used to obtain arrays of spherical microlenses for optical MEMS. This thermoplastic property also allows molding, making polymer MEMS a cheap alternative to silicon based system, particularly for micro-fluidic application. Recently the availability of photosensitive polymers like SU8 \[22\] than can be spun to thickness exceeding 100 µm and patterned with vertical sides has further increased the possibility to build polymer structure.

This quick introduction to MEMS materials needs to mention metals. If their conductivity is of course a must when they are used as electrical connection like in IC, metals can also be used to build structures. Actually, their ability to be grown in thin-films of good quality at a moderate temperature is what decided Texas Instruments to base the complete DLP micro-mirror device on a multi-layer aluminum process. In other applications, electroplated nickel will produce excellent micro-molds, whereas gold reflective properties are used in optical MEMS and nitinol (NiTi), presenting a strong shape memory effect, easily becomes a compact actuator.

### 3.3 Bulk micromachining, wet and dry etching

Bulk micromachining refers to the formation of micro structures by removal of materials from bulk substrates. The bulk substrate in wafer form can be silicon, glass, quartz, crystalline Ge, SiC, GaAs, GaP or InP. The subtractive process commonly used to remove excess material are wet and dry etching, allowing varying degree of control on the profile of the final structure.

#### 3.3.1 Isotropic and anisotropic wet etching

Wet etching is obtained by immersing the material in a chemical bath that dissolves the surfaces not covered by a protective layer. The main advantages of this subtractive technique are that it can be quick, uniform, very selective and cheap. The etching rate and the resulting profile depend on the material, the chemical, the temperature of the bath, the presence of agitation, and the etch stop technique used if any. Wet etching is usually divided between isotropic and anisotropic etching. Isotropic etching happens when the chemical etches the bulk material at the same rate in all directions, while
<table>
<thead>
<tr>
<th>Material</th>
<th>Young’s modulus</th>
<th>Poisson ratio</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stainless Steel</td>
<td>200 GPa</td>
<td>0.3</td>
<td>7900 kg/m³</td>
</tr>
<tr>
<td>Silicon (Si)</td>
<td>&lt;100&gt;130 GPa</td>
<td>0.25</td>
<td>2300 kg/m³</td>
</tr>
<tr>
<td></td>
<td>&lt;111&gt;187 GPa</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>PolySilicon (PolySi)</td>
<td>120-175 GPa</td>
<td>0.15-0.36</td>
<td>2300 kg/m³</td>
</tr>
<tr>
<td>Silicon Dioxide (SiO₂)</td>
<td>73 GPa</td>
<td>0.17</td>
<td>2500 kg/m³</td>
</tr>
<tr>
<td>Silicon Nitride (SiN)</td>
<td>340 GPa</td>
<td>0.29</td>
<td>3100 kg/m³</td>
</tr>
<tr>
<td>Glass (BK7)</td>
<td>(82) GPa</td>
<td>0.206</td>
<td>2500 kg/m³</td>
</tr>
<tr>
<td></td>
<td>(SF11) 66 GPa</td>
<td>0.235</td>
<td>4700 kg/m³</td>
</tr>
<tr>
<td>Gold (Au)</td>
<td>78 GPa</td>
<td>0.42</td>
<td>19300 kg/m³</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>70 GPa</td>
<td>0.33</td>
<td>2700 kg/m³</td>
</tr>
<tr>
<td>SU8</td>
<td>4.1 GPa</td>
<td>0.22</td>
<td>1200 kg/m³</td>
</tr>
<tr>
<td>PDMS</td>
<td>0.0004-0.0009 GPa</td>
<td>0.5</td>
<td>0.97 kg/m³</td>
</tr>
</tbody>
</table>

Table 3.2: Material properties.

Anisotropic etching happens when different etching rate exists along different directions.

However the etching rate never reaches 0, and it is actually impossible to obtain etching in only one direction. This is commonly quantified by estimating the underetch ($w/d$), that is the lateral etching under the edge of the mask with respect to the vertical etching, as shown in the figure. This parameter may range between 1 for isotropic etching to about 0.01 for very anisotropic etch, obtained for example by etching Silicon in a KOH bath. For substrates made of homogeneous and amorphous material, like glass, wet etching must be isotropic, although faster surface etching is sometimes observed. However, for crystalline materials, e.g. silicon, the etching is either isotropic or anisotropic, depending on the type of chemical used. In general, isotropic etchants are acids, while anisotropic etchants are alkaline bases.
Figure 3.4 compares isotropic and anisotropic wet etching of silicon. The top-left inset shows isotropic etching of silicon when the bath is agitated ensuring that fresh chemical constantly reaches the bottom of the trench and resulting in a truly isotropic etch. Isotropic wet etching is used for thin layer or when the rounded profile is interesting, to obtain channels for fluids for example. For silicon, the etchant can be HNA, which is a mixture of hydrofluoric acid (HF), nitric acid (HNO$_3$), and acetic acid (CH$_3$COOH). In HNA the nitric acid acts as an oxidant and HF dissolves the oxide by forming the water soluble H$_2$SiF$_6$. The two steps of the simplified reaction are:

$$\text{Si} + \text{HNO}_3 + \text{H}_2\text{O} \rightarrow \text{SiO}_2 + \text{HNO}_2 + \text{H}_2$$

$$\text{SiO}_2 + 6\text{HF} \rightarrow \text{H}_2\text{SiF}_6 + 2\text{H}_2\text{O}$$

The etching rate for silicon can reach 80 µm/min, and oxide can be used as mask material as its etch rate is only 30 to 80 nm/min. Etching under the mask edge or underetch is unavoidable with isotropic wet etching. Moreover, the etch rate and profile are sensitive to solution agitation and temperature, making it difficult to control the geometry of the deep etch usually needed for MEMS.

Anisotropic etching developed in the late 60s can overcome these problems. The lower part of Figure 3.4 shows features obtained by etching a (100) wafer with a KOH solution. The etched profile is clearly anisotropic, revealing planes without rounded shape and little underetch. Potassium hydroxide
(KOH), tetramethyl ammonium hydroxide (TMAH) and ethylene diamine pyrocatechol (EDP) are common chemicals used for anisotropic etching of silicon. The etching anisotropy has its roots in the different etch rates existing for different crystal planes usually because they present different density of atoms and hence of electrons.

Actually to describe planes and direction in a crystal we commonly use the Miller indices. In a crystal the system of coordinate normally used has its axes placed along the edge of the crystal lattice and, for example, in a cubic crystallographic system (as Si, AsGa...), the coordinate system is Cartesian with the usual $(x,y,z)$ axes. The positions of the atom are then identified by their coordinates, and for Si which crystallizes in the diamond cubic structure, the elementary cell has two atoms, one placed in $(0,0,0)$ and one in $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ which are repeated following a face-centered cubic (fcc) arrangement, one of the 14 possible Bravais arrangements in crystals. The crystal is then built by the repetition of this elementary lattice in all three directions, with the coordinate in the lattice normalized to the elementary lattice dimension. A plane can then be identified by indices $(hkl)$ which are obtained by considering the intersections of the plane with the crystal axes. Actually the number $h, k, l$ are always integers and are obtained by using the reciprocal of the intersection coordinates for the three axes and reducing them to the smallest possible integers by clearing common factors. If the integer is negative, it is represented by placing a bar on its top.

Three important crystal planes, the $(100)$ plane, $(110)$ plane and $(111)$ plane have been illustrated in Figure 3.5. For example the $(100)$ plane intercept the $x$ axis in $1$ (the reciprocal is $1/1 = 1$!), and along $y$ and $z$ the $0$ arises because in these cases the plane is parallel to the axis and thus will intercept it... at infinity - and because we take the reciprocal of the intercept coordinate we get $1/\infty = 0$. Note that if the plane intercepts the axes at the origin, it will need to be translated to be able to compute the indices.

The same indices are used to represent crystallographic direction as well. In this case the indices are obtained by considering the vector coordinate between two points of the lattice placed along the chosen direction. The coordinate are then reduced to the smallest set of integer and placed between brackets $[hkl]$ to differentiate them from the plane indices which were placed between parenthesis $(hkl)$.

Interestingly, the direction normal to the $(hkl)$ plane is the $[hkl]$ direction.
Moreover the angle $\alpha$ existing between two directions is given by taking the dot product between the two vectors:

$$\alpha = \cos^{-1} \frac{h_1 h_2 + k_1 k_2 + l_1 l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2} \sqrt{h_2^2 + k_2^2 + l_2^2}}$$

In general, crystal symmetries result in different directions having the same physical properties and there is no need to distinguish them. In this case we use the $<hkl>$ notation to represent any of these equivalent directions, whereas for plane with equivalent orientation we use the $\{hkl\}$ notation. For example, it is customary to give the crystallographic orientation of a silicon wafer by indicating the equivalent direction of the normal to the top surface. A $<100>$ wafer means that the normal direction is equivalent to the $[100]$ direction, which could be $[\bar{1}00]$ or even $[0\bar{1}0]$. In these different cases, the top surface of the wafer would be $(100)$, $(100)$ and $(010)$, series of plane that would present the same properties in the face-centered cubic lattice. For other lattices with less symmetries, care should be taken to use exact direction $[hkl]$ to indicate the precise crystallographic direction of the top wafer surface.

![Lattice points coordinate, planes and directions in the cubic lattice of silicon.](image)

Figure 3.5: Lattice points coordinate, planes and directions in the cubic lattice of silicon.

The anisotropy can be very large and for example, for silicon and KOH, the etch rate ratio can reach 400 between $(100)$ and $(111)$ planes and even 600 between $(110)$ and $(111)$ planes - meaning that when the etch rate for the $(100)$ plane is about $1 \mu\text{m/min}$ then the $(111)$ plane will etch at only
2.5 nm/min effectively allowing to consider it as an etch-stop plane. With different combinations of wafer orientations and mask patterns, very sophisticated structures such as cavities, grooves, cantilevers, through holes and bridges can be fabricated. For example, if the (100) wafers in Figure 3.4 shows an angle of 54.7° between the (111) plane and the surface, typically producing V-grooves, (110) oriented wafer will present an angle of 90° between these planes resulting in U-grooves with vertical walls. To obtain these grooves, the mask pattern edges need to be aligned with the edge of the (111) planes. For a (100) wafer it is simple because the groove edge are along the <110> direction, that is parallel to the main wafer flat. Moreover the four (111) planes intersect on the (100) surface at 90° and a rectangular pattern will immediately expose four sloping (111) planes and provide a simple way to obtain precisely defined pits and define square membranes. (110) wafers are more difficult to handle, and to obtain a U-groove the side should be tilted by an angle of 125.2° with respect to the <110> wafer flat. In addition, to obtain a four-sided pit, the two other sides should make a 55° angle with the flat direction - defining a non-rectangular pit that is seldom used for membranes.

If the control of the lateral etching by using the (111) planes is usually excellent, controlling the etching depth is more complicated. The first possibility is to use the self limiting effect appearing when two sloping (111) planes finally contact each other, providing the typical V-grooves of Figure 3.4. Controlling the etching time is another alternative. However producing the flat membranes of precise thickness needed for pressure sensors required a better approach that what can be achieved by this simple method. MEMS technologist have tackled this problem by developing different etch stop techniques that reduce by one or two order of magnitude the etch speed when the solution reach a particular depth.

The electrochemical etch stop works by first creating a diode junction for example by using epitaxial growth or doping of a n-layer over a p-substrate. Proper polarization of the substrate and the chemical bath allows for the etching to completely stop at the junction. This process yields an excellent control over the final membrane thickness that is only determined by the thickness of the epitaxial layer, and thus can be better than 1% over a whole wafer. Another popular method that does not require epitaxial growth is to heavily dope the surface of silicon with boron by diffusion or implantation, triggering a decrease of the etch rate by at least one order of magnitude. However, if diffusion from the surface is used to obtain the boron layer, the resulting high boron concentration (> 10^{19} cm^{-3}) at the surface will decrease substantially the piezoresistive coefficient value making piezoresistors less sensitive. Ion implantation can overcome this problem by burying the doped
layer a few \( \mu m \) under the surface, leaving a thin top layer untouched for the fabrication of the piezoresistors.

Actually, the seemingly simple membrane process often requires two tools specially designed for MEMS fabrication. Firstly, to properly align the aperture of the backside mask with the piezoresistor or other features on the front side (Figure 2.5) a double-side mask aligner is required. Different approaches have been used (infrared camera, image storage, folded optical path...) by the various manufacturers (Suss Microtec, OAI, EVGroup...) to tackle this problem, resulting in a very satisfying registration accuracy that can reach 1\( \mu m \) for the best systems. Secondly, etching the cavity below the membrane needs a special protection tool, that in the case of electrochemical etch stop is also used for ensuring the substrate polarization. Actually the presence of the cavity inevitably weakens the wafer and to avoid wafer breakage, the membrane is usually etched in the last step of the process. At that time, the front side will have already received metalization which generally cannot survive the prolonged etch and needs to be protected. This protection can be obtained by using a thick protective wax, but more often a cleaner process is preferred based on a mechanical chuck. The chuck is designed to allow quick loading and unloading operation, using O-ring to seal the front-side of the wafer and often includes spring loaded contacts to provide bias for electrochemical etch-stop.

The chemical used during anisotropic etching are usually strong alkaline bases and requires a hard masking material that can withstand the solution without decomposing or peeling. In general polymer (like photoresist) can not be used to protect the substrate, and if some metals can be used effectively, in general a non-organic thin-film is used. For example, silicon oxide mask is commonly used with TMAH, while silicon nitride is generally used with KOH. Table 3.3 summarizes the characteristics of some anisotropic etching solution.

Of course anisotropic wet etching has its limitation. The most serious one lies with the need to align the sides of the pattern with the crystal axes to benefit from the (111) plane etch-stop, severely limiting the freedom of layout. A typical example is when we want to design a structure with convex corners - that is instead of designing a pit, we now want an island. The island convex corners will inevitably expose planes which are not the (111) planes and will be etched away slowly, finally resulting in the complete disappearance of the island. Although techniques have been developed to slow down the etch rate of the corner by adding protruding ‘prongs’, these structures take space on the wafer and they finally cannot give the same patterning freedom as dry etching techniques.
<table>
<thead>
<tr>
<th>Solution</th>
<th>(100) Si etch rate (µm/min)</th>
<th>Etch rate ratio</th>
<th>Mask etch rate (nm/min)</th>
<th>Boron etch stop (cm⁻³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KOH / H₂O</td>
<td></td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44g / 100ml (30 wt.%) @ 85°C¹</td>
<td>400 for (100)/(111)</td>
<td>600 for (110)/(111)</td>
<td>3.5 (SiO₂) &lt; 0.01 (Si₃N₄)</td>
<td>&gt; 10²⁰ rate/20</td>
</tr>
<tr>
<td>TMAH / H₂O</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28g / 100ml (22 wt.%) @ 90°C²</td>
<td>30 for (100)/(111)</td>
<td>50 for (110)/(111)</td>
<td>0.2 (SiO₂) &lt; 0.01 (Si₃N₄)</td>
<td>4 · 10²⁰ rate/40</td>
</tr>
<tr>
<td>EDP</td>
<td></td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ethylene diamine / pyrocatechol / H₂O) 750ml / 120g / 240ml @ 115°C³</td>
<td>35 for (100)/(111)</td>
<td>≈0 (Au, Cr, Ag, Cu, Ta)</td>
<td>0.5 (SiO₂) 0.1 (Si₃N₄)</td>
<td>7 · 10¹⁹ rate/50</td>
</tr>
</tbody>
</table>

Notes:
1 + largest etch rate ratio; –K ions degrade CMOS; –etch SiO₂ fast
2 + SiO₂ mask; + CMOS compatible; – large overtech
3 + SiO₂ mask; + no metal etch; + CMOS compatible; – large overtech; – toxic

Table 3.3: Characteristics of some anisotropic etchants for silicon.

### 3.3.2 Dry etching

Dry etching is a series of subtractive methods where the solid substrate surface is removed by gaseous species. Plasma is usually involved in the process to increase etching rate and supply reacting ions and radicals. The etching can be conducted physically by ion bombardment (ion etching or sputtering and ion-beam milling), chemically through a chemical reaction occurring at the solid surface (plasma etching or radical etching), or by mechanisms combining both physical and chemical effects (reactive ion etching or RIE). These methods have various etching selectivity and achieve different etching profiles and usually the etching is more anisotropic and vertical when the etching is more physical, while it is more selective and isotropic when it is more chemical. Most of these methods are discussed in standard microelectronics process books, but they take a different twist when they are applied to MEMS fabrication as in general MEMS necessitates deep (> 5µm) etching.
The Figure illustrates an important parameter commonly used to describe dry etching: the aspect ratio. Actually we can define an aspect ratio for features \((h/w_r)\) and for holes \((h/w_h)\) with most technologies giving better results with features than with holes - but generally with only a small difference. Typical values for this parameter would range between 1 (isotropic etch) and 50, for very anisotropic etching like the DRIE process.

To improve the aspect ratio of the etching, several techniques based on RIE have been developed, usually trying to increase the anisotropy by protecting the sidewalls during etching. For example, the SCREAM process developed in Cornell University alternate steps of etching and growth of oxide that remains longer on the sidewall, while for the RIE cryogenic process, very low temperature in a SF\(_6\)/O\(_2\) plasma is used to obtain continuous deposition of a fluoro-polymer on the sidewall during the etch. Finally the innovative Bosch process uses alternating cycle of etching and protection in the plasma chamber to achieve the same function. This important process is the corner stone of DRIE micromaching and will be described in more details in a following section.

### 3.3.3 Wafer bonding

A review of MEMS fabrication technique cannot be complete without mentioning wafer bonding. Wafer bonding is an assembly technique where two or more precisely aligned wafers are bonded together. This method is often used simultaneously for device fabrication and also for its packaging - belong both to front-end and back-end process, another peculiarity of MEMS, but at this stage it is not surprising anymore!

Wafer bonding has the potential to simplify fabrication method because structures can be patterned on both wafers and after bonding they will be part of the same device, without the need for complex multi-layer fabrication process. The main issues that need to be considered to evaluate a wafer-bonding technique are: the bonding temperature (high temperature may damage the materials or structure on the processed wafer), the difference in coefficient of thermal expansion between the bonded materials (affecting stress sensitive systems during use) and the permeability to gas and humidity of bond and bonded wafer (affecting long term reliability).

The bonding techniques are usually split between intermediate layer bonding technique, where an intermediate layer is used to form the bond between the two wafers, and direct bonding methods where there is no such layer.

The simplest form of intermediate layer bonding is of course epoxy bond-
ing. Although epoxy cures at low temperature (< 100°) and is cheap, such bond poses performance problems as epoxy have large thermal expansion coefficient and are permeable to gas and moisture. Intermediate-layer eutectic bonding is based on forming an eutectic alloy that will diffuse into the target wafer and form intermetallic to create a strong bond. For silicon-to-silicon bonding the intermediate layer is often gold which form a eutectic alloy with silicon at 363°C. Actually, if the two sides have been coated with gold, an even lower temperature can be used (≈ 250°) by applying pressure (50 MPa) at the same time. This process is called thermocompression bonding and was originally developed for gold wire-bonding in the late 1950s.

Silicon-to-silicon fusion bonding allows bonding two silicon wafers directly effectively achieving seamless bond possessing an exceptional strength and hermeticity. However the technique requires excellent flatness and high temperature, two hurdles that limit its use.

The most commonly used MEMS bonding methods is probably anodic bonding, or field assisted bonding, which is mainly used to bond silicon wafers with glass wafers. The technique work by applying a high voltage to the stacked wafers that induce migration of ion from glass to silicon, allowing a strong field to appear at the interface. The field will pull both surface in intimate contact, helping to form the bond. This technique is commonly used to fabricate sensors allowing for example to obtain cavities with controlled

Figure 3.6: Silicon pressure sensor SP15 bonded with glass cover (Courtesy SensoNor AS - An Infineon Technologies Company).
pressure for pressure sensor as shown in Figure 3.6. At the same time, the glass wafer provides wafer level packaging, protecting sensitive parts before back-end process.

Before closing this section, it should be noted, that wafer bonding is also used to fabricate MEMS substrates such as SOI and SOG (silicon on glass) wafers. For SOI fabrication a thinned Si wafers is bonded to an oxidized Si substrate before it can be polished to the desired device thickness, resulting in the typical silicon-oxide-silicon stack.

![Figure 3.7: Basic process sequence of surface micromachining.](image)

### 3.4 Surface micromachining and thin-films

Unlike bulk micromachining in which microstructures are formed by etching into the bulk substrate, surface micromachining builds up structures by adding materials, layer by layer, on the surface of the substrate. The thin-film layers are typically 1 ∼ 5 µm thick, some acting as structural layer and others as sacrificial layer. Dry etching is usually used to define the shape of the structure layers, and a final wet etching step releases them from the substrate by removing the supporting sacrificial layer.

A typical surface micromachining process sequence to build a micro bridge is shown in Figure 3.7. Phosphosilicate glass (PSG) is first deposited by
LPCVD to form the sacrificial layer. After the PSG layer has been patterned, a structural layer of low-stress polysilicon is added. Then the polysilicon thin-film is patterned with another mask in CF$_4$ + O$_2$ plasma. Finally, the PSG sacrificial layer is etched away by an HF solution and the polysilicon bridge is released.

![Figure 3.8: A micro optical stage built by surface micromachining.](image)

As a large variety of materials such as polysilicon, oxide, nitride, PSG, metals, diamond, SiC and GaAs can be deposited as thin film and many layers can be stacked, surface micromachining can build very complicated micro structures. For example Sandia National Laboratories is proposing a process with four polysilicon structural layers and four oxide sacrificial layers, which has been used for fabricating complex locking mechanism for defense application. Figure 3.8 demonstrates surface micromachined micro-mirrors fabricated using two polysilicon structural layers and an additional final gold layer to increase reflectivity. They have been assembled in 3D like a pop-up structure, using a micromanipulator on a probe-station.

### 3.4.1 Thin-film fabrication

The choice of the thin-film and its fabrication method is dictated by many different considerations: the temperature budget (limited by the maximum temperature that the substrate can withstand and the allowable thermal stress), the magnitude of the residual stress in the thin-film (too much stress cause layer cracking), the conformality of the thin-film (how the thin-film follows the profile of the substrate as shown in Fig. 3.9), the roughness of the
thin-film, the existence of pinholes, the uniformity of the thin-film, the rate of fabrication (to obtain cost-effective thick thin-film)...
Table 3.5: Combination of materials and etchant for surface micromachining.

<table>
<thead>
<tr>
<th>Structural material</th>
<th>Sacrificial material</th>
<th>Etchant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polysilicon</td>
<td>Oxide (PSG, LTO, etc)</td>
<td>Buffered HF</td>
</tr>
<tr>
<td>Si₃N₄</td>
<td>Poly-Si</td>
<td>KOH</td>
</tr>
<tr>
<td>SiO₂</td>
<td>Poly-Si</td>
<td>EDP/TMAH</td>
</tr>
<tr>
<td>Aluminum</td>
<td>Photoresist</td>
<td>Acetone/O₂ plasma</td>
</tr>
<tr>
<td>Polyimide</td>
<td>Cu</td>
<td>Ferric chloride</td>
</tr>
<tr>
<td>Ti</td>
<td>Au</td>
<td>Ammonium iodide</td>
</tr>
<tr>
<td>SiO₂, Si₃N₄, metal</td>
<td>Poly-Si</td>
<td>XeF₂</td>
</tr>
</tbody>
</table>

The quality of the thin-film is often linked with its microstructure, and we differentiate them according to an increasing degree of order as amorphous, polycrystalline and single crystal films.

A single crystal film present the highest order, as the atoms are periodically arranged in a precise manner following one of the lattice possibility allowed by geometry, that is, one of the 14 Bravais' lattice. In this case the material properties of the crystal are highly reproducible but will generally depend on the orientation with respect to the crystal axes, the material is then anisotropic. Single crystal thin-films are relatively hard to obtain, and only a few process like epitaxy will yield such films.

In the case of polycrystalline films, the material does not crystallise in a continuous film, but in small clusters of crystal, each cluster having a different orientation than its neighbour. In general the atom cluster size range from about 10 nm to a few µm, and they are called grains. The grains may not be completely randomly oriented and some direction may be favored depending on the growth process, resulting in highly varying material properties for different process condition. If the distribution of grain orientation is known, a good approximation of the properties of the material can be obtained by using the weighted average of the single crystal properties along different directions. Finally, in amorphous films, the material grows in a disordered manner, with clusters of crystal being of a few atoms only. In this case, the material prop-
properties are not the same as those present in single crystal or in polycrystalline films, and usually present inferior characteristics: lower strength, lower conductivity... Interestingly, the material properties of such films are normally more stable with the process parameters and are independent of the direction in the film as the material is isotropic.

Oxidation

Oxidation belongs to the modifying processes, sharing with them a generally excellent conformity. Oxidation is a reactive growth technique, used mostly on silicon where silicon dioxide is obtained with a chemical reaction with a gaseous flow of dry or wet dioxygen. Using dry dioxygen results in a slower growth rate than when water vapour are added, but it also results in higher quality films. The rate of growth is given by the well-known Deal and Groove’s model as

\[ d_o = \frac{A}{2} \sqrt{1 + \frac{t + \tau}{A^2/4B}} - \frac{A}{2} \]

where \( B \) is called the parabolic rate constant and \( B/A \) the linear rate constant that are obtained for the long and the short growth time limit respectively. Typical value for these constant at 1000°C are \( A = 0.165\mu m, B = 0.0117\mu m^2 \) and \( \tau = 0.37h \) in dry O\(_2\) and \( A = 0.226\mu m, B = 0.287\mu m^2 \) and \( \tau = 0 \) in wet O\(_2\). It should be noted that the model breaks down for thin dioxide (\(<300\AA\)) in dry atmosphere because of an excessive initial growth rate that is modeled through the use of \( \tau \). The passage from a linear growth rate to a parabolic rate is dictated by the need for the oxygen atoms to diffuse through the growing layer of silicon dioxide - the thicker the layer the slower its growth rate. The diffusion rate is increased for wet oxidation, resulting in faster growth.

Actually, it is possible to control locally the oxidation by blocking in selected places the diffusion of oxygen to the surface. In this process, Si\(_3\)N\(_4\) is deposited and patterned on the silicon surface to act as a barrier against oxygen diffusion. During the oxidation process in the furnace, oxide grows only in the bare regions while Si\(_3\)N\(_4\) prevents oxidation in the covered regions, effectively resulting in a patterned oxide layer. It could be noted that at the edge of the region, the nitride film is lifted by the oxide lateral growth resulting in a “bird’s neck” profile of the oxide, sensibly different to what happens during etching of oxide.

For MEMS applications, an interesting feature of oxidation is that it results in a net volume change as the density is lower for oxide than for silicon. Actually, the volume increases by about 53% of the grown oxide.
That is, for an infinite plane, the growth of a thickness $d_o$ of dioxide results in the consumption of a thickness $d_{Si} = 0.46d_o$ of silicon, but a net expansion of $d_o - d_{Si} = 0.53d_o$ during growth. If this phenomenon may produce unwanted stress, it is also used to close holes in silicon or poly-silicon layers.

The relatively large range of variation of the oxidation growth rate, allows to obtain thick layer (up to 2 $\mu$m in 10h) or very thin layer (a few nm) with a good control. The thicker films can be used as mask for wet etching or as sacrificial layer, and the thinner ones serve to produce nano-structures with high accuracy. This versatility and the high quality of the film produced give oxidation an important role in MEMS manufacturing.

**Doping by diffusion and ion implantation**

Doping is a process where impurities are introduced into a material to modify it, and as such belongs to the modifying processes. The impurities can be directly introduced during the fabrication of the substrate (for example, As can be introduced into the melt during the growth of the silicon ingot to obtain n-doped silicon wafer), but we are here interested by techniques that could be used selectively for doping locally and in-situ a thin layer of material. The two main techniques that are used are diffusion and ion implantation.

Diffusion is performed by placing the substrate in a high temperature furnace in presence of the doping species. In the main process currently used the doping species are present in gaseous form in the furnace or have been deposited as a thin film directly on to of the substrate. The high temperature will agitate atoms strongly and allow the impurities atoms to move slowly inside the substrate, until their concentration is uniform.

Actually, the diffusion is governed by the Fick’s laws which are derived from the statistical study of the random motion of particle due to thermal energy. The first law relates the flux of the diffusing species ($j$) with the gradient of concentration $C$. The proportionality constant $D$ is the diffusion constant, a material constant depending on the substrate, the diffusing atoms and the temperature.

$$j = -D \text{grad}C$$

which in system with one dimension gives $j = -D \partial C/\partial x$. This equation translates the fact that the average flow of impurities will last until the concentration is equal everywhere and the gradient is null. The second law allows to relate the evolution of the concentration with time.

$$\frac{\partial C}{\partial t} = D \Delta C$$

which in one dimension is given by $\partial C/\partial t = D \partial^2 C/\partial x^2$. These partial
differential equations can be solved for different set of initial and boundary condition, depending on the diffusion configuration.

Of particular interest is the case of infinite source, where the concentration at the surface remains constant during the complete diffusion, as happen during diffusion from a gaseous source in a furnace. In this case the concentration is given as:

\[ C(x, t) = C_S \operatorname{erfc} \left( \frac{x}{2\sqrt{Dt}} \right) \]

where \( \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = 1 - 2/\sqrt{\pi} \int_0^x e^{-\zeta^2} d\zeta \) is the complementary error function. From this equation we can obtain the diffusion depth \( d \) as:

\[ d = 2\sqrt{Dt} \operatorname{erfc}^{-1} \left( \frac{C_d}{C_S} \right) \]

where \( C_d \) is the concentration in \( d \). We verify here that the depth varies as the square root of time, that is, to go twice as deep, the diffusion needs to be 4 times longer. The surface concentration \( C_S \) remains constant during the diffusion and is generally given by the solubility limit of the impurities in the substrate (e.g., \( 2 \cdot 10^{20} \text{at/cm}^3 \) for Boron in Silicon at 1100°C). Actually, the Fick’s equations describe only the evolution of the concentration inside the material and not what happen at the interface between the substrate and the upper medium. The solubility limit gives the maximum concentration that is reached before the impurity forms cluster and small crystals which will require a lot more energy and generally does not happen. In this way it also gives the maximum impurities concentration that can be obtained by diffusion in a material.

Another common case is the case of finite source, where, after some time, the source is completely consumed by the diffusion, as happen when a thin-film on the substrate surface acts as the diffusion source. In this case, the relevant quantity is the total amount of dopant present \( Q \) that will finally fully diffuse in the substrate. In this case the solution to the Fick’s second law is:

\[ C(x, t) = C_S \exp \left( -\frac{x^2}{4Dt} \right) \]

with in this case the surface concentration \( C_S = Q / \sqrt{\pi Dt} \), decreasing steadily with the diffusion time. We obtain in this case a diffusion depth given by

\[ d = 2\sqrt{Dt} \sqrt{\ln(C_d/C_S)} \]

varying again as a function of the square root of the time.

Ion implantation is comparatively a more versatile technique, but it requires a much more complex set-up (Figure 3.10). In this case the impurities
are introduced into the substrate by bombarding the substrate with ions of the impurity traveling at high velocity. The ions will penetrate and be stopped after a short distance under the surface (at most a few μm) by interaction with the electrons and atoms of the substrate material. The ion implanter is thus composed of a collimated source of ion, a section using high voltage to accelerate the ions, a mass spectrometer to sort the ions and only select the desired species and finally an electrostatic scanning system allowing to direct the ion beam toward any place on the wafer surface.

![Figure 3.10: Schematic of an ion implanter.](image)

The parameters governing ion implantation are simply the ion energy, expressed in eV and depending on the acceleration voltage, and the dose, that is the number of implanted atoms per unit area. The energy will determine the depth of implantation, with deeper implantation resulting in broader distribution of atoms. The dose is simply the ion current density \(j\) multiplied by the implantation duration \(t\) and divided by the charge of one ion. It can also be expressed by using the total current \(I\),

\[
D = \frac{jt}{q} = \frac{It}{Aq}.
\]

The implantation being a rather violent process, the collision of the impurity ions with the stationary atoms of the substrate cause them to recoil and generally results in amorphization of the doped portion of the substrate. In general recrystallization is needed and thus the implantation process needs to be followed by an annealing process at high temperature (800°C to 1200°C) under an inert atmosphere.

If ion implantation allows to tailor, by varying dose and energy, doping profile much more precisely than diffusion, it is not exempt of drawbacks. The
Implantation is a rather directional process and it is affected by shadowing behind tall structures and reflection on the side wall, making uniformity more problematic with high aspect ratio structures. Additionally, inside crystals there are often directions with less chance of nucleus collision where ions will be channeled much deeper. For example in Si, such phenomenon appears along the <111> orientation, and when ion implantation replaced diffusion for microelectronics at the turn of the 1980's, the preferred Si wafer orientation changed from <111> to <100>. Finally, ion implantation is unable to provide deep doping profile (i.e., a few µm) and, contrary to microelectronics, diffusion remains widely used in MEMS fabrication.

![Concentration profile](image)

Figure 3.11: Concentration profile for (left) implanted ions of different energy (right) diffused atoms with different diffusion time and with (dots) infinite and (solid) finite source.

In general, it is possible to control locally the doping by placing a protective layer over the zone that should not be doped before the doping process is performed. Actually the protecting layer will be doped at the same time as the exposed substrate, but will be removed in a later step leaving only the exposed substrate with doping. If the doping is obtained by diffusion, at the edge of the pattern lateral (isotropic) diffusion will occur enlarging the original pattern, but implanted layer will have more precisely defined edges.

**Spin-coating**

Spin-coating is a simple and fast technique that allows to deposit thin-films in liquid form. In general it is used to deposit polymer in a solvent, and
particularly photoresist, but it can also be used for glass (spin-on glass) or other material (PZT...) with the sol-gel technique.

![Spin-coating principle](image)

Figure 3.12: Spin-coating principle.

As shown in Figure 3.12, the substrate is mounted on a chuck that is able to spin at high speed before the liquid is dispensed in a puddle in the substrate center. In a first spinning step at relatively low speed (a few 100 rpm), the liquid spreads over the entire substrate. Then a fast acceleration brings it to a high speed spinning for about 30 s where the layer reach its desired thickness. In general the high speed spin has to be kept between 1000 rpm and 5000 rpm to give optimum results, with increasing speed giving thinner films. The viscosity of the liquid is the main factor that determine the final layer thickness at a certain spinning speed, requiring a good temperature control (viscosity changes rapidly with temperature) to obtain reproducible results.

Photoresists exist in a wide range of viscosity, from a few cSt to 50000 cSt or more, allowing in a single spin to obtain thickness between about 100 nm and up to a few 100 µm. A standard layer would be around 1 µm, but thick resist are very attractive in MEMS fabrication. However, spin-coating very viscous polymer is difficult and will be helped by using a spinner with a co-rotating cover. In that case the top cover above the substrate is rotating at the same speed, allowing to avoid the effect of air flow that creates uneven layer with particularly thicker edge called edge-bead. The edge-bead can be removed in a post-process conducted just after spin-coating, but their minimization is still important.

The main limitation with spin-coating is that the material need to be in liquid form, restricting the range of available material to a relatively small list. In general, as it is the case for photoresist, the material is dissolved in a solvent which is then evaporated after spin-coating leaving a hard film on the substrate. Alternatively, monomer can be spin-coated and then polymerized by heating or using UV exposure to form a film. Finally, in sol-gel method
like the spin-on-glass (SOG), suspension can be spin-coated and will form a network upon thermal treatment.

The method works best over a flat substrate as its conformality is not good. Actually, when the liquid is spun over a substrate with topography, the liquid surface tension will smooth out the sharp edge, rounding convex corner and filling concave ones. This phenomena can be put to good use to smooth out the topography of a wafer and has been use for local planarization. However in general this phenomena is a problem and a few methods have been developed to solve it in the case of photoresist where a good control of thickness is very important. One of such method is the spray coating method, where the photoresist is sprayed over the surface avoiding presence of thinner layer at corners. Photoresist has also been deposited by using electroplating. In that case a conductive substrate (i.e. metal coated) is placed in a special electrolytic bath and after application of a current, a layer of photoresist is grown uniformly on all the surface of the substrate, providing excellent conformality.

Physical Vapor Deposition (PVD) techniques

There are two main techniques for physical vapor deposition: evaporation and sputtering. These techniques are low temperature and most often used to deposit conductive materials like metals (Au, Pt, Ti, Cr, Al, W, Ni...). However they can also be used for oxide or semi-conductive materials, resulting almost invariably in amorphous thin-films.

The principle of evaporation is very simple. The wafer is first placed with the source material in a vacuum chamber. The material is heated above its evaporation temperature sending vaporized atoms across the chamber. The atoms then condense on the colder surfaces: the wafer and the chamber walls. The material source is kept in a small crucible and heated using resistive heating (Joule’s effect) or using an electron beam. In the later case, a high velocity beam of electron is directed towards the source. When the electron collide with the surface atom they transfer their kinetic energy, bringing the material to high temperature. The choice between the two techniques is not only governed by the evaporation temperature, which would make the e-beam technique superior as it can reach much higher temperature than resistive heating. Actually, some materials are easier to evaporate with one method or the other, and for example, some rare earth oxide can not be evaporated nicely with e-beam whereas they are successfully evaporated with resistive heating in a tungsten crucible.

For both heating techniques, the substrate has to be kept at a relatively...
large distance from the hot source to prevent the substrate from heating uncontrollably. However as the source material in the crucible has a relatively small size, the large distance result in a line of sight deposition with poor conformality. As we see in the figure, the horizontal surface will be uniformly coated, but the vertical surface will receive less material as their projected surface is smaller. The vertical side can even be left in the ‘shadow’ and receive no material at all. This last problem can be generally avoided (except in narrow trenches) by rotating the substrate during the deposition. Of course the shadowing effect is not always a problem and actually lift-off process can benefit from it. Actually, if nothing deposits on the sacrificial material side this will facilitate its dissolution by providing a better access to the etching liquid.

The sputtering process is one of the many process based on the utilization of a plasma, an equal mixture of positive ions and electrons with some neutral atoms. In a plasma, new pairs of ion and electron are continuously formed by ionization and destroyed by recombination. Plasma can be created by a glow discharge when a low pressure gas is submitted to a large electrical field. In the glow discharge when the plasma is ignited, electrons close to the cathode are accelerated by the electrical field across the Crooke’s dark space until they reach the velocity necessary to ionize gas atoms. When the electrons hit a neutral gas atom they knock electrons out of their outer shell, ionizing them. The space region where this happen is called the negative glow and it is where most of the plasma is present. Actually the glow, whose colour (or wavelength) is characteristic of the plasma, comes from the photon emitted when a previously ionized atom and an electron recombine. The positively ionized atoms are in turn accelerated by the electrical field in the Crooke’s dark space, but because of their charge they move toward the cathode. When they hit it, the collision produces secondary electrons, which can be accelerated again to ionize more neutral atoms, creating a self-sustaining phenomena.

In the region after the negative glow, the electron have lost most of their kinetic energy during the collision with the atoms and they accelerate through the Faraday dark zone before having enough energy to ionize atoms again and creating more glow zone in what is called the positive column. In general this part of the glow discharge region is not useful and the anode is kept close to the Faraday zone edge.

The pressure in the plasma has to be kept low so that the accelerated electrons do not encounter neutral atoms before they have acquired enough kinetic energy to ionize them. In the other hand if the pressure is too low the probability of the electrons to ionize an atom is too low and the plasma

\[1\] This is the same phenomena that makes the temperature depends on the latitude on earth: at higher latitude the curve of the globe let the rays of the sun shine obliquely on the earth surface and thus make them illuminate a larger surface, bringing less heat there.
may not sustain itself. There is thus understandably a preferred range of operation pressure where the plasma is the brightest.

The basic operation of the DC sputter is basically the same as the glow discharge, but operated at a higher voltage. In that case, when the positive gas ions are accelerated in the Crooke’s dark space and hit the target surface they have enough momentum to eject one or more target atom in addition to the secondary electrons. The neutral atom will then fly through the chamber and land onto the wafer.

The most commonly used gas for sputtering is argon, an inert gas, avoiding any chance for the gas to chemically react with the substrate of the target. The target is located at the cathode which is biased at a few kV where the sputtering efficiency is highest for most materials. The substrate bias is also
often set to a negative value ($\approx -100$ V, that is, of course, much lower than the cathode voltage), to attract ions that will help compact the deposited film and remove the loose atoms. Alternatively a positive bias could be used to attract the electrons and repel the ions, resulting in different thin-films properties. Clearly, the substrate bias should not be made too negative otherwise sputter of the substrate will occur. In fact, this possibility is often used for a short time by biasing the substrate to 1 kV before the deposition is started in order to clean the substrate surface. Additionally, the wafer is normally heated at a temperature set halfway to the melting temperature ($T \approx 0.5T_m$) to increase the atom mobility at the surface. This result in film having less porosity and lower stress. However, the main factor affecting stress is the gas pressure, and lower stress will be obtained with low pressure.

The magnet is used to increase the ionization yield of gas atoms. Actually the magnetic field traps the moving electrons (Lenz’s law) in an helical path, increasing the length of their trajectory and thus increasing the chance they hit a gas atom. The yield increases dramatically ($> 10\times$) resulting in a similar increase of the deposition rate. Most modern sputter are using this principle and are called magnetron sputter.

One issue with the DC sputter is that it can not be used efficiently for depositing insulating material. Actually, the charge building-up on the insulating target surface would finally repel the incoming ions and require a dramatic increase of the voltage to sputter the target material. However, instead of a DC potential, the sputter can also be operated with a pulsed field in the radio frequency (RF sputter) range (in general at 13.56 MHz). In this case the potential of the target is maintained negative for most of the period and then briefly turned positive. The short duration of the positive field does not modify much the momentum of the heavy ions that behave mostly as seen before in the DC field. In the other hand the much lighter electrons are affected by the positive field and they move backward toward the target. This neutralizes the build-up of positive charge which would happen in DC sputter with non conductive target and allows deposition of insulating or semi-conducting materials.

The atoms coming from the target follows mostly a line of sight path, but the conformality is still better than with evaporation. Actually, as can be seen in Figure 3.14, owing to the large dimension of the target and its proximity with the wafer, target atoms arrive on the wafer within a relatively broad solid angle spread, decreasing the shadow effect usually observed with evaporation. Additionally, the atoms from the sputter have a higher velocity that atoms obtained by evaporation, resulting in layer with better adhesion to the substrate. Finally the deposition speed is higher with sputter, making it the tool of choice for metal deposition in the industry.
Chemical Vapor Deposition (CVD) techniques

Chemical vapor deposition techniques are split according to the operating pressure, from atmospheric pressure CVD (APCVD), low-pressure CVD (LPCVD) to finally ultra-high-vacuum CVD (UHCVD). However they all work on the same principle: the decomposition of a gas on the heated surface of the wafer in the absence of any reagent, a phenomena called pyrolysis. CVD is performed in a simple furnace with a gas inlet and connected to a vacuum pump to maintain a controlled pressure as shown in Figure 3.15.

![Figure 3.15: LPCVD furnace for thin-film deposition.](image)

Depending on the gas, or gas mixture, used it is possible to deposit a wide variety of thin-films. The most commonly deposited films are polysilicon using the decomposition of silane (SiH₄) at 620°C, silicon nitride using a mixture of di-chlorosilane (SiH₂Cl₂) and ammonia (NH₃) at 800°C and low temperature oxide (LTO) using silane and oxygen (O₂) at 450°C.

In general the resulting films are of good quality showing a very good conformality and having no pinholes and good deposition rate (1µm/h). The stress in the film can be controlled by varying the temperature and the gas flow rate or composition. For example if stoichiometric silicon nitride will usually present a very large tensile stress (>1 GPa) making it crack easily for thickness above 200nm, increasing the amount of dichlorosilane in the gas mixture will result in silicon-rich silicon nitride films with a much lower stress that can be grown up to 1µm and above.

Mixing of gas can result in other interesting variation in thin-films. For example it is possible to deposit oxyxnitride thin-films with an index of refraction anywhere between oxide (n=1.46) and nitride (n=2.1) by simply varying the ratio of oxygen and ammonia in a mixture with dichlorosilane.

The main concern with the LPCVD is the relatively elevated temperature needed for the pyrolysis of the gas. Actually, even the LTO deposition is actually rarely used as it shows a low conformality and usually oxide is deposited using pyrolysis of TEOS at 650°C. To circumvent this temperature
problem, it is possible to use Plasma Enhanced CVD (PECVD). In this case a plasma is used to add kinetic energy to the thermal energy to facilitate the chemical reactions allowing the furnace to run at much lower temperature (<450°C). The low process temperature has made the PECVD reactor a popular tool for research and industry, used for many materials as shown in Table 3.6.

The PECVD furnace is completely different and more complex than its LPCVD counterpart, as in addition to temperature and gas control it needs the circuitry to excite and maintain a plasma inside the chamber above the wafer. As such, the deposition can only happen for one wafer at a time, and because of the horizontal position of the wafer, is more prone to contamination from falling particles. Additionally, the quality of the thin-films is usually lower - but adjusting the plasma parameter allows a better control on the mechanical properties of the film.

<table>
<thead>
<tr>
<th>Material</th>
<th>Technique</th>
<th>Gas</th>
<th>T [°C]</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxide (SiO₂)</td>
<td>LPCVD LTO</td>
<td>SiH₄ + O₂</td>
<td>425</td>
<td>low density</td>
</tr>
<tr>
<td></td>
<td>LPCVD TEOS</td>
<td>Si(OC₂H₅)₄</td>
<td>700</td>
<td>good film</td>
</tr>
<tr>
<td></td>
<td>PECVD</td>
<td>SiH₄ + N₂O</td>
<td>300</td>
<td>contain H</td>
</tr>
<tr>
<td>Nitride (SiNx, Si₃N₄...)</td>
<td>LPCVD Si₃N₄</td>
<td>SiH₂Cl₂ + NH₃</td>
<td>800</td>
<td>high stress</td>
</tr>
<tr>
<td></td>
<td>LPCVD SiₙNₓ</td>
<td>SiH₂Cl₂ + NH₃</td>
<td>800</td>
<td>low stress</td>
</tr>
<tr>
<td></td>
<td>PECVD SiₙNₓ</td>
<td>SiH₄ + NH₃</td>
<td>300</td>
<td>contain H</td>
</tr>
<tr>
<td>Silicon (PolySi, a-Si...)</td>
<td>LPCVD PolySi</td>
<td>SiH₄</td>
<td>620</td>
<td>small grain</td>
</tr>
<tr>
<td></td>
<td>LPCVD a-Si</td>
<td>SiH₄</td>
<td>570</td>
<td>amorphous</td>
</tr>
<tr>
<td></td>
<td>PECVD a-Si</td>
<td>SiH₄</td>
<td>280</td>
<td>contain H</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>LPCVD W</td>
<td>WF₆ + SiH₄</td>
<td>440</td>
<td>good conf.</td>
</tr>
</tbody>
</table>

Table 3.6: CVD of common thin-films.

Actually LPCVD furnaces, even with vertically standing wafers, are also subjected to contamination from particles falling from the furnace walls. As we mentioned earlier the walls are hot and thus experience the same deposition as the wafer, accumulating across multiple run and finally falling on the wafer as particles. There are a few solution to this problem, and for example the LPCVD furnace may be oriented vertically instead of horizontally. Another possibility is to keep the wall cold so that no deposition occurs on
them. This is not possible with LPCVD furnaces, but it can be achieved with Rapid Thermal Processing (RTP) furnaces. In a RTP furnace wafers are processed one by one horizontally, similar to what is done in a PECVD furnace. However here the high temperature is obtained using strong lamp whose NIR radiation is absorbed in the wafer. The walls can thus be kept at low temperature by water circulation preventing any deposition there.

It should be noted though, that for MEMS the problem of particles contamination is not as bleak as it could seem because the dimension of the feature are usually much larger than for ICs. Actually fabrication process generally generates much more small particles than big ones: in a typical cleanroom air, when the size of particles goes from 1µm to 0.01µm, the particle density may increase by 2 order of magnitude! As such, if we accept the rule of thumb that particle should be smaller than about one third of the critical dimension in the layout, we understand easily that MEMS with feature size rarely below a few µm are much less impacted by the particle contamination than ICs with 50nm design rules. We can also remark that the expected gain in yield achieved by shrinking the chip dimension (smaller area has less chance to see one particle) can be easily offset by the dramatic increase of smaller particle number noted above: to keep the same yield, the environment should actually be cleaner, and the number of smaller particle should decrease proportionally with the scaling factor.

**Epitaxy**

Epitaxy is a CVD technique, as it generally relies on furnace very similar to what is used for LPCVD, but actually it presents features that makes it unique. The main difference between epitaxy and other CVD techniques is that in the case of epitaxy the structure of the thin-film depends on the substrate, and particularly, epitaxial growth allows to obtain single crystal layers. Actually if in the case of CVD the deposition is relatively random and independent of the substrate, generally resulting in amorphous or polycrystalline films, with epitaxy the thin-film will grow in an ordered manner determined by the lattice of the substrate.

If the material of the epitaxial layer is the same as the substrate, the process is called homoepitaxy and heteroepitaxy otherwise. Moreover, depending on the match between the lattice period of the substrate and the film, we can distinguish three types of epitaxial growth:

- **commensurate growth** when the substrate and the layer have the same crystal structure and lattice constant,
- **incommensurate growth** when they don’t have the same lattice constant
resulting in point defects at the interface,

**pseudomorphic growth** when they don’t have the same lattice constant but the epitaxial layer strains to match the lattice of the substrate.

The growth of high quality crystals, like silicon, is generally obtained by the Czochralsky method, which consists in pulling slowly from a melt very large single crystal starting from a small seed crystal. Actually, this method could be described as an extreme case of homoepitaxy with commensurate growth.

The epitaxy process needs a furnace similar to a LPCVD furnace, but in practice the process is relatively more complicated to control. On silicon the main process is based on the reduction in a H$_2$ atmosphere of SiCl$_4$ at 1200°C with HCl as a by-product. However the high temperature makes it hardly useful, except as a first process step, and lower temperature process using dichlorosilane above 950°C have been developed, but are harder to control resulting often in polycrystalline layers.

The main interest of the technique is in the high quality of the grown layer, which results in good electronics properties, important for optoelectronics (solar cell, laser diode...) and some specific electronics circuits, and good mechanical properties (low stress) more interesting for the MEMS application. The relative difficulty of the technique makes it rarely used in MEMS fabrication, with the notable exception of the process used by Bosch for their multi-user foundry process. In the MPW process the structural layer is a 10.5µm polycrystalline layer grown by epitaxy (called epipoly). In this case the interest is the growth speed (that could exceed 0.3µm/min) that can be obtained without sacrificing the low stress present in the layer.

### 3.4.2 Design limitation

The flexibility of surface micromachining is not free of unique problems that need addressing to obtain working devices.

During layer deposition, a strict control of the stress in the structural layer has to be exerted. Compressive stress in a constrained member will cause it to buckle, while a gradient of stress across a cantilevered structure causes it to warp, resulting in both case in probable device failure. The possibility to stack several layers brings freedom but also adds complexity. Actually there is large chance that the topography created by the pattern on underlying layer will create havoc with the upper layer, as illustrated in Figure [3.16](#).

A common problem is the formation of strings of structural material, called ‘stringers’, during the patterning of the upper layer. Actually the high
anisotropy of the etching by RIE leaves some material where the layer is thicker because of the conformal deposition of the structural material. To avoid the problem during fabrication, the RIE etching time needs to be substantially increased to fully etch the layer where it is thicker. For example the MUMPS surface micromachining process proposed by the foundry MemsCap is using an overetching of 100%, that is, the etching lasts twice the time needed to clear the material in the flat zone. Another common issue is the likelihood of structure interference between the stacked layers. In Figure 3.16 we see that the topography creates an unintended protrusion below the top structural layer that will forbid it to move freely sideways - probably dooming the whole device. This problem can be tackled during layout, particularly when the layout editor has a cross-section view, like L-Edit from Tanner Research. However even a clever layout won’t be able to suppress this problem completely and it will need to be addressed during fabrication. Actually it is possible to polish using Chemical-Mechanical Polishing (CMP) the intermediate sacrificial layer to make it completely flat, will avoid all interference problems. For example, Sandia National Laboratory uses oxide CMP of the second sacrificial layer for their four layers SUMMiT V process. However, sometimes the interference may be a desired effect and for example the so called 'scissors' hinge [23] design shown in Figure 3.17 benefits greatly from it. The scissors hinge is designed to provide a hinge functionality with micromachining process and as we see here the protrusions below the upper layer help to hold the hinge axis tightly. If we had to rely on lithography only, the gap between the axis and the fixed part in the first structural layer would be at best 2 $\mu$m, as limited by the design rules, and the axis will have too much play. However the protrusions below the staple reduce the gap to
0.75 µm, the thickness of the second sacrificial layer, and the quality of the hinge is greatly increased.

Figure 3.17: Tight clearance obtained by layer interference in a hinge structure.

The final step in surface micromachining process is the release - and this critical step has also a fair amount of issues that need to be considered.

### 3.4.3 Microstructure release

The release step is the source of much technologist woes. Release is usually a wet process that is used to dissolve the sacrificial material under the structure to be freed. However the removal rate is usually relatively slow because the sacrificial layer is only a few µm thick and the reaction becomes quickly diffusion limited. Then the depth of sacrificial layer dissolved under the structure will increase slowly with the etching time as

\[ d_{\text{release}} \propto \sqrt{t_{\text{etch}}} \]

Simply said, releasing a structure twice as wide will take 4 times more time. However if the etching lasts too long the chemical may start attacking the device structural material too. A first measure to avoid problems is to use
compatible material and chemical, where the sacrificial layer is etched quickly but other material not at all. A typical example is given by the DLP (Digital Light Processing) from Texas Instruments, where the structural layer is aluminum and the sacrificial layer is a polymer. The polymer is removed with oxygen plasma, and prolonged release time will only slightly affect the metal. This ideal case is often difficult to reach and for example metals have often a finite etch rate in HF, which is used to remove PSG sacrificial layer. Thus to decrease the release time we have to facilitate etching of the sacrificial layer by providing access hole for the chemical through the structural layer. In the case of Figure 3.8 for example, the mirror metal starts to peel off after about 10 minutes in HF. However in about 5 minutes HF could only reach 40 µm under a plain plate, and the designer introduced ‘release holes’. These holes in the structural layer are spaced by roughly 30 µm in the middle of the mirror plate (the white dots in the figure) allowing for the HF to etch all the oxide beneath in less than 5 minutes.

![Figure 3.18: Stiction phenomenon during release.](image)

The problems with wet release continue when you need to dry your sample. The meniscus created by the receding liquid/air interface tends to pull the structure against the substrate. This intimate contact give rise to other surface forces like Van der Walls force, which will irremediably pin your structure to the substrate when the drying is complete, effectively destroying your device. This phenomenon is referred as stiction (Figure 3.18). Strategies that have been used to overcome this problem have tackled it at design and fabrication level. In surface micromachining the idea has been to reduce the contact surface by introducing dimples under the structure. From the fabrication side, super-critical drying, where the liquid changes to gas without
creating a receding meniscus, has also been applied successfully. Coating the structure with non-sticking layer (fluorocarbon, hydrophobic SAM...) has also proved successful and this method, albeit more complex, has the added advantage to provide long lasting protection again sticking that could arise during use.

Finally, a completely different approach is to avoid wet release altogether and instead to perform a dry release with a gas or a vapour, suppressing entirely the stiction concerns. For example the Multi-Project-Wafer (MPW) process run by Bosch uses HF vapour to remove the oxide layer below the polycrystalline structures. The reaction is roughly as follow:

\[
\begin{align*}
\text{SiO}_2 + 2\text{H}_2\text{O} & \rightarrow \text{Si(OH)}_4 \\
\text{Si(OH)}_4 + 4\text{HF} & \rightarrow \text{SiF}_4 + 4\text{H}_2\text{O}
\end{align*}
\]

where all the final by-products are, of course, gaseous. The main issue with the technique is the high toxicity of the HF vapour and in Table 3.5 we describe two other popular methods which present less risk: dissolving polymer sacrificial layer with O\textsubscript{2} plasma, and using xenon difluoride (XeF\textsubscript{2}) to etch sacrificial silicon. The xenon difluoride is a gas showing an excellent selectivity, having etching rate ratio close to 1000 with metal and up to 10000 with oxide. The gas has thus been used successfully to release very compliant or nano-sized oxide structures where silicon was used as the sacrificial material. The process does not use plasma, making the chamber rather simple, and several manufacturers like XactiX (in cooperation with STS), in the USA or PentaVacuum in Singapore are proposing tools exploiting the technology.

### 3.5 DRIE micromachining

Deep reactive ion etching (DRIE) micromachining shares features both from surface and bulk micromachining. As in bulk micromachining the structure is etch in the bulk of the substrate, but as in surface micromachining a release steps is used to free the microstructure. Figure 3.19 shows a simplified process of bulk micromachining on silicon-on-oxide (SOI) wafer using deep reactive ion etching (DRIE), a special MEMS dry etch technique allowing large etch depth with very vertical side walls. The SOI wafers used in MEMS usually have a device layer thickness between 10 and 200µm where the structure is defined. After photolithography, the wafer is etched with DRIE to form high aspect ratio silicon structures, and the buried silicon dioxide acts as an effective etch stop. Stripping off the protective photoresist by O\textsubscript{2} plasma and then etching the sacrificial layer of the oxide using HF to release the microstructure finish the device. This simple, yet powerful,
technique needs only one mask to obtain working devices, and it is understandably used in commercial products. The best known example is the optical switch produced by Sercalo, a company founded by C. Marxer the inventor of this fabrication technique. DRIE has reached a large popularity in recent years among MEMS community and the tools produced by Adixen (Alcatel), Surface Technology Systems (STS) and Oxford System produce high aspect ratio structures (>25) with vertical sidewalls (>89°) at a decent etch rate (6µm/min or more).

A standard DRIE setting uses high density inductively coupled plasma (ICP) as the plasma source, and usually adopts the patented "Bosch process". The Bosch process is a repetition of two alternating steps: passivation and etching. In the passivation step, C₄F₈ gas flows into the ICP chamber forming a polymer protective layer (n(-CF₂-)) on all the surfaces. In the following etch step, the SF₆ gas in the plasma chamber is dissociated to F-radicals and ions. The vertical ion bombardment sputter away the polymer at the trench bottom, while keeping the sidewall untouched and still protected by the polymer. Then the radicals chemically etch the silicon on the bottom making the trench deeper. By carefully controlling the duration of the etching and passivation steps, trenches with aspect ratio of 25:1 have been routinely fabricated - and aspect ratio as high as 100:1 have been reached. Figure 3.20 is a SEM picture of some structures fabricated by DRIE on a SOI wafer.

The DRIE is a very versatile tool and allows a good control on the etched
profile slope, mostly by varying the etching/passivation duration and also by varying the substrate biasing voltage. However, it is affected by common RIE problems like microloading, where etching rate is slower for high density patterns than for low density ones. This effect is linked with the transport speed of reactant and products to and from the surface, which can be improved somewhat by increasing the flow rate of etching gas. But this is not the only issue, and other common issues are shown in Figure 3.21.

Figure 3.21: Some issues affecting DRIE process: scalloping, notching and lag (ARDE).

One issue with DRIE is the presence of regular ripple with an amplitude over 100nm on the vertical edge of trenches, a phenomena referred to as
scalloping. The ripples come from the repetition of isotropic etching and passivating steps, and can be annoying for etching nano-pillars with a few 100 nm diameter or for obtaining vertical wall with smooth mirror finish. Actually, they can be mostly removed by shortening the etching step to 1 s, instead of a standard 7 s, and by reducing the passivation step duration accordingly. This of course results in a much slower etching rate, but is a usual practice of trading etching speed for improving another etching parameter. The existence of DRIE lag is also a nuisance that needs to be considered. Actually, in narrow trenches, the limited evacuation rate of reaction products lowers the etching rate as compared to what happen in wide trenches. This is described as the aspect-ratio dependent etching (ARDE) effect. This effect again can be controlled by properly tweaking the recipe and trading a bit of etching speed.

Another major issue existing in DRIE is the fast silicon undertech that happens in SOI trenches when the etch reaches the buried oxide layer. Actually after the silicon has been completely etched, the oxide get exposed, and positive charges build-up in the insulating layer. This local space charge deviates the incoming ions laterally, causing an increased etch at the lower portion of the sidewall of the trench, an effect called notching. The most recent DRIE tools have managed to tackle this problem satisfactorily, by using a low frequency biasing scheme. Actually the normal RIE plasma frequency (13.56 MHz) is too high to have any effect on the ions, but by lowering the frequency to 380 kHz the ions bombardment will follow the field. In a way similar to what happens in a RF sputter, but this time for the ions, the ions during bias pulse won’t be anymore directed toward the substrate, letting the plasma electrons recombine within the charged insulator and suppress the spatial charge. By varying the cyclic ratio of the low frequency bias pulse it is thus possible to control the etching/uncharging timing, and obtain optimal etching rate while avoiding notching.

It should be noted that the notching effect can be put to good use and help produce an ‘etch and release’ process. Actually it has been found [21] that the notching effect is self limiting and the depth of the notch is roughly equal to the width of the trench as soon as the trench has an aspect ratio larger than 2 (for smaller aspect ratio there is no notching effect). In this way, by carefully designing the geometry of the layout, it is possible to etch the structure and finally obtain anchored or free structures within the same etching step. This simplifies further the DRIE fabrication process, and the device can now be operated right after emerging from the DRIE - without the need for a separate release etch!

The SOI wafer used often in DRIE machining is still expensive and it is possible to obtain the thick silicon structural layer by growing it using epitaxy.
on an oxidized wafer. Even more simply, DRIE has been used to etch through the Si wafer for a dry etched version of bulk micromachining - but allowing complete freedom over layout as there is no more crystallographic orientation concerns. In this case wafer bonding can be used to provide movable part.

3.6 Other microfabrication techniques

3.6.1 Micro-molding and LIGA

Other methods exist where no material is removed but this time molded to achieve the desired pattern. LIGA, a German acronym for lithography (Lithographie), electroforming (Galvaniformung), and molding (Abformung) is the mother of these methods.

LIGA makes very high aspect ratio 3-D microstructures with non-silicon materials such as metal, plastic or ceramics using replication or molding. LIGA process begins with X-ray lithography using a synchrotron source (e.g. energy of 2.4 GeV and wavelength of 2 Å) to expose a thick layer of X-ray photoresist (e.g. PMMA). Because of the incredibly small wavelength, diffraction effects are minimized and thick layer of photoresist can be patterned with sub-micron accuracy. The resist mold is subsequently used for electroforming and metal (e.g. nickel using NiCl$_2$ solution) is electroplated in the resist mold. After the resist is dissolved, the metal structure remains. This structure may be the final product but to lower down the costs, it usually serves as a mold insert for injection molding or hot embossing. The possibility to replicate hundreds of part with the same insert opens the door to cheap mass production. When the sub-micrometer resolution is not much of a concern, pseudo-LIGA processes can be advantageously used. These techniques avoid using the high cost X-ray source for the mold fabrication by replacing it by the thick photoresist SU8 and a standard UV exposure or even by fabricating a silicon mold using DRIE.

3.6.2 Polymer MEMS

Bulk and surface micromachining can be classified as direct etch method, where the device pattern is obtained by removing material from the substrate or from deposited layers. However, etching necessitates the use of lithography, which already includes patterning the photoresist, then why would we want to etch the lower layer when the pattern is already here? Actually lithography for MEMS has seen the emergence of ultra-thick photoresist that can be spun up to several 100 µm and exposed with a standard
mask aligner, providing a quick way to the production of micro-parts. SU8, a high-density negative photoresist can be spun in excess of 200 µm and allows the fabrication of mechanical parts [22] of good quality. It is used in many applications ranging from bioMEMS with micro-parts for tissue scaffold or channels, for example to packaging, where it is used as buffer layer.

Another application of thick photo-patternable polymer is the fabrication of microlenses using reflow at elevated temperature of thick positive photoresist pillars (e.g. AZ9260). Actually polymers usually have better optical properties than silicon in the visible, and there is a lot of opportunity for polymer micro-optical elements and system, a domain that is sometimes called Polymer Optical MEMS or POEMS.

Next to these major techniques, other microfabrication processes exist and keep emerging. They all have their purpose and advantages and are often used for a specific application. For example, quartz micromachining is based on anisotropic wet etching of quartz wafers to take benefit of its stable piezoelectric properties and build sensors like gyroscopes.

### 3.7 Problems

1. An optical telecommunication devices manufacturer wants to use microfabrication to produce V-groove for holding optical fiber. The most likely process to use is:
   - silicon substrate, photoresist mask and HF etchant
   - glass substrate, chromium mask and RIE etching with SF₆
   - silicon substrate, silicon nitride (Si₃N₄) mask and KOH etchant
   - silicon substrate, silicon dioxide (SiO₂) mask and HF etchant

2. A mask has the pattern of Figure 3.22(a). Which photoresist and pattern transfer technique could be used to pattern a thin film (in black) as shown in Figure 3.22(b)?

![Mask and film pattern](image.png)
- positive photoresist and lift-off
- negative photoresist and wet etching
- positive photoresist and RIE etching
- negative photoresist and lift-off

3. A circular hole in silicon is closed by growing oxide. The hole has a diameter of 2 µm.
   - Approximate the Deal and Grove’s equation when \( t \) is much larger than \( \tau \) and \( A^2/4A \).
   - How long will it take to close the hole if the long duration oxidation is performed at 1100°C in wet O\(_2\)? (Note: we have \( B = 0.51 \mu m^2/h \) at 1100°C)

4. Suggest a complete microfabrication process that can be used to fabricate the channel shown in Figure 3.23.

![Figure 3.23: Sealed micro-channel in glass.](image)

5. Propose a complete process based on KOH etching that could be used to produce the structure of Figure 3.24. Justify the use of a sandwich of SiO\(_2\) and Si\(_3\)N\(_4\). Could the two layers order be inverted?

![Figure 3.24: Membrane obtained by KOH etching.](image)
Chapter 4

MEMS technology

4.1 MEMS system partitioning

At the early stage of MEMS design an important question to be answered will be: hybrid or monolithic? Actually the decision to integrate the MEMS directly with its electronics or to build two separate chips has a tremendous impact on the complete design process. Most MEMS observer will advocate the use of separate chips and only in the case of a definite advantage (performance, size, cost) should a MEMS be integrated together with its electronics. From past industry examples, only a handful of companies, like Analog De-

Figure 4.1: Hybrid integration in a pressure sensor (Courtesy SensoNor AS - An Infineon Technologies Company).

device for its range of accelerometer or Motorola for its pressure sensors, have promoted the integrated process - and all are big companies having market reaching millions of chips. The hybrid approach in the other hand is used by many more companies on the market. For example Figure 4.1 shows a
hybrid solution from SensoNor, the pressure sensor SP15. The MEMS chip on the left is connected to the ASIC on the right using tiny wire and both are mounted in a metal frame before encapsulation in the same package. The advantage of this solution, dubbed ‘system in the package (SIP)’, is that both chips can use the best process without compromise and may achieve a better overall yield. However compactness and reliability suffers from the additional elements and the packaging becomes slightly more complicated. Moreover, the electronic is somewhat further from the sensing element and the tiny wires used for connection may introduce additional noise if the signal is small. It is this last argument that has pushed AD to develop its fully integrated accelerometer range, the iMEMS.

Figure 4.2: Flip-chip assembly of an IC chip on a MEMS chip.

More recently an intermediate solution has emerged that tries to reconcile both approaches: the flip-chip integration. In that case, shown in Figure 4.2, the electronics chip and the MEMS chip are interconnected using small solder balls that allows very tight assembly between them, with connection lead only a few m long. This technique is very promising, but is still more expensive to use and has the drawback to hide the MEMS chip, preventing its inspection or open use.

If a MEMS system can thus be partitioned between a pure MEMS part and electronics, it is also true that MEMS devices are a mosaic of diverse elements: mechanical structures, actuator, sensors, etc. For example, the MEMS optical switch shown in Figure 4.3 without its control electronics can be divided into a set of optical waveguides, a pair of electrostatic actuator, multiple springs and hinges, a lock and many linkage bars.

Actually, an interesting way to split the different structures used for the
Figure 4.3: MEMS optical switch composed of (red) optical waveguides (yellow) hinges and lock (purple) springs and suspension (green) electrostatic actuator (blue) alignment structure.

Design of MEMS is to consider separately passive and active structures:

- Active structures are at the core of actuator and sensor operation. Fundamentally their role is to allow a certain form of energy transfer between the environment and the system.

- Passive structures are used to support, guide, channel, etc. providing indispensable building blocks for the realization of complete systems. Their main role is to transport energy within the system.

4.2 Passive structures

4.2.1 Mechanical structures

Most MEMS required the development of special micro-mechanical elements to achieve standard mechanical function like linkage, suspension, articulation, etc.

The suspension, where the traditional fabrication of coiled springs is generally too complex to be considered, has seen many original development based on the elastic properties of beams. Actually the elasticity of a simple beam is well known\[9\]. We list in Table 4.1 for different cases shown in Figure 4.4.
Figure 4.4: Beams with different boundary conditions (cantilever, clamped-clamped beam, clamped-guided beam) in bending.

<table>
<thead>
<tr>
<th>Type</th>
<th>Deflection</th>
<th>Max deflection</th>
<th>Spring constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cantilever</td>
<td>$y = \frac{Fz^2}{6EI}(3L - z)$</td>
<td>$y(L) = \frac{FL^3}{3EI}$</td>
<td>$\frac{3EI}{L^3}$</td>
</tr>
<tr>
<td>Clamped-clamped beam</td>
<td>$y = \frac{F}{192EI}(12Lz^2 - 16z^3)$</td>
<td>$y(L/2) = \frac{FL^3}{192EI}$</td>
<td>$\frac{192EI}{L^3}$</td>
</tr>
<tr>
<td>Clamped-guided beam</td>
<td>$y = \frac{F}{12EI}(3Lz^2 - 2z^3)$</td>
<td>$y(L) = \frac{FL^3}{12EI}$</td>
<td>$\frac{12EI}{L^3}$</td>
</tr>
</tbody>
</table>

Table 4.1: Characteristics of beams in bending.

the deflection and spring constant when a point force is applied on the beam.

There $E$ is representing the Young’s modulus for the material of the beam and $I$ is the second moment of inertia for the beam cross-section, which is given by $I = \int \int x^2 dA = \frac{wh^3}{12}$ for a beam with a rectangular cross-section as shown in the inset.

But of course these beams are not enough and need usually to be combined to provide suspension with added flexibility. Actually for choosing a suspension there are usually four main characteristics to watch: the spring constant in the direction of use, the compliance in the other directions (it needs to be low to keep the motion in the desired direction), the tolerance to internal stress (long beam may buckle during fabrication) and the linearity during large deformation. There is of course a trade-off to be observed and different designs have been pursued (Figure 4.5) which have the characteristics shown in Table 4.2. As we see here the folded beam suspension is versatile and particularly suitable for process where there is a risk of buckling (it will stand large internal stress, as those appearing in surface micromachining) but more compact design may be suitable with other process. For example a clamped-clamped beam is an excellent choice in process with little or no internal stress like DRIE micromachining.
In the case of suspension made of combined beams the computation of the spring constant may be complicated but often existing symmetries allow to decompose the suspension into elementary beams connected in series and in parallel. For two beam connected in series, the equivalent spring constant is simply the sum of the two beams spring constants whereas if the two beams are in series, the resulting spring constant is the inverse of the sum of the inverse of the spring constants - in fact the spring constant behaves in a similar way as capacitor in electronic circuit. For example, it could be noted that the clamped-guided beam is in fact two cantilevers of half length connected in series, and the spring constant for the former can easily be deducted from the spring constant of the later.

Of course the force on the beam is not always concentrated and in some application, particularly in fluidic application, the beam is subject to a uniform line pressure, $q$. In that case there is no real ‘spring constant’ that can be defined (there is no force, $q$ is a pressure per unit of length in N/m) but the deflection can still be obtained and is given in Table 4.3 corresponding to the cases shown in Figure 4.6.

In addition to beams, MEMS often uses membranes or diaphragms for which the equation are usually more complicated. We give here the simpler case of round and square membranes clamped at the edge (Figure 4.7). These
Figure 4.6: Beams under pressure.

<table>
<thead>
<tr>
<th>Type</th>
<th>Deflection</th>
<th>Max deflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cantilever</td>
<td>$y = \frac{qz^2}{24EI}(z^2 - 4Lz + 6L^2)$</td>
<td>$y(L) = \frac{qL^4}{8EI}$</td>
</tr>
<tr>
<td>Clamped-clamped beam</td>
<td>$y = \frac{qz^2}{24EI}(z^2 - 2Lz + L^2)$</td>
<td>$y(L/2) = \frac{qL^4}{384EI}$</td>
</tr>
</tbody>
</table>

Table 4.3: Deflection of beams under pressure.

cases are corresponding to typical application for pressure sensor or valves, showing the general dependence of the characteristics with the geometry (Table 4.4).

Figure 4.7: Deflection of membranes.

Of course the case of deflection of membrane under pressure is of particular interest for pressure sensors, where the measurement of the deflection or the stress will be used to infer the pressure.

Besides suspension, other mechanical function, like hinge or joint, are often needed in MEMS. However the fundamental inability to miniaturize hinge because of the low relative manufacturing accuracy, forces designers to often use flexible micro-joint instead. This joints will have excellent wear characteristic but they will usually restrict rotation. These flexure hinge may use a simple cantilever or more complex beam arrangement. As we can see schematically in (Figure 4.8), the fork hinge [24] has the advantage to present a larger rotation angle for the same horizontal displacement than a standard cantilever beam with the same stability (resistance to buckling). Of course, if the angle of rotation need to be really large ($> \pm 20^\circ$), the alternative is to use a free hinge as shown in Figure 3.17 but the manufacturing complexity will increase substantially - and the reliability will drop.
<table>
<thead>
<tr>
<th>Type</th>
<th>Max Deflection</th>
<th>Spring constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round (Force)</td>
<td>$y_C = \frac{Fr^2}{16\pi D}$</td>
<td>$k = \frac{16\pi D}{r^2}$</td>
</tr>
<tr>
<td>Round (Pressure)</td>
<td>$y_C = \frac{gr^4}{64D}$</td>
<td></td>
</tr>
<tr>
<td>Square (Force)</td>
<td>$y_C = \frac{\alpha_F Fa^2}{Et^3}$</td>
<td>$k = \frac{Et^3}{\alpha_F a^2}$</td>
</tr>
<tr>
<td>Square (Pressure)</td>
<td>$y_C = \frac{\alpha_P qa^4}{Et^3}$</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: Deflection of round and square membrane (Plate constant $D = Et^3/[12(1 - \nu^2)]$, $\alpha_F = 0.014$ ($\nu = 0.3$), $\alpha_P = 0.061$ ($\nu = 0.3$).)

Figure 4.8: Flexure hinge and equivalent free hinge for (left) standard cantilever hinge (right) improved fork hinge

### 4.2.2 Fluidic structures

In microfluidic devices the ubiquitous passive element is the channel which is used to transport fluids. In general the problem of fluid flow, even using simpler incompressible fluid, is rather complex as it is governed by the Navier-Stokes equation. However, in some practical cases at micro-scale this equation can be simplified and has even a few analytic solutions. The distinction between these cases is based on a series of dimensionless coefficients, principally the Knudsen number and Mach number ($Ma = \frac{u}{c_s}$) for gas and the Reynolds number ($Re = \frac{\rho u L}{\eta}$) for gas and liquid. The fluid properties are noted as $u$ for the velocity of flow, $c_s$ the speed of sound in the fluid, $\rho$ the density of the fluid, $\eta$ the dynamic viscosity ($\eta = \rho \nu$ with $\nu$ the kinematic viscosity) and $L$ a characteristic dimension of the channel (usually its width). Interestingly, for value of Mach number smaller than 0.3, a gas can be considered as an incompressible fluid and use the same simplified equations as
Cross-section Fluid velocity and flow

<table>
<thead>
<tr>
<th>Cross-section</th>
<th>Fluid velocity and flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circular</td>
<td>$u(r) = -\frac{dp r_0^2 - r^2}{dx} \frac{dx}{4\eta}$ and $\dot{Q} = -\frac{dp \pi r_0^4}{dx} \frac{dx}{3\eta}$</td>
</tr>
<tr>
<td>Rectangular</td>
<td>$u(y, z) = -\frac{dp}{dx} \frac{16a^2}{\pi^2\eta} \sum_{i=1,3,\ldots}^{\infty} \left( -1 \right)^{i-1} \frac{1}{r^i} \left( 1 - \frac{\cosh(\pi z/a)}{\cosh(\pi b/a)} \right)$</td>
</tr>
<tr>
<td></td>
<td>and $\dot{Q} = -\frac{dp}{dx} \frac{40a^3}{3\eta} \left( 1 - \frac{192a}{\pi^5 b} \sum_{i=1,3,\ldots}^{\infty} \frac{\tanh(\pi b/2a)}{r^i} \right)$</td>
</tr>
</tbody>
</table>

Table 4.5: Fluid velocity and flow in channels with circular and rectangular cross-section (the channel is placed along $x$, $p$ is the pressure, $r_0$ the channel radius, $a$ the channel width and $b$ its height).

for liquids.

In cases of microchannel at the low flow velocity usually observed in microdevices, the Reynolds number is small ($<1500$) and the flow is laminar (that is, there is no turbulence). This complicates the tasks for producing mixers, as turbulence is the most efficient way to mix two fluids, but in that case, the flow induced by a difference of pressure (Poiseuille’s flow) can be obtained analytically as shown in Table 4.5 [16].

With the equations in the table it is possible to obtain the flow and the velocity across the channel section according to the pressure drop and to the geometry.

However it is not always necessary to know them precisely at any point across the channel and often it is enough to obtain the average velocity $\bar{u}$. The flow $\dot{Q}$ is related to the average fluid velocity by:

$$\dot{Q} = \bar{u} A \quad (4.1)$$

where $A$ is the channel cross-section area. In the case of pressure driven flow, the average fluid velocity can simply be expressed as:

$$\bar{u}_{pf} = \frac{2D_n^2}{Re f \eta L} \Delta p \quad (4.2)$$

where $Re f$ is the product of the Reynold’s number ($Re$) and $f$ the friction factor. This product can be computed from first principle for circular channel and we have $Re f = 64$. Experience shows that this value needs correction for micro-channels, where it is actually more in the $50 - 60$ range.

If the channel is not circular it is still possible to use the same equation, but we then we need to use an effective diameter, which is called the hydraulic...
diameter. For a channel of any cross-section we have:

\[ D_h = \frac{4A}{P_{\text{wet}}} \]  

(4.3)

where \( P_{\text{wet}} \) is the wetted perimeter, that is, the perimeter of the channel that is in direct contact with the fluid. We note that for a circular channel of diameter \( D \) we have \( D_h = 4A/P_{\text{wet}} = 4\pi(D/2)^2/2\pi(D/2) = D \), as we would expect.

These equations will give the pressure drop along a channel, or the flow rate if the pressure drop is known.

In micro-channel there is another phenomena that need to be considered, the problem of wetting. At the interface between gas, liquid and solid the equilibrium of surface tension forces (that is the resultant cohesive forces between neighboring molecules in the gas the liquid and the solid) can be written as:

\[ \sigma_{\text{sl}} + \sigma_{\text{lg}} \cos \theta = \sigma_{\text{sg}} \]  

(4.4)

where \( \sigma_{\text{sl}} \), \( \sigma_{\text{lg}} \) and \( \sigma_{\text{sg}} \) are the surface tensions at the solid-liquid, the liquid-gas and the solid-gas interfaces respectively, and \( \theta \) is the contact angle. The contact angle is used in particular to distinguish between hydrophobic and hydrophilic surfaces, with the exact criterion being that the contact angle should be larger than \( 90^\circ \) for the former and lower for the later. However in general this concept is often used in a less rigorous manner and describe relative value of different concepts as shown in Table 4.6

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Hydrophobic surface</th>
<th>Hydrophilic surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drop behavior</td>
<td>&gt;90°</td>
<td>&lt;90°</td>
</tr>
<tr>
<td>Contact angle</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Adhesiveness</td>
<td>poor</td>
<td>good</td>
</tr>
<tr>
<td>Wettability</td>
<td>poor</td>
<td>good</td>
</tr>
<tr>
<td>Solid surface free energy</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>

Table 4.6: Properties of hydrophobic and hydrophilic surface.

The wetting properties of a micro-channel results in a drop of pressure
across a liquid-gas interface given by the Young-Laplace equation as:

\[ \Delta p = \sigma_{lg} \left( \frac{1}{R_1} + \frac{1}{R_2} \right) \]  

(4.5)

where \( R_1 \) and \( R_2 \) are the two radii of curvature of the interface (this surface is three dimensional and for a circular channel we have \( R_1 = R_2 \)).

Using these two equations we can find the force exerted on a liquid in a circular micro-channel - also called a capillary. Actually in that case the pressure drop at the liquid-gas interface simplifies to:

\[ \Delta p = \frac{2\sigma_{lg}}{r \cos \theta} \]  

(4.6)

where \( r \) is the radius of the capillary. This resulting pressure difference can draw liquid in narrow channel (where \( r \) is small, thus \( \Delta p \) large), even balancing the gravitational force induced hydrostatic pressure (\( \Delta p = \rho gh \)) and thus allowing the liquid to rise in the capillary to the height \( h \). This property of fluid to rise in narrow channel made of material that they wet (actually the contact angle need to be smaller than 90° for the force to pull the fluid) is called capillarity. Capillarity makes it difficult to fill hydrophobic channels, as it tends to push the liquid back, or to empty hydrophilic ones, where it pulls the liquid inside. This effect is more pronounced for liquid with larger surface tension (Table 4.7), like water, which has one of the largest surface tension of common material due to the hydrogen bond between water molecules.

### 4.3 Sensor technology

Sensing is certainly a quality that we associate with living being. A stone does not sense, but can a silicon circuit do it? Of course, the answer is yes, and MEMS have increased tremendously the number of physical parameters that are sensed by silicon.

Sensing can be formally defined by the ability to transform the energy present in the environment in any form to energy inside a system. An example will be to convert the air temperature (heat) to an electrical signal by using a thermo-couple. At the heart of the sensor is the ability to perform the energy transformation, a process usually called transduction. MEMS sensor ability to measure different parameters as pressure, acceleration, magnetic field, force, chemical concentration, etc is based on a limited number of transduction principles compatible with miniaturization.
<table>
<thead>
<tr>
<th>Liquid</th>
<th>Liquid-gas surface tension (N/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>0.425</td>
</tr>
<tr>
<td>Sodium Chloride 6M</td>
<td>0.082</td>
</tr>
<tr>
<td>Water</td>
<td>0.072</td>
</tr>
<tr>
<td>Hydrochloric Acid (conc.)</td>
<td>0.070</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>0.048</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>0.023</td>
</tr>
<tr>
<td>Ethanol</td>
<td>0.022</td>
</tr>
<tr>
<td>Perfluorohexane</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 4.7: Surface tension for selected fluids at 20°C (the surface tension will decrease with the temperature).

4.3.1 Piezoresistive sensing

The oldest MEMS sensor that gained huge popularity was the pressure sensor and it was based on the piezoresistive effect. Piezoresistivity can be described by the change of resistance of a material when it is submitted to stress. This effect is known since the 19th century in metals, but it was only in the mid 1950s that it was recognized that semiconductor and particularly silicon had huge piezoresistive coefficient compared to metal[4]. The MEMS designer will place resistors obtained by doping silicon where the stress variation is maximal, for example at the edge of a membrane in the case of a pressure sensor. Then a simple Wheatstone bridge circuit (Figure 4.9) could be used to convert the resistance change $\delta R$ to a voltage difference. Actually, it is simple to show that if there is a single variable resistor in the bridge and if $\Delta R << R$ then

$$V_{out} \approx \frac{V_{in}}{4R} \Delta R.$$  

Moreover, if a judicious choice of variable resistors allows reaching the configuration shown in the right (where the variation of two variable resistors is opposite to the variation of the two other but with the same magnitude), then the sensitivity of the bridge increases fourfold and we exactly have

$$V_{out} = \frac{V_{in}}{R} \Delta R.$$  

77
For resistors with large aspect ratio in plane it is possible to write the relative change of resistance as:

\[
\frac{\Delta R}{R} = \pi_l \sigma_l + \pi_t \sigma_t
\]

where \( \pi_l \) is the piezoresistive coefficient and \( \sigma_l \) the stress component respectively, along the direction parallel to the current flow (l longitudinal) or perpendicular to it (t transverse). However the anisotropy in silicon, and actually in most crystals, makes it difficult to obtain the piezoelectric coefficients. Actually, all the physical parameters of silicon, like Young’s modulus or conductivity, depends on the direction with respect to the crystal axes in which they are measured. Thus, a complete treatment of piezoresistivity will involve complex mathematical object called tensors. Moreover, the piezoresistive coefficients in silicon depend on the type (n- or p-type) and concentration of doping, being generally larger for p-type resistors, and also on temperature. However for the most important cases the expression can be found in the literature and for example for p-type resistors placed in a n-type substrate along the (110) direction, that is, parallel to the wafer flat in (100) wafers, we have \( \pi_l \approx 71.8 \cdot 10^{-11} \text{ Pa}^{-1} \) and \( \pi_t \approx -66.3 \cdot 10^{-11} \text{ Pa}^{-1} \).

On a square membrane, for symmetry reasons, the stress in the middle of a side is essentially perpendicular to that side. Piezoresistor placed parallel or perpendicular to the side at that point will be, respectively, under transverse or longitudinal stress. If the membrane sides have been aligned with the (110) direction, the \( \pi_l \) and \( \pi_t \) are about the same magnitude but of opposite sign and the resistance of the two resistors under longitudinal stress in Figure 4.10 will increase when the membrane deforms while the resistance of the two resistors under transverse stress will decrease. It is thus possible to connect the four identical resistors in a full bridge configuration, as in Figure 4.9, and
Figure 4.10: Typical position of piezoresistors for a square membrane on (100) Si wafer.

the bridge sensitivity simplifies to:

\[ V_{\text{out}} \approx \frac{70 \cdot 10^{-11} V_{\text{in}} \sigma_{\text{max}}}{R} \]

where \( V_{\text{in}} \) is the bridge polarization voltage, \( \sigma_{\text{max}} \) the maximum stress in the membrane and \( R \) the nominal value of the piezoresistors. Although it seems advantageous this configuration is seldom used in practical devices because it suffers from a low manufacturability as the resistor positioning requires a very good accuracy. In general only one or two sensing resistors are used allowing simpler bridge balancing with trimmed external thick-film resistors. Piezoresitivity is not only used for pressure sensor but find also application in acceleration or force sensors. Unfortunately, the simplicity of the method is counterbalanced by a strong dependence on temperature that has to be compensated for most commercial products by more complex circuitry that the elementary Wheatstone bridge.

4.3.2 Capacitive sensing

Capacitive sensing is independent of the base material and relies on the variation of capacitance happening when the geometry of a capacitor is changing. Capacitance is generally proportional to \( C \propto \epsilon_0 \epsilon_r \frac{A}{g} \) where \( A \) is the area of the electrodes, \( g \) the distance between them and \( \epsilon_r \) the permittivity of the material separating them (actually, for a plane capacitor as shown above, the proportionality factor is about 1). A change in any of these parameters will be measured as a change of capacitance and variation of each of the three variables has been used in MEMS sensing. For example, whereas chemical or humidity sensor may be based on a change of \( \epsilon_r \), accelerometers have been based on a change in \( g \) or in \( A \). If the dielectric in the capacitor is air,
capacitive sensing is essentially independent of temperature but contrary to piezoresitivity, capacitive sensing requires complex readout electronics. Still the sensitivity of the method can be very large and, for example, Analog Devices used for his range of accelerometer a comb capacitor having a suspended electrode with varying gap. Measurement showed that the integrated electronics circuit could resolve a change of the gap distance of only 20 pm, a mere 1/5th of the silicon inter-atomic distance.

### 4.3.3 Other sensing mechanism

A third commonly used transduction mechanism is based on piezoelectricity. Piezoelectricity occurs when stress applied on a material induces the apparition of charge on its surface. Silicon does not present piezoelectricity but crystalline quartz has a large piezoelectric coefficient and other material like ZnO or PZT can be deposited in thin films possessing piezoelectric properties. The advantage of piezoelectricity is that it can be used to sense stress but also as an actuator too. Actually a difference of potential applied on two sides of a piezoelectric layer will induce its deformation. Thus piezoelectric material can be excited in vibration and the vibration sensed with the same structure. This has been the heart of the quartz watch since its invention in the 1970’s, but it is also used for different inertial MEMS sensor like gyroscope.

Magnetic sensing, although less often used, has its supporters mainly because it is a non-contact sensing mechanism with a fairly long range. Its main application has to be found in the (giant)magnetoresistive effect used inside the hard-disk head. However other uses of magnetic sensing have been tested and for example some sensors have been based on the Hall effect taking advantage of the simplicity to manufacture this sensing element.

### 4.4 Actuator technology

Since the industrial revolution we have understood that machines can perform task with more force and endurance than humans. Bulldozers moving around with their huge engine and pushing big rocks with their powerful pneumatic actuators are probably a good example of what a big machine can do. But what will be the function of a micro-sized actuator?

The main parameters useful to describe an actuator are its force and its stroke. However we have seen previously that all forces decrease with the scale, thus we can not expect to move big rocks around - but only micro-rocks. The micro-actuators are currently used to act on micro-object, typically one
part of a MEMS device, and generate forces in the micro to milli Newton range with a stroke from a few µm to several hundreds µm. It would be interesting to have enough force and stroke to allow actuator to help interface human and machine by providing force feedback for example, but micro-actuators are still unable to do that properly.

Still a wide range of principles exists that would transform internal energy of a system (usually electrical energy) to energy in the environment (in the case of MEMS, generally mechanical energy). Sometimes the conversion from electric energy to mechanical energy is direct but often another intermediate energy form is used. For example, the heatuator, a form of thermal actuator, uses current to generate heat which in turn becomes strain and displacement.

The MEMS actuators can be conveniently classified according to the origin of their main energy form. In Table 4.8 we compare the most common MEMS actuators, where Efficiency refers to the loss existing in the actuator conversion of electrical energy to mechanical energy and Manuf. is the manufacturability or the simplicity of micro-fabrication.

<table>
<thead>
<tr>
<th>Type</th>
<th>Force</th>
<th>Stroke</th>
<th>Efficiency</th>
<th>Manuf.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electromagnetic</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gap-closing</td>
<td>0</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Electrostatic</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Comb-drive</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>SDA</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Piezoelectric</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bimorph</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Heatuator</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Shape memory alloy (SMA)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Thermo-fluidic</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.8: Comparison of common micro-mechanical actuators.

4.4.1 Magnetic actuator

Electromagnetic actuation is well known for providing the actuator used in house appliances, toys, watches, relays... The principle of electromagnetic
motor is well known and it is tempting to miniaturize such a versatile device to use it in the micro-world. However an electromagnetic motor with its coils, armature and bearings prove a tremendous task for micro-fabrication and so far nobody has been able to batch produced a motor less than 1mm diameter.

Still magnetic actuation has many proponents and some version of magnetic linear actuator have been used in different devices. Such a mobile armature actuator is shown in Figure 4.11 where by increasing the current in the coil the mobile armature is attracted along the x direction to align with the fixed armature.

![Figure 4.11: Mobile armature linear magnetic actuator.](image)

The magnetic force produced on the mobile armature is linked to the change of reluctance and is given approximately by [12]:

$$F_{ma} = \frac{(nI)^2}{2w} \left( \frac{\mu_0 A}{\mu + \mu_0 L/\mu} \right)$$

From this equation it is clear that the force is non linear with the current, and assuming a constant resistance for the coil, the force will also depend on the square of the coil voltage.

Although this force does not scale very favorably, the possibility to increase the current at small scale, because the heat can be dissipated more quickly, still allows producing relatively strong force. However the main difficulty preventing the widespread use of this type of actuator in a MEMS component is the fabrication of the coil. In that case the most convincing approach proposed so far are most probably those using a hybrid architecture, where the magnetic circuit is fabricated using micro-fabrication but the coil is obtained with more conventional techniques and later assembled with the MEMS part. Actually some design have shown that the coil does not need to be micro-fabricated at all and can be placed in the package, taking benefit of the long range action of the magnetic field.
Finally it should be noted that magnetic actuation can be used in conjunction with ferro-magnetic material to provide bistable actuator where two positions can be maintained without power consumption. A permanent magnet placed in the package is used to maintain the magnetized ferro-magnetic material in place. Then, when we send a current pulse of the right polarity in a coil wound around the ferro-magnetic material we invert its magnetization and the actuator switch to its second state. NTT has been producing since at least 1995 a fiber optic switch based on a moving fiber with a ferro-nickel sleeve that has two stable positions in front of two output fibers [13]. The device will consume power only during the brief time where the current pulse is sent and can maintain its position for years.

4.4.2 Electrostatic actuator

A physical principle that leads itself well to integration with MEMS technology is electrostatics. Actually by applying a potential difference between two electrodes, they develop charges of opposite sign and start attracting each other through the Coulomb’s force. This principle has known several application among which, the comb-drive actuator, the gap-closing actuator and the scratch drive actuator are the most commonly used (Figure 4.12). From energy consideration it is easy to show that the force developed between two electrodes is proportional to the change (derivative) of their capacitance multiplied by the square of the voltage ($F_{\text{elec}} \propto dC/dx V^2$). This fundamentally shows that electrostatic actuators develop force non-linear with voltage.

![Figure 4.12: Different type of electrostatic actuators.](image)

The comb-drive actuator was invented by W. Tang [14] at UC Berkeley and it generally allows motion in the direction parallel to the finger length. The force produced by $n$ fingers in the rotor is approximately given by

$$F_{\text{cd}} \approx n \varepsilon_0 \frac{h}{g} V^2$$

where we see the expected dependence with the square of the voltage and notice that it is independent of the displacement $x$. The proportionality
factor is $\epsilon_0$, a small quantity indeed, hinting to a small force generated per finger, in the order of a few 10 nN. Of course the number of fingers can reach 100 or more and the actuator aspect-ratio can be made larger (i.e., increase $h/g$) to increase its force proportionally. This actuator has been used repeatedly in MEMS component, for example in the original Analog Devices accelerometer or in the fiber optic switch from Sercalo.

The origin of the force parallel to the electrodes results from the unbalance of the Coulomb’s force along some part of the finger. Actually the rotor charges located in (a) will experience a symmetrical attraction resulting in an absence of net force, whereas, charges located in (b) will, as a result of the unbalanced attraction, create a net force pulling the rotor parallel to the stator. We note that the Coulomb’s force also results in a force perpendicular to the surface, however the motion toward the stator is prevented by the rotor suspension and by the balancing force of the stator finger placed on the other side of the rotor finger.

The gap-closing actuator actually makes use of this force perpendicular to the electrodes surface and usually delivers a larger force (proportional to $A$). This force is again non linear with the applied voltage, but additionally it now depends on the displacement $x$.

$$F_{gc} \approx \epsilon_0 \frac{A}{2x^2} V^2$$

It can be shown that, when the actuator is used in conjunction with a spring (usually a bending beam) used to polarize and retain the rotor electrode, the rotor position can only be controlled over a limited range. Actually, as can be seen in Figure 4.13 as soon as the rotor has moved by one third of the original gap width ($g$), snap-in suddenly occurs and the rotor comes into contact with the stator (in the figure we show two blocks in black used to prevent the contact between the electrodes and a short-circuit, the position of these blocks will determine the pull-out voltage when the voltage is decreased).

The pull-in voltage is given by:

$$V_{\text{pull-in}} = \sqrt{\frac{8}{27} \frac{k g^3}{\epsilon_0 A}}$$

where $g$ is the original gap width and $k$ the rotor suspension spring constant.

This behavior can be advantageous if the actuator is used for bi-stable operation, but as here, preventive measures should be taken to avoid electrodes short-circuit. Actually, the actuator behind the Texas Instruments’ DLP is a gap-closing electrostatic actuator working in torsion with the two stable
Figure 4.13: Voltage-Displacement curve for a gap-closing actuator showing the pull-in phenomena.

states position fixed by resting posts. By biasing the actuator at a voltage in the middle of the hysteresis curve, it needs only a small swing of voltage to allow a robust bi-stable actuation.

The scratch drive actuator is a more recent invention by T. Akiyama [15] and although it is actuated by electrostatic force, the friction force is the real driving force. As we can see in the diagram, the electrostatic energy is stored in the SDA strain while its front part, the bushing, bends. When the voltage is released, the strain energy tends to decrease and the bushing returns to its rest orientation producing displacement. The main advantage of this actuator is that it is able to produce a rather large force (100 \( \mu N \)), which can be even increased by connecting multiple actuators together. Actually the SDA has been used as an actuator in the 2D optical switch matrix that was developed by Optical Micro Machines (OMM) and which received the stringent Telcordia certification.

Actually, electrostatics can also be used to move liquids using two phenomena: electro-hydrodynamic which works with non-conductive fluids and electro-osmosis which works with conductive fluids. Electro-osmosis pumps are of larger significance because biological fluids are actually solute with different ions (salts) and are thus conductive. In an electro-osmosis pump a stationary electrical field is applied along a channel and result in an overall motion of the conductive fluid. Although the liquid is globally neutral, this motion can happen because of a so-called double layer of charged particles near the walls of the channel. In general with conducting fluids a surface charge appears on the insulating walls, usually made of polymer or glass, resulting in a negative potential (the zeta potential \( \zeta \)) ranging generally between -20 mV and -200 mV. Those trapped
charge attracts positive ions in the solution close to the wall creating the
double layer of a few nm thick, as shown in Figure 4.14. Then, the external
electrostatic field makes this layer move, entraining, because of the shear
force in the liquid with non-zero viscosity, all the liquid in the channel. The
flow velocity profile, contrary to pressure driven flow, is uniform across the
channel (plug flow) and we have:

\[ \bar{u}_{\text{eof}} = u_{\text{eof}} = \frac{\epsilon \zeta}{\eta} E \]  

(4.7)

where \( \epsilon \) is the dielectric constant of the fluid and \( \eta \) its viscosity. Actually
the proportionality constant between the velocity and the field is usually
rather small (\( \mu_{\text{eo}} = \epsilon_0 \epsilon_r \zeta / \eta \), \( \mu_{\text{eo}} \) electro-osmotic mobility) and in practice the
voltage used for electro-osmosis flow across a 10 mm-long channel will be in
the order of 1000 V.

### 4.4.3 Thermal actuator

The thermal energy used by this class of MEMS actuator comes almost invari-
ably from the Joule effect when a current flows through a resistive element.
These actuators are generally relatively strong and their main drawback is
most probably their speed, although at micro-scale the heat is quickly radi-
ated away and operating frequency up to 1 kHz can be achieved. Bimorph
actuators are the most common type of thermal actuator. The bi-material
actuator, well known from the bimetallic version used in cheap temperature
controller, and the heatuator (Figure 4.15) are both bending actuator where
bending is induced by a difference of strain in two members connected to-
gether.

The bi-material actuator obtains this effect by using two different materi-
als with different coefficients of thermal expansion that are placed at the same
temperature resulting in a misfit strain \( \epsilon_m = (\alpha_1 - \alpha_2) \Delta T \). The curvature of
a bi-material actuator is given by:

$$\kappa = \frac{6E_1E_2(h_1 + h_2)h_1h_2\epsilon_m}{E_1^2h_1^4 + 4E_1E_2h_1^3h_2 + 6E_1E_2h_1^2h_2^2 + 4E_1E_2h_2^3h_1 + E_2^2h_2^4}$$  \(4.8\)

where $E_i$ is the Young’s modulus for the two materials and $h_i$ their thickness.

Actually the difference in thermal expansion existing for materials deposited at different temperatures (e.g., polysilicon and metal) makes any bilayer curl when it is released at room temperature. This effect is often annoying, and if it can be controlled to some extent, it is the main issue behind the use of bilayer actuator. Actually, for such actuator, upon release the bilayer will curl (up if it has been well designed!) and as its temperature changes (e.g. using Joule’s heating by flowing a current) its radius of curvature evolves – but it will be difficult to make it flat. Still, the initial stress induced curvature has been put to good use to fabricate curling beam that naturally protrude high above the surface of wafer in a permanent way. They have been used to lift other MEMS structure (e.g. micro-assembly in Figure 5.1) or micro-parts (e.g. conveying system).

The heatuator \textsuperscript{17} does not have this problem as it uses a single material. This simplifies the fabrication, and the difference in strain is obtained by maintaining different temperature in the two arms. Actually as the current flow through the actuator the wider ‘cold’ arm will have a lower resistance and thus generate less heat than the other narrow ‘hot’ arm.

It should be noted that the force produced by these two actuators decreases with the stroke: at maximum stroke all the force is used to bend the actuator and no external force is produced. One heatuator can produce force in the 10 µN range and they can be connected together or made thicker to produce larger force.

The thermo-pneumatic actuator is another actuator where the expansion of a heated fluid can bulge a membrane and produce a large force. This principle has been used to control valve aperture in micro-fluidic components. In the extreme case, the heating could produce bubble resulting in large
change of volume and allowing to produce force as in the inkjet printer head from Canon or HP.

Finally the shape memory effect is also controlled by temperature change and traditionally belongs to the class of thermal actuator. The shape memory effect appears in single crystal metal like copper and in many alloys among which the more popular are NiTi (nitinol) or Ni$_x$Ti$_y$Cu$_z$. In such shape memory alloys (SMA) after a high temperature treatment step, two solid phases will appear one at low temperature (martensite phase) and the other at high temperature (austenite phase). The alloy is rather soft and can be easily deformed at low temperature in the martensite phase. However, upon heating the alloy above its phase transition temperature it will turn to austenite phase and returns to its original shape. This process creates large recovery forces that can be used in an actuator. The temperature difference between the two phases can be as low as 10°C and can be controlled by changing the composition of the alloy. In principle the alloy can be ‘trained’ and will then shift from a high temperature shape to a low temperature shape and vice-versa when the temperature is changed. In practice, training is difficult and micro-actuators based on SMA are one way actuator, the restoring force being often brought by an elastic member, limiting the total deformation. The most common application of such material has been for various micro-grippers, but its use remains limited because of the difficulty in controlling the deposition of SMA thin-films.

4.5 Problems

1. Establish the expression of the spring constant of the folded beam suspension. You may want to consider the symmetry existing in the structure and decompose it as a set of cantilever beams connected in series and in parallel.

2. We consider a micro-cantilever of length $L$, width $w$ and thickness $h$ bending under its own weight.

   - What is the expression of the weight per unit of length of the cantilever assuming the material has a density of $\rho$?
   - What is the general expression of the deflection at the tip of the cantilever?
   - What is the length of a 2 $\mu$m thick silicon cantilever whose tip deflects by 2 $\mu$m? (Note: the density and Young’s modulus of
silicon are \( \rho = 2.33 \cdot 10^3 \text{kg/m}^3 \) and \( E = 106 \text{GPa} \), the acceleration of gravity is \( g = 9.81 \text{ m/s}^2 \).

- What are the practical implications of this deflection for the cantilever?

3. A force sensor is based on a piezoresistor placed on a cantilever. For which of the layout in Figure 4.16 will the force sensitivity be the highest when the force is applied at the tip of the cantilever? (Note: we have \( \pi_l = -31 \cdot 10^{-11} \text{Pa}^{-1} \) and \( \pi_t = -17 \cdot 10^{-11} \text{Pa}^{-1} \).)

![Figure 4.16: Design of force sensor.](image)

4. A channel etched in silicon by KOH is then sealed with wafer bonding resulting in the cross-section shown in the Figure 4.17. What will be the expression of the flow as a function of the pressure drop for such a channel?

![Figure 4.17: Channel cross-section.](image)

5. Establish the equation (eq. 4.6) for the pressure drop in the capillary shown in Figure 4.18, starting from the expression of the Young-Laplace’s equation (eq. 4.5).
6. A micro channel of diameter 100µm is placed vertically above a reservoir. How high will water rise in the channel if its surface is bare silicon (contact angle 63°) or if it has been coated with silicon nitride (contact angle 24°)? What happens in the capillary if water is replaced by mercury?

7. Establish the expression of the pull-in voltage in the electrostatic gap-closing actuator shown in Figure 4.13. You can start first by writing the equilibrium equation between the restoring spring force and the electrostatic force.

Figure 4.18: Capillary cross-section.
Chapter 5

MEMS packaging, assembly and test

MEMS packaging, assembly and test – collectively called back-end process – problems are the aspects of the MEMS technology that even now remain the less mature. Actually, although the bookshelves appear to be replete with books discussing all the aspects of MEMS technology, we had to wait until 2004 to finally have a reference book really discussing these three issues with real MEMS examples [26]. However it is hard to stress enough how MEMS packaging and test are important for obtaining a successful product at a low cost. Figure 5.1 shows two real life examples of MEMS based micro-sensor where packaging, assembly and test are a dominant part of the total cost.

![Figure 5.1: Cost break-up in (a) a pressure sensor in plastic package (b) an accelerometer in surface-mount package](image)

As a matter of fact, the choice taken for packaging and test may dictate how to design the MEMS chip itself! This is in sharp contrast with micro-electronics packaging where packaging is a somewhat independent activities
than chip processing or design. On the contrary, in the MEMS case (no pun intended!), the influence of the package on the microsystem behavior may be very significant.

The Example 5.1 shows clearly that early consideration of the packaging solution could lead to a successful product... while ignoring it could lead to a dramatic failure.

The different operations that could appear in a somewhat complete MEMS back-end process are presented in Figure 5.2, and we find the assembly, packaging and testing steps. However as we stressed earlier, nothing is less typical than a ‘standard’ MEMS back-end process - and in practice, some steps may not be present or their order be different.

![Figure 5.2: Standard packaging steps for MEMS](image)

The general tendency, to decrease the cost, is to perform as many steps as possible at the wafer level and not for individual die. Accordingly, recent years have seen the emergence of techniques like wafer Level Packaging (WLP), where the back-end process can hardly be distinguished with the front-end process. However, in most cases, many operation still need to be performed component by component, explaining largely the cost of this step in the MEMS production.

### 5.1 Assembly

Assembly of MEMS part is normally not a good idea as it traditionally requires part-by-part processing and loose the cost advantage attached to batch processing.

However there are examples that show it can still be used and we have discussed previously [4.1] that hybrid integration is the normal answer to electronics integration with MEMS. The IC chip is assembled with the MEMS chip either quasi-monolithically using flip-chip solder bump, or more simply using in-package assembly (Figure 3.6) : the MEMS part is placed inside the package along the electronics chip and connected using wire-bonding. Such systems are referred to SIP (System In the Package) and contrast strongly with the SOC (System On Chip) approach, where MEMS and electronics
We consider a thermal sensor, used in a flow sensor and look at two different scenarios for the packaging: either the sensing element is directly in contact with the fluid or, because the fluid is corrosive, it is protected by a sheath.

In the first case the sensor is a first order system, where $R$ is the thermal resistance (the way it oppose change in temperature) of the sensing element and $C$ its thermal capacitance (how it stores heat).

Thus its transfer function is given by:

$$\frac{T_s}{T_f} = \frac{1}{1 + \tau_s s}$$

with $\tau_s = R_s C_s$

However, when a sheath is placed around the sensing element it brings an additional thermal resistance ($R_p$) and thermal capacitance ($C_p$), and the sensor becomes a second order system.

with the transfer function :

$$\frac{T_s}{T_f} = \frac{1}{1 + (\tau_p + \tau_s + \tau_s R_p/R_s)s + \tau_s \tau_p s^2}$$

with $\tau_p = R_p C_p$ and $\tau_s = R_s C_s$

The dramatic differences existing between responses of first and second order systems clearly underscore the necessity to include the packaging at an early stage of the design.

**Example 5.1:** Packaging affecting MEMS response
are built together on the same die like in the ADXL accelerometer from Analog-Devices.

Assembly is sometimes a viable solution, and it is actually even possible to have real MEMS part assembly. After all, watch-makers assemble small parts and still manage to reach very low cost using automation. Accordingly, different automated MEMS assembly systems have been demonstrated, and some showed success for tasks with reduced manipulation need. However complete free 3D assembly of MEMS part (6DOF) with the necessary accuracy and precision (< 1µm) is still too slow to be of practical use.

However MEMS offers a much clever path to assembly than serial processing: the batch assembly, which maintains the advantage of batch processing. This is accomplished along two paths: self-assembly or integration of assembly micro-actuator.

In the first case, natural forces like capillary forces, are used to pull and position the MEMS element in place.

In the second approach, at the same time the MEMS is developed, micro-actuator are integrated to realize the assembly task. In this way, we can recoup some of the overhead brought by assembly, because now, the mechanical assembly can be performed by batch. Actually for allowing complete assembly, the MEMS should not only include an assembly actuator but also some locking structure that will keep the assembly in position even after the special actuators are no more powered. As they are used only for a short time, these actuators can have a short lifespan (e.g., SDA actuator) even working one time only (e.g., stress based actuator as in Example 5.2), require large power (e.g., heatuators) or high voltage (e.g. large force comb-drive actuator) without posing too much problem.

5.2 Packaging

MEMS packaging, unlike the well-established and standardized IC packaging technology, is still largely an ad-hoc development. The main packaging efforts have been conducted within MEMS manufacturer companies, and they have jealously kept their secret considered, with reason, as the most difficult step to bring MEMS to market.

Still, the purpose of packaging in MEMS is in many ways similar to IC, and we can list a series of functionalities that should be brought by the package:

- Support: provide a standard mechanical support for handling during assembly of the MEMS part into a system.
Alcatel-Lucent Bell labs did develop at the turn of the millennium (and in 18 months!) the Lambda router (marketed until mid 2002 as the core component in their Wavestar all-optical routing system) using an intricate structure of integrated stress-driven vertical actuator to assemble array of two-axis gimbal mirrors.

The complex mirror structure needed a fast electrostatic command that could be easily obtained with surface micromachining. However they also needed a large tilting angle that this process could not directly provide as deposited sacrificial layer are too thin to obtain enough space below the 500 µm wide mirrors. They had to develop a clever assembly process to obtain from the flat surface-micromachined layers the 3D structures they needed.

![Diagram of dual-gimbal mount with blocking support](image)

The mirror is actually pushed above the surface of the wafer using beams curling up permanently under stress gradient (see Section 4.4.3) and held in operating position by additional blocking structure. The Figure shows the principle of the device which get automatically assembled during the release etch: at first (1) the narrow bilayer beams get freed positioning the locking structure above the gimbal mount ‘ears’, then (2) the wider bilayer beams are freed, curling up and pushing the mirror against the blocking structure that self-align using the V-shaped frames.

Example 5.2: Stress-driven automated assembly of tunable micro-mirror.

- Protection: protect the chip from the environment (dust, stress, shock, moisture...). For MEMS the most important parameter to be controlled is often stress.
- Interfacing: bring signal in and out of the chip. For MEMS, signals will not only be electrical but may be fluids, radiation, fields...

- Heat removal: ensure the heat generated inside the chip is properly evacuated to the environment (it is actually only a special kind of interfacing, but not for a signal). Actually, this point is much less severe with MEMS than with IC and is usually relatively easy to fulfill.

The need for protection from the environment is actually for MEMS not only for reliability (e.g. preventing corrosion of metal contact) but it often serves to solve simultaneously a functional problem. For example, hermetic encapsulation protects die from contamination but it is also used in pressure sensor to obtain a reference pressure in a cavity, or to control damping in resonant system. But the required protection level in MEMS is often higher than for IC. Actually the presence of water vapor - that could condensate during use on any mobile part - may make hermetic package a must have and not only for pressure sensor. For example, a water droplet appearing inside the TI’s DMD mirror array would induce defects hard to accept in a product of that price.

To understand some of the challenge that lies in the design of MEMS packaging we may have a look at the Table 5.1 adapted from [25], where, next to some typical MEMS micro-sensors, we have also figured the requisite for micro-electronics packaging.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>maybe</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Flow</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Accel.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>maybe</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Yaw-rate</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>maybe</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Sound</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Light</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>maybe</td>
<td>no</td>
</tr>
<tr>
<td>Temp.</td>
<td>yes</td>
<td>maybe</td>
<td>no</td>
<td>maybe</td>
<td>no</td>
<td>maybe</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>IC</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>moisture</td>
<td>no</td>
<td>maybe</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 5.1: Packaging and testing requirement for some micro-sensors (adapted from [25]).

It is obvious that the challenges brought by micro-sensors packaging are completely different from those encountered by electronics packaging. The necessity for the measurand to reach the sensing element brings in transparent windows, fluid port, gas hermetic sealing, stress isolation... unheard of in the IC industry, and unfortunately, in the IC packaging industry.
In fact, besides protection, the major hurdle often rests in interfacing to the external environment. Actually this problem is diametrically opposed to the preceding point: interfacing requires us to open a way in (and out) through the protection to come close to the MEMS die. For inertial sensors, such as accelerometers and gyroscopes, the packaging problem is not too severe because they can be fully sealed and still sense the measurand they are to probe provided they are rigidly attached to the package. In that case, the use of stress relieving submount and a bonded cap is all what’s needed to be able to use modified IC packaging procedure. But this is for the simplest cases, for chemical and biological sensors, which must be exposed to fluids, the task is much more complex and the package can represent as much as 90% of the final cost. Actually, the diversity of issue encountered for interfacing has for the moment received no standard solution and the packages are then designed on a case-by-case basis. In many cases the package will condition the response of the MEMS, particularly in the case of microsensor, and the package must be considered during design at the earliest possible stage. See for example the gas sensors developed by Microsens in Figure 5.3. The package use a charcoal filter placed inside the cap for a very important function: decreasing cross-sensitivity by allowing only small molecule gas to go through and to reach the sensing element. The time for the gas to diffuse through the filter determines mostly the response of the sensor, that behaves as a first order system with a time constant in the order of 10 s, whereas the response of the sensing element (the ‘MEMS’ die) is shorter than 1 s. The package has definitely a dramatic effect on the system response!

![Figure 5.3: A gas sensor and its package developed by Microsens](image)

97
5.2.1 Encapsulation

Encapsulation is used to protect the MEMS from mechanical contact, dust, water vapour or other gases that could affect the reliability of the MEMS. Encapsulation is performed by using plastic, ceramic or metal, with cost increasing in the same order. For example, a pressure sensor from Novasensor packaged in plastic may be sold for less than US$5, while a steel housed sensor with metallic membrane for harsh environment may exceed the US$100 mark! NovaSensor is proposing the three types of packaging: the plastic package for a low-cost low performance sensor, a ceramic package for compensated structure for medical application, a metal case (modified from a standard TO-8 cap from the early IC industry) and a full metal package (the interface to the environment is a metallic membrane) for corrosive or harmful media. But we should note that the MEMS sensing element used in these three sensors is exactly the same - the value of the sensor is mostly in the package!

Figure 5.4: Box and lid package in metal (TO-header) and ceramic (CERDIP).

As MEMS generally have mobile parts, the main difficulty of MEMS encapsulation is to avoid blocking this motion. This originally forbid the simple use of injection molding after mounting on a lead frame as in standard IC packaging. The earlier idea were to use existing ‘box and lid’ type of casing, like the older TO series of package used for transistors or the ceramic CERDIP package, as we see in Figure 5.4. This package have been adapted to the specific MEMS application, and for example we have figured an original Transistor Outline (TO) case modified by welding a fluid entry port to the cap for using it in a pressure sensor. Later on, to reduce cost, injection molded plastic ‘box and lid’ cases were used, most notably in Motorola’s consumer grade range of pressure sensors. The CERDIP case solves the second main issue generally encountered in MEMS packaging: thermal stress. Many MEMS devices are actually stress sensitive, and besides the obvious piezore-
Table 5.2: Characteristics of some packaging materials (T: tensile, S: shear).

<table>
<thead>
<tr>
<th>Material</th>
<th>CTE (ppm/K)</th>
<th>Modulus (T/S)</th>
<th>Yield strength (T/S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>4.1</td>
<td>150</td>
<td>N/A</td>
</tr>
<tr>
<td>Al (pure/alloy 1100)</td>
<td>23</td>
<td>69</td>
<td>35/60</td>
</tr>
<tr>
<td>Al (Si alloy)</td>
<td>6.5-13.5</td>
<td>100-130</td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>17</td>
<td>117</td>
<td>70</td>
</tr>
<tr>
<td>Ni</td>
<td>13</td>
<td>207</td>
<td>148/359</td>
</tr>
<tr>
<td>In</td>
<td>29</td>
<td>11</td>
<td>1.9/6.1</td>
</tr>
<tr>
<td>Alumina (Al₂O₃)</td>
<td>6.7</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Kovar</td>
<td>5.9</td>
<td>131</td>
<td>340</td>
</tr>
<tr>
<td>Borosilicate glass</td>
<td>4.5</td>
<td>65/26</td>
<td>29</td>
</tr>
<tr>
<td>Epoxy (pure)</td>
<td>60</td>
<td>2.4/0.9</td>
<td>54</td>
</tr>
<tr>
<td>Epoxy (carbon fiber)</td>
<td>-1.1</td>
<td>186</td>
<td></td>
</tr>
</tbody>
</table>

\[ \sigma_T \propto E \alpha T \]
As polymer have usually small Young’s modulus, they won’t exert much force on the silicon die as they will deform readily and absorb much of the induced strain. In fact soft polymer are often used as a stress relieving buffer to attach the die to the casing for example. More surprisingly, some material, that possess a high CTE and a relatively high Young’s modulus can still be used in packaging. The best example is given by indium. This metal can be used for solder or as a paste for die attachment because, although it has a relatively large CTE and Young’s modulus, its yield strength is very low. Thus as the thermal strain changes, the alloy will quickly deform plastically and again won’t induce any excessive stress on the MEMS die.

If MEMS encapsulation is still too often an adhoc development, some strategies are maturing to keep as much as possible of the cost advantage brought up by batch fabrication. As such, more and more MEMS devices use first level encapsulation, where a glass or silicon wafer is bonded to the chip, helping to maintain the MEMS integrity during dicing and further mounting in the package. In the packaging with wafer bonding technique (cf. Sec. 3.3.3) the cap wafer is first patterned with a simple cavity, or even a hole if an access to the environment is needed. Then, alignment of the MEMS wafer and the cap wafer brings the cavity in front of the MEMS part before the bonding is finally performed (Figure 5.5). This pre-encapsulation technique has the advantage to be wafer level and to provide an inexpensive way to allow the use of the epoxy overmolding to package the component. Actually, after dicing the wafer in chips, the capped MEMS is rather sturdy and can be processed using standard IC packaging procedure.

The first step, shown in Figure 5.6, consists in placing each die on a lead frame, a long strip of identical metal structures punched from a thin foil. The lead frame is used as a support for the die and to obtain electrical connection that can be soldered on a printed circuit board (PCB). The dies
are first glued on each of the die bonding site using polymer or indium. Then electrical wiring is done to connect the pad on the MEMS chip to the lead frame contacts that reroutes them to the chip contacts. This part of the process is serial in nature, but can benefit heavily from automation (pick-and-place and wire-bonding machines) as it is a simple task, making it surprisingly cost effective.

![Figure 5.6: Die mounting and wiring on a lead frame.](image)

The encapsulation steps is, in the other hand, done by batch, using directly the lead frame strip as shown in Figure 5.7. The lead frames with their wired dies are placed in a mold having multiple cavities before injection molding can be used. A thermosetting monomer ‘puck’ is heated above its glass transition temperature and pushed by a piston in all the cavities of the mold, completely covering the dies. Then heating is maintained until the thermosetting monomer is fully cross-linked and becomes a hard polymer. The polymer normally used is based on epoxy, although it has relatively poor thermal properties and needs to incorporate different additives (flame retardant for example) to meet environmental regulations.

### 5.2.2 Hermetic encapsulation

The protection function of the package is normally understood as relatively simple to ensure if one can make the package fully hermetic: in that case no contaminant (gas, moisture, dust...) can contact the die and it will be protected. However we know that even the most airtight metal box of cookies eventually got them spoiled, thus what does ‘hermetic’ really mean?

Actually fully hermetic package, that is a package without any exchange with the environment for any period of time, hardly exists at all. In practice, given enough time, some gases will creep through some defects by diffusion like process and reach the inside of the package. Of course the existence of leaks, as can be found at seal interface, makes the thing worse, but plain materials without cracks will anyhow let fluids sip in. Then the choice of
Figure 5.7: Main steps in overmolding process using thermosetting polymer (epoxy).

The material is crucial to obtain good hermetic package, as the permeation of gas and moisture through the material itself will be the ultimate limit to the leak rate in any package.

Actually, the flow of gas \( \dot{Q} \) through a barrier made of a certain material can be linearized and using the unit of mole per unit of time (\( \dot{Q}_M = \frac{\partial N_M}{\partial t} \)) given the form:

\[
\dot{Q}_M = P_0 \frac{A \Delta P}{d}
\]

where \( P_0 \) is the intrinsic permeability for the material, \( A \) the exposed surface, \( \Delta P \) the pressure difference, and \( d \) the barrier thickness. There is no standard unit for \( P_0 \) and it mostly changes with the pressure unit used.\(^1\) We express flow of matter in mol/s but the literature often reports a mass flow in kg/s instead. Divide the kg/s value by the gas molar mass (e.g. 0.018 kg/mol for water) to obtain mol/s.

\(^1\)We express flow of matter in mol/s but the literature often reports a mass flow in kg/s instead. Divide the kg/s value by the gas molar mass (e.g. 0.018 kg/mol for water) to obtain mol/s.
mol s\(^{-1}\) atm\(^{-1}\) m\(^{-1}\). In Figure 5.8 we show the permeability of a barrier for some class of materials typically used in packaging.

Figure 5.8: Typical intrinsic water permeability of a few packaging materials with the estimated corresponding time for the humidity inside the package to reach 50% of exterior humidity.

As a matter of fact, for electronic circuit high reliability is obtained even with non-hermetic packaging (polymer) by coating the surface of the wafer with protecting layer (e.g., low stress silicon nitride and other dielectrics) that acts as excellent barrier to moisture and gas. We can see in Figure 5.8 that 2 µm of such materials is theoretically equivalent to almost 10 cm of epoxy! Of course the reality will be different as thin films tend to have more defects, but nevertheless the figure allows to highlight a big difference with MEMS packaging: in general mobile part will prevent using protection layers deposited on the die and reliability will be harder to guarantee as it will more directly depend on the hermeticity of the package itself.

To assess and model this behavior we will develop a model of leak where we describe the flow of gas with classical theory using SI units of mol/s. We note that the literature for leak or flow often use a unit of pressure × volume (at a certain temperature) instead of mol for the quantity of matter. This unit
Table 5.3: Conversion of flow units between customary units and mol/s for different temperatures.

<table>
<thead>
<tr>
<th>Unit</th>
<th>0°C mol/s</th>
<th>20°C mol/s</th>
<th>25°C mol/s</th>
<th>300K mol/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 atm cc/sec</td>
<td>4.46 \times 10^{-5}</td>
<td>4.16 \times 10^{-5}</td>
<td>4.09 \times 10^{-5}</td>
<td>4.06 \times 10^{-5}</td>
</tr>
<tr>
<td>1 Pa m³/s</td>
<td>4.40 \times 10^{-4}</td>
<td>4.10 \times 10^{-4}</td>
<td>4.03 \times 10^{-4}</td>
<td>4.01 \times 10^{-4}</td>
</tr>
<tr>
<td>1 mbar l/s</td>
<td>4.40 \times 10^{-5}</td>
<td>4.10 \times 10^{-5}</td>
<td>4.03 \times 10^{-5}</td>
<td>4.01 \times 10^{-5}</td>
</tr>
</tbody>
</table>

is ambiguous (in addition to not being a SI unit) as the quantity of matter represented by the pressure \times volume units depends on the experimental temperature which is often not reported, making precise comparison between different systems impossible. If the temperature is known, using the perfect gas relationship, we can convert the customary units to mol/s as shown in Table 5.3.

The leaking behavior of packages can be described relatively simply by considering leaks as very narrow channels. Actually when there is a difference of pressure between two sides of a barrier maintained at the same temperature, it means that there is a difference of density in gas molecules on both sides. The relationship between gas molecule molar density and pressure is simply given by using the perfect gas’s law:

\[ PV = N_M RT \Rightarrow P = \frac{N_M}{V} RT \Rightarrow P = n_M RT \]

where \( N_M \) is the number of mole of gas molecule, \( n_M \) the molar density (in mol/m³) and \( R = 8.31 J/K \).

But what happens now if there is an aperture in the barrier? In that case, in the average, the number of molecule entering the aperture on the high density (i.e. pressure) side will be larger than on the lower density (i.e. pressure) side, resulting in a net flow of molecule from high pressure to low pressure. This flow is then simply governed by the gas molar density difference \( n_{M2} - n_{M1} \), and can be written in the form:

\[ \dot{Q}_{M1} = \frac{\partial N_{M1}}{\partial t} = C(n_{M2} - n_{M1}) \]

where \( C \) is expressed in m³/s and is the conductance of the channel that takes into account the probability that a molecule is going through the aperture and not coming back.
The conductance depends on the flow regime, which in turn depends on the dominant collision mode for the gas molecule flowing through the aperture: are the wall collisions or the inter-molecule collisions dominating? To estimate this effect we compare the dimension of the channel $d$ with the gas molecules mean free path $\lambda$. At one end of the regime ($\lambda/d < 0.01$ : large channel), as the wall contact is infrequent and the inter-molecules collisions dominant we have viscous Poiseuille’s flow ( cf. Sec. 4.2.2). Then, as the interaction with the wall becomes dominant ($\lambda/d > 1$ : narrow channel) we observe molecular and diffusion flow. The equations governing these different type of flow, shows that for a long channel of diameter $d$, Poiseuille’s flow varies as $d^4$ (cf. Eq. 4.2), molecular flow as $d^3$ and diffusion flow as $d^2$, which could allow to experimentally find what regime is dominant in a particular case. But, for small hermetic package, only the study of fine leaks is of interest (larger leak will change pressure inside the package too rapidly) and we can reasonably suppose that molecular flow regime is the dominant flow regime.

The channel conductance then takes a simple form as shown by Knudsen and the equation governing the flow becomes:

$$\dot{Q}_{M1} = F_m \sqrt{\frac{T}{M}} (n_{M2} - n_{M1})$$

where $\dot{Q}_{M1}$ is the molar flow in region 1 counted positive if it enters the region, $F_m$ is the molecular conductance of the leak (which for a single ideal circular channel of diameter $d$ and length $L$ is given by $F_m = \sqrt{R\pi/18d^4/L}$), $T$ the absolute temperature, $M$ the molecular mass of the gas and $n_{Mi}$ the molar density in region $i$. The conductance of the conduit (also called the standard or true leak rate), for molecular flow is then defined as:

$$C = F_m \sqrt{\frac{T}{M}}$$

We note that when the conductance for one gas is known, it can be obtained for any other gases, provided its molecular mass is known. Helium gas having the smallest molecular mass it will leak the fastest and, at the same temperature, all other gas will leak slower by a factor given by $C_2 = C_1 \sqrt{M_1/M_2}$. Table 5.4 gives this ratio for some commonly encountered gases. The air value is for rough computation as the molar density (or partial pressure) of each component of air should be used with their corresponding leak rate to

---

2Diffusion flow will be dominant over molecular flow in the case where $L \ll d$, that is a short conduit between the two regions which is not a common case in encapsulation.

3We use here molar density instead of pressure as it makes the theory more sound and removes some ambiguities in existing derivation.
estimate the effect of the leak (78% N\textsubscript{2}, 31% O\textsubscript{2}...). Actually, because of the difference in gas conductance, the leaked ‘air’ will have a different composition than normal air (N\textsubscript{2} will leak faster than O\textsubscript{2}) until the molar density equilibrates. The molar leak rate $\dot{Q}_M$ is easily computed if the molar density difference is constant, but actually for a closed package a gas leaking inside will gradually increase the molar density (or lower if it leaks outside) continuously changing the leak rate, until the molar density on both side becomes equal.

If we consider the standard situation for a package where the molar density (and pressure) of the gas in the environment is unaffected by what can leak from the package (i.e., $n_{M2} = n_{M20}$ is constant), and by using the volume $V$ of the package to relate the mole number to molar density $n_{M1} = N_{M1}/V$, we can solve the flow equation Eq. [5.2.2] and obtain the evolution with time of the difference in molar density inside and outside the package:

$$n_{M20} - n_{M1} = (n_{M20} - n_{M10})e^{-\frac{C}{V}t}$$

Then the molar leak rate is obtained as:

$$\dot{Q}_M = C(n_{M20} - n_{M10})e^{-\frac{C}{V}t}$$

The molar density evolution equation can be written directly using instead the partial pressure in the package using Eq. [5.2.2] as:

$$p_1 = p_{20} + (p_{10} - p_{20})e^{-\frac{C}{V}t}$$

This equation is the one generally obtained by using the standard derivation of the molecular flow theory, however it is only valid if the temperature...
inside and outside the package is the same (which is not necessarily the case as heat is usually generated inside the package) and shows the interest of our approach.

Typically, the equation is used with two different initial conditions:

- the package is in vacuum at $t = 0$ and air slowly leak inside. Then we have:
  \[ p_{\text{pack}} = p_{\text{air}}(1 - e^{-CVt}) \]

- the package is pressurized with air at $p_0$ at $t = 0$ and leaks outside. We have:
  \[ p_{\text{pack}} = p_{\text{air}} + (p_0 - p_{\text{air}})e^{-CVt} \]

Of course, other situation will involve more complex results, as for example if He is used to pressurize the chip: as the He will escape through the leak, air will try to enter the package (the partial pressure of nitrogen or oxygen is 0 inside) - but will be impeded in its inward flow by the out-flowing He.

To maintain tight hermeticity the best method is probably to use wafer bonding technologies with limited permeability to gas, like glass to silicon anodic bonding or metal to silicon eutectic bonding (cf. 3.3.3). However all MEMS can not be treated in this way, and for example the Texas Instrument DLP’s packaging is more complex because the tiny mirror would not survive harsh elevated temperature treatment including glass bonding. Thus, a full chip-by-chip hermetic package in metal with a transparent glass window had to be designed. The package is sealed using brazed metal can in a clean room under a dry nitrogen atmosphere with some helium to help check leaks, and incorporates strip of getter material, a special material for removing the last trace of humidity (Fig. 5.9). The getter is a material with high porosity that will react with the target gas (usually water vapour, but oxygen or other gases can also be targeted) forming a solid compound at the getter surface. In general a special trick is used to activate the getter after
the package is closed (heating it above a certain temperature, for example) so that the getter does not get quickly saturated in the open air during packaging operation. Getter can be deposited by PVD or CVD in thin films or pasted in the package, but they have a finite size and thus will work best for a finite amount of gas molecules: in general the trace of gas adsorbed on the inner surface of the package. On the longer run, they will retard the degradation due to the permeation of water vapour but even very small leaks will saturate the getter rapidly.

In view of the complexity - and cost - of fully hermetic packaging, it is lucky that all MEMS do not need such packages and the end of the 1990s Ken Gileo and others introduced a new concept: near-hermetic packaging. If this concept resists a formal definition (what leak rate defines quasi-hermeticity?) a heuristic definition would say that the package should be “good enough” for the MEMS operation. Accordingly, relaxing the constraint on the hermeticity (possibly by supplementing it with a getter) opens up the range of techniques that can be used, and for example polymer encapsulation or bonded wafer using solder bonding or even polymer bonding, would be generally good enough while allowing a much simpler bonding procedure (as, for example, the flatness requirement with a relatively thick solder paste is heavily relaxed compared to what is needed for anodic, or even worse, for fusion bonding).

Finally, for hermetic or quasi-hermetic package the remaining question will be how to test the hermeticity? Gross leak can easily be detected by a simple bubble test where the package is heated and immersed in a liquid: if there is a gross leak the heated gas will escape through the leak and form bubble. But this test does rarely make sense in MEMS packaging, where gross leak could be spotted during visual inspection. For fine leak, the standard test is the He-leak test using what is known as the “bombig test”. Here, the package is first placed into a high pressure chamber with Helium for some time to force the gas into the package. Then the package is taken out of the high pressure chamber and the He leak rate is measured with a calibrated mass-spectrometer. This procedure is able to measure fine leak rate in package with a volume of a fraction of a cm$^3$ down to a leak rate of about $5 \cdot 10^{-17}$ mol/s (or $10^{-12}$ mbar l/s). Still, in general MEMS package are too small (and the amount of He too little) to use directly the test on real packages and special test bed (using the same material and bonding technique but having a larger cavity) need to be built to perform this test. More advanced techniques will use the measurement of the Q factor of a mechanical resonator microfabricated on the chip itself. As the pressure inside the package increases (when it is left at normal pressure or at higher pressure for accelerated tests), the Q factor of the resonator decreases, allowing to estimate the leak rate. The advantage of this technique is that it is sensitive
enough to measure the leak directly on the MEMS packages with their actual dimension.

### 5.2.3 Electrical feedthrough

The main technique used for connecting die to the contact on the lead frame (cf. Figure 5.6), or to other die in case of in-package assembly, is wire-bonding. Originally developed as thermocompression gold to gold bonding (still used for wafer-to-wafer bonding cf. Sec. 3.3.3), it evolved to take benefit from ultrasonic force. We can distinguish now 3 different types of techniques (Table 5.5, with the dominant one in IC manufacturing being thermosonic bonding, while ultrasonic bonding is often used for MEMS because of its low process temperature, although the high ultrasonic energy may pose problem to mobile mechanical part.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Pressure</th>
<th>Temp.</th>
<th>US</th>
<th>Mat.</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermocompression</td>
<td>High</td>
<td>300-500°C</td>
<td>No</td>
<td>Au/Au/Au/Au</td>
<td>B-W</td>
</tr>
<tr>
<td>Ultrasonic</td>
<td>Low</td>
<td>25°C</td>
<td>Yes</td>
<td>Au/Au/Au/Au</td>
<td>W-W</td>
</tr>
<tr>
<td>Thermosonic</td>
<td>Low</td>
<td>100-150°C</td>
<td>Yes</td>
<td>Au/Au/Au/Au</td>
<td>B-W</td>
</tr>
</tbody>
</table>

Table 5.5: Comparison of wire-bonding techniques

The type of the bond (ball or wedge) depends on the tool used, as shown in Figure 5.10. Wedge bonding would usually allow higher contact density (pitch < 50µm), but is slightly slower than ball bonding.

The most common material used for the bond are gold and aluminum. The wire is 50-75 µm diameter while the pad minimum dimension should be about 4 times the wire diameter for ball bonding, or 2 times for wedge bond. The Al/Au combination may form intermetallic and develop Kirkendall voids at moderate temperature that are detrimental to the reliability of the bond. Accordingly, Al/Au or Au/Au combinations should be preferred as they are free from these problems. Other materials could be used as well in special cases. Copper wires are attractive because of their low price and excellent conductivity, but they are harder to bond requiring more force and inert atmosphere as they oxidize readily. The Au/Ag combination has shown
excellent high temperature reliability and has been used for years with Ag plated lead-frame.

Wirebonding is harder to apply if the MEMS has been capped for protection or for hermeticity. Actually, by covering the surface of the MEMS wafer, the cap wafer hides the pads that would have normally been used for the interconnect. Accordingly, many techniques have been developed to get the electrical contact fed under the cap. In general the techniques are split in three different classes, with typical examples in each cases shown in Figure 5.11:

Through the cap in that case the contact are taken through the cap. The advantage of this technique is that it does not interfere with the MEMS process, as the bulk of the additional processes needed is performed on the cap itself.

Through the wafer in that case the contact are fed through the MEMS wafer to its back-surface. This require additional process on the MEMS wafer, but it can be interesting to do if there are already through-the-wafer holes in the main MEMS process.

Lateral here the contact run below the surface of the MEMS wafer at the edge of the cap. The difficulty here is to have a conductor running on the surface that won’t compromises the hermeticity of the bond. A solution proposed by Sensonor for their MPW process is to use boron doped conductors diffused at the surface of a SOI wafer before bonding.

The wafer bonding approach, has the potential to be used for a true wafer level packaging technique (WLP) - without the need for overmolding - that could allow packaging and of all the chips on the wafer in a single batch.
Through the cap
Through the wafer
Lateral

Figure 5.11: Example of electrical feedthrough techniques for capped wafers.

operation. Actually the only remaining feature that is needed after wafer bonding is the possibility to connect the chip to a printed circuit board (PCB) so that it can be assembled in a complex system. This feature can be obtained by using the ball bonding technology seen in flip-chip integration that we have seen in Figure 4.2. The solder ball can be used to interconnect the IC and the MEMS chip, but also the resulting stack of chips to the PCB, a step forward in 3D packaging technique. The advantage of solder ball bonding is that it is performed in batch, as the balls will solder the two wafers after they have been simply aligned in contact with a pad and heated in an oven. Besides, as we can see in Figure 5.12, the deposition of solder balls, also called bumping, can be performed at wafer level in batch. The combination of ball-bonding and wafer bonding (including the inclusion of via to bring contact from one side of the wafer to the other) may result in packaging that is the same size as the die, and we then speak of chip-size packaging (CSP), the ultimate goal for extreme system miniaturization.

## 5.3 Testing and calibration

Testing is required to increase the reliability of the packaged MEMS. Different type of tests are performed during the complete process, and we distinguish qualification tests to detect failed dies while burn-in tests screen out low reliability dies. Burn-in tests are performed at chip level often after packaging, while qualification tests are being performed both at wafer and chip level to screen the chips that should not be packaged.

In addition to qualification test performed at wafer level, the testing phase allows to perform calibration, which will be normally conducted after packaging. The purpose of calibration is to compensate some of the defects into the MEMS characteristic to make it work more ideally. Actually the linearity of MEMS sensor or actuator may not be perfect or, more often, the cross-sensitivity with some environmental parameters (usually temperature) would require to be accounted for, another operation called compensation.

The calibration and compensation will help to decrease the influence of
unavoidable defects of the micro-sensor but, most of the time, the calibration and the compensation will be decided at the factory. In the future it is expected that all the system will include a capability to perform self-calibration. If this is not always possible to implement (for example, applying a reference pressure for a pressure sensor, is not an easy task), some system have so much drift that they won’t operate in any other manner. A good example is provided by some chemical micro-sensors that can only compare concentration in two fluids (thus need a calibration for each measurement) but can hardly give ‘absolute’ readings.

5.3.1 Testing

However if the overall reliability is mostly considered to be governed by the fabrication process, the reality is different. Actually all the process steps added during packaging and test may affect adversely the final reliability of
A good insight at the importance of the calibration and compensation may be gained by comparing two products from the range of pressure sensors from Motorola. The MPX10 is an uncalibrated and uncompensated pressure sensor, while the MPX2010 is passively calibrated and compensated. We report in the following table some of their characteristics, extracted from the manufacturer’s datasheets (MPX10/D and MPX2010/D).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>MPX10</th>
<th></th>
<th></th>
<th>MPX2010</th>
<th></th>
<th></th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scale Span</td>
<td>20</td>
<td>35</td>
<td>50</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>mV</td>
</tr>
<tr>
<td>Offset</td>
<td>0</td>
<td>20</td>
<td>35</td>
<td>-1.0</td>
<td>-</td>
<td>1.0</td>
<td>mV</td>
</tr>
<tr>
<td>Temp. eff. full scale span</td>
<td>-18</td>
<td>-</td>
<td>-13</td>
<td>-1.0</td>
<td>-</td>
<td>1.0</td>
<td>%FS</td>
</tr>
<tr>
<td>Temp. eff. offset</td>
<td>–</td>
<td>± 1.3</td>
<td>–</td>
<td>-1.0</td>
<td>-</td>
<td>1.0</td>
<td>mV</td>
</tr>
<tr>
<td>Temp. coeff. full scale span</td>
<td>-0.22</td>
<td>-</td>
<td>-0.16</td>
<td>-1.0</td>
<td>-</td>
<td>1.0</td>
<td>%/°C</td>
</tr>
<tr>
<td>Temp. coeff. offset</td>
<td>–</td>
<td>± 15</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.5</td>
<td>µV/°C</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>–</td>
<td>3.5</td>
<td>–</td>
<td>2.5</td>
<td>–</td>
<td>–</td>
<td>mV/kPa</td>
</tr>
</tbody>
</table>

We see the effect of calibration of the sensor using laser trimmed resistor on the full scale span, that is properly normalized, and on the offset that is almost suppressed. The compensation technique use additional resistor that are also laser trimmed to reduce tremendously the influence of temperature on the full scale span. However, we could note that the effect of temperature on the offset is not really compensated this way.

What is not shown here is the large difference in price between this two sensors, which could be an important element of choice!

**Example 5.3:** Using calibrated and compensated sensor or not?
the device in sometimes unsuspected ways. For example, it has been shown that the qualification tests performed at wafer level may affect reliability of wire bonding. Actually, these tests are performed in a probe station using sharp needles to contact the pads and apply different test signals to the device. The contact of the test probe on the gold pads would leave a small scratch on the pad surface, which has been shown to affect the bonding strength, ultimately possibly decreasing the reliability of the packaged system.

The main problem faced by MEMS testing is that we now have to handle signal that are not purely electrical, but optical, fluidic, mechanical, chemical... Then, verifying the absence of defect needs the development of specialized system and new strategies.

Texas Instrument DLP chip may have as many a 2 millions mirrors and simple math shows that testing them one by one during 1 s would take approximately three weeks at 24 h/day – clearly not a manageable solution. TI has thus developed a series of test using specific mirror activation patterns that allow testing mirrors by group and still detect defect for individual mirror, like sticking or missing mirror. After testing at wafer level the chip is diced, put into packages and then goes through a burn-in procedure. They are then tested again before being finally approved. TI noticed that the encapsulation step decreased the yield if the environment wasn’t clean enough and they have to use a class 10 clean-room for the packaging of their DLP chips.

Testing is also a major hurdle for micro-sensor and to facilitate it additional test features may have to be included into the MEMS design. A good example is given by the integrated accelerometer range from Analog Devices. The system use a micromachined suspended mass whose displacement is monitored by using induced change in capacitance. However one part of this electrodes has been configured as an actuator. By applying a voltage on this electrodes it is possible to induce a movement of the mass without external acceleration. This is used before the packaging to verify the mechanical integrity of the accelerometer... and allow to save a lot of money, compared to a set-up that would need to apply a real acceleration. Moreover that function may be used during operation in a smart system to verify the integrity of the accelerometer.

Testing is conducted at different stage during the fabrication of the MEMS, but the final test are normally conducted after the packaging is done. The reason is simple: the packaging process always introduce stress (or damping) that change the characteristics of the sensing elements, but need to be accounted for. This final test can be used for burn-in and usually allow the final calibration and compensation of the sensor.

114
5.3.2 Calibration

The calibration of the MEMS is the adjustment needed to deliver the most linear possible transfer function, where the output goes from 0% to 100% when the input varies in the same range. It generally means two different things:

1. linearization of the transfer function (i.e., having a constant sensitivity)

2. suppression of the offset

It is a particularly important step for microsensors, although it is understood that all the other environmental variables (e.g., temperature, humidity, pressure, stress...) are kept constant during the calibration and will be specified. As such, the response of the system will generally be the best at calibration point. We note that the cancellation of the change caused by these other environment variables on the sensitivity or on the offset is left to the compensation procedure. Thus it is possible to have a calibrated but uncompensated system, while the reverse is often much less interesting.

To perform these tasks it is possible to trim integrated additional analog circuit or to use a digital signal and a CPU. The choice between both is a mix of speed, complexity and cost analysis, but obviously the integrated analog approach needs more wit! Obviously it is possible to mix the two methods, that are not mutually exclusive, and for example offset removing is often performed analogically while sensitivity trimming is more easy to do with digital techniques. In both case the calibration is performed after having the result of the calibration test that will tell to which extent the sensor needs to be adjusted.

With analog calibration technique, the basic tuning elements are trimmable resistors. The technique use a laser or electric fuse to trim the value of resistor controlling MEMS sensitivity and offset. This method has the advantage to be relatively cheap and to provide sensor with very high speed. Actually the laser trimming method is less useful as it does not allow recalibration nor allow the calibration after packaging, and electrical trimming is preferred. The trimmable resistor does not allow to compensate for complex non-linearity, and if previously there was a lot of effort devised in developing linearizing scheme with complex analog circuits, now the trend is toward digital calibration for the more complex cases.

Digital calibration technique is quite straightforward to implement if sufficient computation power is available, but will always be slower because it needs analog to digital conversion and data processing. The principle is to perform a calibration test, and to use the data recorded to compute the
output of the sensor. The CPU may use a model of the sensor, usually implemented using a high order polynomial, or use the actual values of the calibration test, stored in a look-up table (LUT). Between to point of the look-up table, the correction is estimated by using a linear approximation. In principle this approach allows to compensate for any non-linearity, and should deliver ‘perfect’ characteristics. However apart from the speed problem, the analog to digital conversion of the signal introduce new parameters that does not allow to correct all the defect in the transfer function characteristic. For example, we show in Figure 5.13 the case of a MEMS sensing element presenting a flatter part in his transfer function. The marked non-

![Figure 5.13: Effect of sensing element characteristic on accuracy after digital conversion](image)

linearity of this transfer function introduce measurement errors than can not be calibrated out. Actually, any input laying in the range between about 5 and 12.5, will give the same A/D converter output: [0110]! We have here an important loss in the accuracy of the sensor that can not be suppressed. Thus if digital calibration allows to correct many defects, it is not possible to eliminate all of them. From this example, we may remember that the maximum error presented by non-linear element will be governed by the region of the transfer function presenting the smallest slope.

Additionally, the digital converter works best by using its full range (otherwise, we ‘loose’ some bits of resolution)... which is not necessarily the same range as the MEMS operation range. In general, it is thus necessary to have a controllable gain amplifier that will adapt the converter signal to the
MEMS range. Additionally it would be interesting for the same reason to have a controllable way to trim an offset voltage. These two functions can be merged together in an amplifier with tunable gain and offset. As an example, Microsensors Inc. is producing an integrated circuit for sensor interfacing, the MS3110, that has an output amplifier with a voltage gain that may be chosen between 1.7 and 2.3 using 8 bits of control for a step of 0.0024 and a trimmable offset that change in a range of ±100 mV using 5 bits of control with 6.25 mV step. Actually, this component will also be very useful for the compensation of the sensor, as we will see now.

5.3.3 Compensation

The compensation refers to the techniques used to separate the MEMS output from an interfering environmental parameter, like the temperature, and we distinguish:

- **structural compensation** where preventive measure are taken at the design stage to decrease the magnitude of cross-sensitivities;

- **monitored compensation** where implicit or explicit measurement of the interfering parameter allow to modify the output of the sensor to compensate for its effect. Implicit measurement approach use additional integrated circuitry with the sensor, while explicit measurement use a sensor for the interfering parameter (e.g., a temperature sensor) and a CPU to adjust the raw measurements according to the value of the interfering parameter.

In the first class of compensation, the layout of the microsystem is important and for example it is possible to insulate it thermally to decrease the influence of temperature. Another very simple structural compensation technique that should not be underestimated is the use of symmetry in the design. The principle is to use a difference signal, while all the other interfering variables will produce a common mode signal, that will thus not be apparent on the MEMS output. The compensation for residual stress in many MEMS sensors is often based on this principle, and observing the design of the sensitive elements will invariably show a marked 2-folds or 4-folds symmetry.

The monitored compensation may be performed at the system level without explicit measurement of the perturbing parameter using completely analog signal, and we talk of implicit compensation or with explicit compensation. In this latter case the compensation is using a CPU and a programmable gain amplifier and programmable offset. The choice between these two approaches is often dictated by the complexity of implementation. An implicit
compensation is often used to compensate for temperature in pressure sensor, but as soon as the dependence on the external factor is complex, or when a very precise compensation is needed the compensation is performed at the system level using a CPU. Generally speaking, an implicit compensation needs more cleverness than an explicit approach that will work ‘all the time’. For example, if it is relatively easy to compensate for an offset induced by the temperature with analog circuit, a change in the gain of the transfer function of the sensor will be much more difficult to compensate and explicit compensation with analog or digital techniques will be required. Still implicit approach will produce smaller control circuit that could be a decisive advantage.

For example, implicit monitored compensation can be applied to shelter from fluctuation in power voltage. In this case the idea is simply to use working principles that are ratiometric to this voltage. For example, a Wheatstone bridge (see Sec. 4.3.1) delivers a voltage that is proportional to the supply voltage $V_{in}$:

$$V_{out} \approx \frac{V_{in}}{4R} \Delta R.$$  

This may seem to be a serious problem as any fluctuation in the supply voltage will result in loss of accuracy, but when we consider the complete measurement chain, this may turn to an advantage. If we consider a pressure sensor based on piezoresistive elements in a Wheatstone bridge, the signal from the sensor needs digitization further down in the measurement chain to be transmitted or recorded. The digitization based on A/D converter requires generally a reference voltage from which the quantization step (quantum) is derived. If for this reference voltage we use the bridge supply voltage $V_{in}$, then any fluctuation in this voltage is automatically compensated by an equivalent change in the A/D converter quantum. In this way the recorded digital information of the pressure will be accurate even if the supply voltage change due to battery charge dwindling or other environmental causes.

A typical example of a circuit using explicit compensation (also called digital compensation) is shown in Figure 5.14. Here the CPU constantly change the output according to the value of the input, using a model or a calibration curve. However if this implementation looks simple it has some marked drawbacks, when, for example, the sensitivity of the system decrease with the interfering parameter as shown for a sensor in figure 5.15. If the sensitivity increase ($T_1 > T$) in this simple scheme, a loss of resolution occurs because the ADC has a limited span and thus saturates. If the sensitivity decreases ($T_2 < T$), and the noise at the output of the sensing element remains constant, the resolution of the sensor drops tremendously. Thus this ‘all
digital’ approach is not completely satisfactory and a mixed digital-analog approach is required. It generally uses a tunable amplifier or voltage source to compensate the output of the sensing element, performing also the signal standardization to match the ADC. A typical implementation of this strategy will using digital to analog converter to control the supply voltage of the sensing element, controlling effectively its sensitivity, and use a gain tunable amplifier (GTA). The additional circuitry used is based on a digital to analog converter (DAC), converting the digital signal from the CPU to an analog signal that is used directly (voltage offset control) or use an additional compensation actuator to convert it to a usable signal (voltage/gain conversion, voltage/current conversion, voltage/force conversion as in the ADXL105). Additionally in such circuit using explicit compensation, in order to reduce
circuitry, the temperature sensor and the sensing element are sharing the same A/D converter using a multiplexer.

The digital signal processor (DSP) will perform the necessary computation to compute the value of the different adjustable parameters, either using a physical or empirical model of the dependence of the sensitivity and offset with the temperature, or, better, using the result from a calibration test. In this latter case, as it is not possible to perform a calibration for every temperature, an interpolation will be used to determine the change between two measured points. As in the case of the calibration the correction factor will be stored as a series of values, but in this case in a two dimensional array, the first dimension corresponding to the measurand and the second, the temperature.

The LUT approach is interesting when only one interfering parameter needs to be compensated, but when there are multiple interfering parameters, as in the case of chemical sensors, the approach becomes quickly intractable. Actually, as it is not possible to perform all the calibration tests needed, the compensation can only be done using a model. This model will be implemented in the system using a series of coefficient of a polynomial (normally with order < 7), limiting the necessary amount of memory at the cost of a larger computation effort.
Chapter 6

Challenges, trends, and conclusions

6.1 MEMS current challenges

Although some products like pressure sensors have been produced for 30 years, MEMS industry in many aspects is still a young industry. The heavily segmented market is probably the main reason why a consortium like SEMI is still to appear for MEMS. However everybody agrees that better cooperation and planning has to happen if the cost of the assembly, test and packaging is to come down. MEMS can currently only look with envy as IC industry seriously considers producing RFID chips for cents - including packaging. Again the path shown by the IC industry can serve as a model, and standardization to insure packaging compatibility between different MEMS chip manufacturers seems the way to go. Considering the smaller market size of most MEMS component, standard is the only way to bring the numbers where unit packaging price is reduced substantially. This implies of course automating assembly by defining standard chip handling procedure, and probably standard testing procedure.

Of course, the diversity of MEMS market makes it impracticable to develop a one-fit-all packaging solution and the division in a few classes (inertia, gas, fluidic) is to be expected. For example, several proposals for a generic solution to fluidic interfacing have been proposed and could become a recommendation in the future.

In the other hand it is not clear if standardization of MEMS fabrication process à la CMOS will ever happen - and is even possible. But currently most of the cost for MEMS component happens during back-end process, thus it is by standardizing interfaces that most savings can be expected.
The relatively long development cycle for a MEMS component is also a hurdle that needs to be lowered if we want more company to embrace the technology.

One answer lies with the MEMS designing tool providers. The possibility to do software verification up to the component level would certainly be a breakthrough that is now only possible for a limited set of cases.

But it is also true that the answer to proper design is not solely in the hand of better computer software but also in better training of the design engineer. In particular we hope that this short introduction has shown that specific training is needed for MEMS engineers, where knowledge of mechanical and material engineering supplements electronic engineering. Actually, experience has often revealed that an electronic engineer with no understanding of physical aspect of MEMS is a mean MEMS designer.

### 6.2 Future trend of MEMS

Looking in the crystal ball for MEMS market has shown to be a deceptive work, but current emerging tendencies may help foresee what will happen in the medium term.

From the manufacturer point of view, a quest for lowering manufacturing cost will hopefully result in standardization of the MEMS interfacing as we discussed earlier, but finally will lead to pursue less expensive micro-fabrication method than photolithography. Different flavors of soft-lithography are solid contenders here and micro-fluidic and BioMEMS are already starting to experience this change. Another possibility for reducing cost will be integration with electronics - but, as we already discussed, the system-on-a-chip approach may not be optimal in many cases. Still, one likely good candidate for integration will be the fabrication of a single-chip wireless communication system, using MEMS switch and surface high-Q component.

From the market side, MEMS will undoubtedly invade more and more consumer products. The recent use of accelerometer in cameras, handphone or in the Segway is a clear demonstration of the larger applicability of the MEMS solutions - and as the prices drop, this trend should increase in the future. Of course medical application can be expected to be a major driver too, but here the stringent requirements make the progress slow. In the mid-term, before micromachines can wade in the human body to repair or measure, biomedical sensors to be used by doctors or, more interesting, by patients are expected to become an important market.

A farthest opportunity for MEMS lies probably in nanotechnology. Actually, nanotechnology is bringing a lot of hope - and some hype - but current fab-
Fabrication techniques are definitely not ready for production. MEMS will play a role by interfacing nano-scale with meso-scale systems, and by providing tools to produce nano-patterns at an affordable price.

6.3 Conclusion

The MEMS industry thought it had found the killer application when at the turn of the millennium 10’s of startups rushed to join the fiber telecommunication bandwagon. Alas, the burst of the telecommunication bubble has reminded people that in business it is not enough to have a product to be successful - you need customers. Now the industry has set more modest goals, and if the pace of development is no more exponential it remains solid at 2 digits, with MEMS constantly invading more and more markets. Although the MEMS business with an intrinsically segmented structure will most probably never see the emergence of an Intel we can be sure that the future for MEMS is bright. At least because, as R. Feynman [27] stated boldly in his famous 1959 talk, which inspired some of the MEMS pioneers, because, indeed, ”There’s plenty of room at the bottom”!
Chapter 7

Readings and References

7.1 Online resources and journals

http://www.smalltimes.com/ the free international press organ of the MEMS–NEMS community.

http://www.mstnews.de/ free news journal on European microsystem technology

http://www.dbanks.demon.co.uk/ueng/ D. Banks renowned ”Introduction to microengineering” website with plenty of information.

http://www.aero.org/publications/aeropress/Helvajian/ The first chapter of ‘Microengineering Aerospace Systems’ co-authored by M. Mehregany and S. Roy and edited by H. Helvajian is online and makes a short, although slightly outdated, introduction to MEMS.

IEEE/ASME Journal of MEMS This journal originally edited by W. Trimmer, is arguably one of the best journal in the field of MEMS (http://www.ieee.org/organizations/pubs/transactions/jms.htm).

Sensors and Actuators A That is the most cited journal in the field, with a copious variety of research work (http://www.elsevier.com/wps/product/cws_home/504103).

Sensors and Actuators B Catering mostly for Chemical Sensor papers, they also have issue on MicroTAS, where you find numerous microfluidics and Bio-MEMS papers (http://www.elsevier.com/wps/product/cws_home/504104).

Smart Materials and Structures  Another IOP journal, with editor V. Varadan, that has a more material oriented approach than his cousin (http://www.iop.org/EJ/S/3/176/journal/0964-1726).

Microsystem Technologies  A Springer journal that favors papers on fabrication technology and particularly on high-aspect ratio (LIGA like) technology (http://link.springer.de/link/service/journals/00542/index.htm).


Biosensors and Bioelectronics  Another Bio-MEMS journal with more emphasis on sensors (http://www.elsevier.com/wps/product/cws_home/405913).

IEEE Transaction on Biomedical engineering  Bio-MEMS and biomedical application can be found here (http://www.ieee.org/organizations/pubs/transactions/tbe.htm).

IEEE Photonics Technology Letters  Highly cited photonics journal publishing short papers, including optical MEMS or MOEMS (http://www.ieee.org/organizations/pubs/transactions/ptl.htm).


7.2 Other MEMS ressources

http://www.memsnet.org/ the MEMS and nanotechnology clearinghouse.
the MEMS industry group aimed at becoming a unifying resource for the MEMS industry. Will hopefully initiate standardization effort and eventually establish a MEMS roadmap.

one of the MEMS industry watch group publishing regular report on the market.
Bibliography


Index

actuation
  electromagnetic, 81
  electrostatic, 83
  thermal, 86
actuator, see MEMS actuator
Agilent, 8
Alcatel-Adixen, 14, 61
amorphous, 42, 55
Analog Devices, 7, 19, 68, 80, 114
anisotropic etching, 11, 31
annealing, 46
ANSYS, 22
aspect ratio, 37
Assembly, 92
AZ9260, 55
beam, 69
bi-material actuator, 80
bimetallic actuator, see bi-material actuator
boron doping, 34
Bosch, 18, 56, 60
  process, 37, 61
bulk micromachining, 29
bumping, 111
calibration, 111
capacitive sensing, 79
capillarity, 76
case
  CERDIP, 98
  TO, 98
chemical vapor deposition, 53
  APCVD, 53
  LPCVD, 53
  PECVD, 54
  UHCVD, 53
Chemical-Mechanical Polishing, 57
chip size packaging (CSP), 111
CMP, see Chemical-Mechanical Polishing
coefficient of thermal expansion, 99
compensation, 111, 117
  actuator, 119
  explicit, 117
  implicit, 117
  monitored, 117
conductance, 105
conformality, 40
Coventor, 21
crystallographic planes, 32
CSP, see chip size packaging
CTE, see coefficient of thermal expansion
CVD, see chemical vapor deposition
Czochralsky process, 56
Deep reactive ion etching, 60
  ARDE, 63
  etching lag, 63
  notching, 63
  ripple, 62
  scalloping, 63
Delphi, 7
design rules, 57
diaphragm, see membrane diffusion, 44
diffusion barrier, 43
DLP, 8 20 29 59 99 107
DRIE, see Deep reactive ion etching
diffusion, 44
diffusion barrier, 43
diaphragm, see membrane
diffusion, 44
diffusion barrier, 43
DLP, 8 20 29 59 99 107
DRIE, see Deep reactive ion etching
diffusion, 44
Diffusion barrier, 43
DLP, 8 20 29 59 99 107
DRIE, see Deep reactive ion etching
microelectronics, 5, 13
integration, 67
 microloading, 62
Microsens, 97
miniaturization, 6
model, see simulation
molecular conductance, 105
Motorola, 7, 98
MUMPS, 57

NTT, 83

OMM, see Optical MicroMachines
Optical MicroMachines, 8, 85
overetch, 57
Oxford System, 61
oxidation, 43

pattern generator, 26
pattern transfer, 24
patterning, 24
PentaVacuum, 60
permeability, 102
photolithography, 26
mask, 25
photoresist, 24
negative, 25
positive, 25
physical vapor deposition, 49
piezoelectricity, 80
piezoresistive effect, 11, 77
piezoresistor, 78
plasma, 50
ICP, 61
POEMS, see MEMS polymer
polycrystalline, 42, 55
polymer, 29
process, 23
additive, 24, 41
back-end, 26, 91
front-end, 26
modifying, 24, 43, 44
subtractive, 24, 29, 36
PVD, see physical vapor deposition
pyrolysis, 53
quartz, 65

rapid thermal processing, 55
reactive ion etching, 36
release etch, 58
reliability, 19, 20
Reynold’s number, 73
RIE, see reactive ion etching
RTP, see rapid thermal processing
sacrificial etching, 41
sacrificial layer, 26, 39, 58
Sandia National Laboratory, 28, 57
scaling laws, 15
scratch-drive actuator, 85
SDA, see scratch drive actuator
Sensonor, 110
sensor, see MEMS sensor
Sercalo, 8, 28
shape-memory alloy, 88
shape-memory effect, 88
silicon, 27
Silicon Light Machines, 8
simulation, 21
SIP, see system in the package
SMA, see shape-memory alloy
SOI, 18, 28
sol-gel, 48
spin-coating, 47
spin-on-glass, 48
spring, see suspension
sputter, 51
DC, 51
magnetron, 52
RF, 52
stiction, 59
structural layer, 39

134
structure
  active, [69]
  passive, [69]
STS, [14]
SU8, [65]
surface micromachining, [39]
Surface Technology Systems, [61]
suspension, [69]
  folded-beam, [70]
Suss Microtec, [35]
Suss microtec, [14]
System in the package (SIP), [92]
system in the package (SIP), [68]
System on chip (SOC), [92]
Tanner Research, [57]
Texas Instruments, [8, 10, 20, 29, 59, 84, 107, 114]
TI, see Texas Instruments
underetch, [30]
Van der Walls force, [59]
wafer bonding, [37]
  thermocompression, [38]
Wafer level packaging (WLP), [110]
wetting, [75]
Wheatstone bridge, [77, 118]
XactiX, [60]
XeF₂, see xenon difluoride
xenon difluoride, [60]
Young-Laplace equation, [76]
zeta potential, [85]