NUMERO: **2558A** ANNO: 2023

A P P U N T I

STUDENTE: Vinci Flaminia

MATERIA: Physics of NanoBioSystems, Theory Lessons - Prof. Ricciardi

Politecnico di Torino

Physics of NanoBioSystems

Master's degree in Nanotechnologies for Smart and Integrated Systems
A.Y. 2022/2023 - Prof. Carlo Ricciardi

Theory Lessons

Author: Flaminia Vinci, s300003

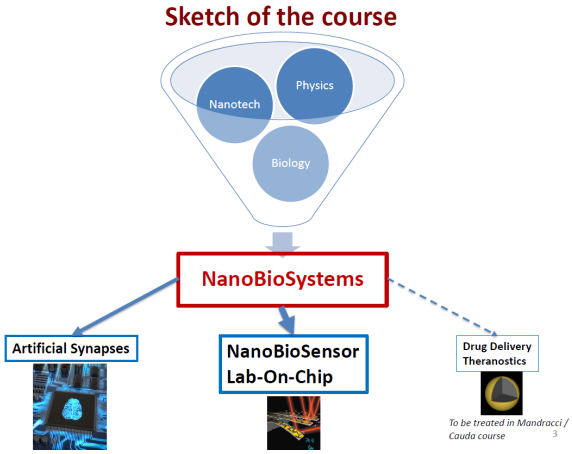
© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55/B - Torino / Pagina 3 di 235

Contents

© Proprietà riservata dell'autore - Digitalizzazione e distribuzione a cura del CENTRO APPUNTI - Corso Luigi Einaudi, 55/B - Torino / Pagina 7 di 235

Introduction

The topic of this course is on NanoBioSystems, i.e. systems of nanometric scale dealing with the biology world.
For example, Coronavirus is a typical example of nanobiosystem, but also an artificial synapse.



The antigenic test is an example of NanoBioSensor, since it exploit nanometric or micrometric capillary forces to move the fluid inside interacting with gold nanoparticles (AuNPs). We will see better its working principle in the following lesson.
Nanoparticles can also be exploited for drug delivery in theranostic application.

We define **Nanotechnologies** the technologies with features having at least one dimension sized between 1 and 100 nm.
Between 100nm and 1μm we are in the submicron scale and if the features have size near 1μm we talk about **Microtechnology**.
Finally **Microelectronics** cover the study of both microtechnology and nanotechnology, since its final aim is to fabricate micrometer-scale or smaller electronic devices, e.g. microchips and MEMS.

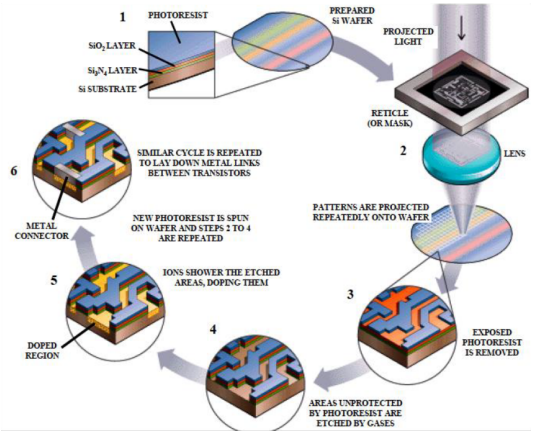


Figure 2: Pattern transfer process

1.2 Microtechnology

Microtechnology or Micromachining, is a generalization of what we said up to now. Indeed, the number of applications in microelectronics can be increased by introducing in the chip **free-standing structures**, like microcantilevers.

1.2.1 MEMS

The most famous example is the MEMS (Micro-Electro-Mechanical-System), present everywhere, in gyroscopes, accelerometers, microphones, DMD[1] and so on.
In particular, let's see the example of the **pressure sensor**, whose layout and fabrication process is shown below.

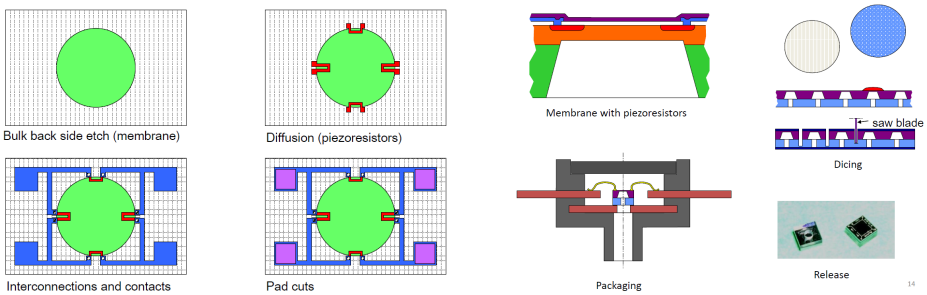


Figure 3: Pressure sensor layout (left) and fabrication process (right)

Like with microchips, also with MEMS, the costs can be reduced when they are parallelized. However, MEMS are much bigger components w.r.t. transistors, since there is a mechanical physical movement instead of just an electron flow. For instance, in the pressure sensor the final package dye is on cm scale.

1.3 Nanotechnology

Nanotechnology embeds all the technologies coming from physics, chemistry, biology and engineering for design and use of materials and devices at **atomic or molecular scale**.

[1] digital mirror device

1.3.1 How big is a nanometer?

The nanometer is 1 bilionth of a meter, but to become aware of how much small it is, let's see it as 1/50.000 of the diameter of a human hair, or the 40% of the diameter of a DNA molecule. It's fundamental to know that also the dimension of the instrument, e.g. a tweezer, must be of the same order of magnitude of the sample, like in biology a virus is recognized by an antibody.

Nanowire

We define a nanowire any solid material in the form of wire with diameter smaller than about 100nm. The nanowire has different properties w.r.t. their microscopic or macroscopic counterparts. The reason behind this is that with nanowire you approach to the atom scale, so its properties are dominated by quantum mechanics and not classical physics.

Nanotube

The nanotube is a hollow nanowire, tupically with a wall thickness on the order of molecular dimensions.
The smallest nanotube is the single-walled carbon nanotube (SWNT) consisting of a single graphene sheet rolled up into a tube.

Nanoparticle (Nanodot)

Finally, the Nanoparticle is a particle having one or more dimensions of the order of 100nm or less.

1.3.2 Nanofabrication

We have said that microtechnology is always top-down. Instead, with Nanotechnology we can also take advantage of the bottom-up approach.

- **Top-down**: you start from the sample substrate and you trow away material or you add several layers to make nanoscale objects.
- **Bottom-up**: the assembly of nanoscale objects comes from even smaller units (e.g. atoms and molecules). This is what nature does, for example, with the DNA molecule, which is originated by simply 4 basis.

REgarding nanobiosystems, you typically use the top-down for the larger part and the bottom-up for the smaller part.

Top-Down

The top down method for nanofabrication is the same for microfabrication, but now the problem is the cost and heat dissipation which can be very large.
To partially compensate the first issue, most of the industries are looking for new technologies like **Nanoimprint Lithography** which consists into the physical transfer of a mold pattern into the resist thickness deposited onto a substrate.

Self Assembly

Self Assembly is the principle behind bottom-up processing. It occurs every time a solid device interacts with an object coming from a bio world. This leads to the formation of an interface layer made of molecule called **self-assembled monolayer**.
Thanks to self assembly we can fabricate very ordered structures of any shape.

See video 01_*sam* about the formation of a SAM on a solid-liquid interface

The fascinating idea is that molecule are able to spontaneously organize maintaining the lowest potential energy, replicating the symmetry of the crystal solid substrate. All of this is mediated by the presence of a functional group attached to the NP.

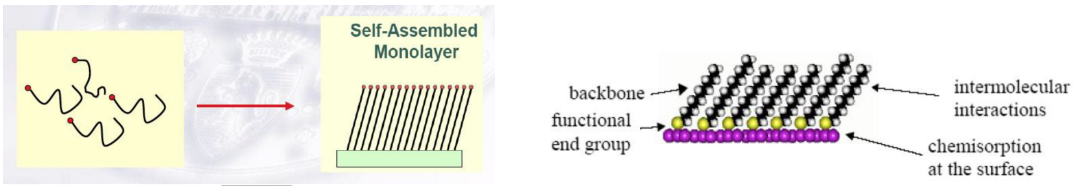


Figure 7: Self Assembled Monolayer (SAM)

This creation is a dynamic process, where a lot of molecules attach between themselves, and a lot of other molecules detach, until the equilibrium condition is reached.
Every dynamic process is mediated by temperature.

1.3.3 Future of top-down and bottom-up processing

Nowadays, the idea is to start from a top down approach, e.g. DUV[2] lithography with a resolution of 200nm, and then move to a bottom up one, e.g. a BCP[3] self assembly forming a random structure. By combining the patterned resist with the PS-b-PDLA, we obtain features with pitch around 10nm, excepting from some errors, which are always present in this method.

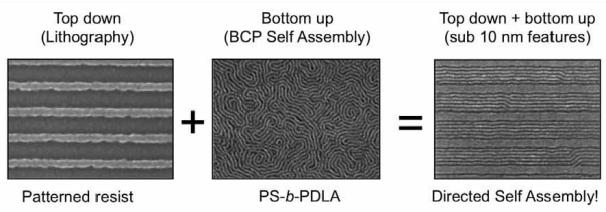


Figure 8: Intel idea of combining lithography and self assembly

1.4 What's new in nanotechnology

At the nanometer scale, intrinsic **properties** become **size-dependent**. For example:

1. Thermal properties. When the nanocrystal size decreases, the surface energy increases and at the end the melting point decreases. e.g. 3nm CdSe nanocrystal melts at 700K compared to bulk CdSe at 1678K.

[2]deep ultraviolet
[3]block coblock polymer

Scaling laws

2.1 Scaling issues

Miniaturization always concerns with not just scaling of dimensions, but also **scaling of forces and properties**.
Competing physical effects do not follow the same scaling and our common intuition of typical macro-scale phenomena (i.e. classical physics) can lead to false deductions when applied to microscale objects:
– Effects negligible become important, and vice versa
– Example: Gravity become negligible in microcantilever w.r.t. its macroscopic counter part in swimming pools. At nanoscale, materials properties are even size-dependent:
– Surface effects become important
– Quantum effects may become dominant when the interatomic distance range is reached

Reducing dimensions by a factor **s** causes:

- **volume forces** (mass) to be reduced by $\boldsymbol{s^3}$
- **surface forces** (pressure, friction, electrostatic..) to be reduced by $\boldsymbol{s^2}$
- **line forces** (surface tension[1]) to be reduced by $\boldsymbol{s}$

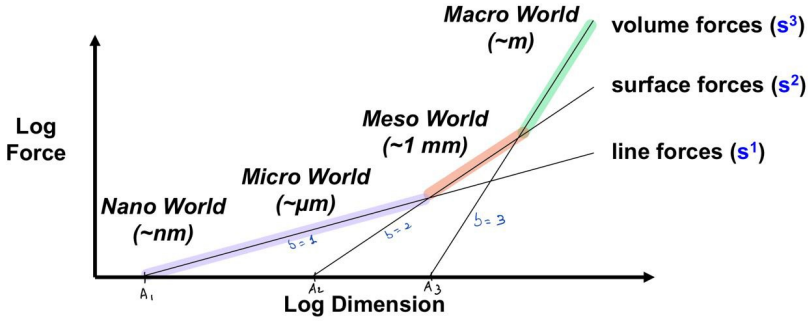


Figure 9: Dependence of forces on dimensions. In the violet portion line forces are dominant, in the orange one surface forces dominates, in the green one volume forces are dominant

Gravitation, for instance, is mass and volume dependant, so all the three dimensions are involved when sizes are shrinked. e.g. if $s = 10^3$, volume forces effect scales as 10^9.

[1] surface tension is a line force because it is given by the ratio between a volume and a surface force

2.2.2 Gravitation and Adhesion

Friction, stiction and adhesion are three separated concept, thus don't confuse them.
Let's start from the **adhesion** between a solid and another solid (or a liquid) is due to forces between atoms and molecules.
The main forces responsible for adhesion are Van der Waals type forces (dipole interactions from neighboring molecules). These are enhanced in a solid/liquid interface, thus in nanobiosystems, which deals always with fluids, it's fundamental to talk about adhesion.
With adhesion we always talk about a chemical property, not physical. No movement is present and, for that, adhesion is a static force.
Being dependent on the number of chemical interactions, F_{VdW} varies like the contact area where molecules are present.

$$F_{VdW} \sim L^2 \qquad and \qquad \frac{F_{VdW}}{F_{gr}} \sim L^{-1} \tag{2.3}$$

i.e. the adhesion force dominates the gravitational force at small L.
From this we can start to talk about the principle of **capillarity**, which is basically the ratio between adhesion (e.g. due to water on a surface of glass) and gravitation (due to the fluid that tends to go down).
The critical value at which both forces are equal depend on the distance between the solids x and on the nature of the medium in between them. Normally, below L = 1 mm, F_{gr} is much less than F_{VdW}.

2.2.3 Friction

On the other side, **friction** is surface independent and it originates from the movement of one solid w.r.t. another.
When two surfaces are sliding about each other, at the **macroscopic** level, the friction force is given by $\boldsymbol{F_{fr} = \mu F_{gr} = \mu mg}$, where μ is the friction coefficient and it is constant.

$$F_{fr} \sim L^3 \tag{2.4}$$

The fact friction is not dependant on the contact area may seem counter intuitive but, actually, once the movement has started there's no surface dependence. The generally admitted physical reason is that two rough bodies touch each other only at three points.

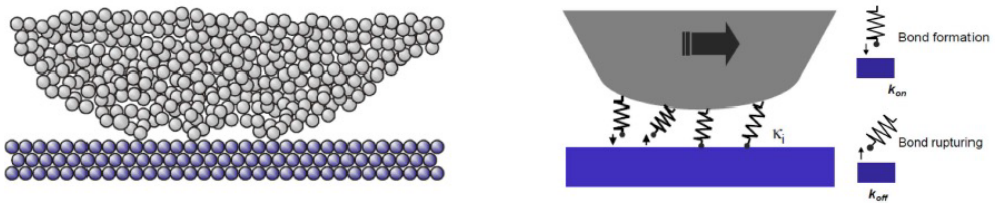


Figure 10: Friction between two solid surfaces

As can be also understood from the figure, no chemical interaction is involved in this process, it's only a matter of physical interaction with few (3 are typically enough) molecules.
The friction principle will return later in resistive switching, where a sort of conductive path is originated inside an insulator. This path is localized in few atoms.

In conclusion, we can say that adhesion ($\sim L^2$) scales less than friction and gravitation ($\sim L^3$). Actually, there's also a third effect which is **stiction**. It is, basically, the static friction and it is more

Allometric scaling: others [OPTIONAL]

Basal metabolism of mammals (that is, the minimum rate of energy generation of an organism) has long been known to scale empirically as (Kleiber's law).

$$B = dQ/dt = const.M^{3/4} \tag{2.9}$$

Also connected to lifetime of mammals.
The brain mass also scales similarly $M_{brain} \sim M^{3/4}$

2.2.5 Beams

Beams, strings or tubes are characterized by resonance frequencies and bending.
The lowest eigenfrequency ν corresponds to a state where the length of the device equals a quarter or half the wavelength λ . From the well known relation $v = \lambda\nu$, one obtains that

$$\nu \sim L^{-1} \tag{2.10}$$

It turns out that resonance frequencies are large in small systems.
The deflection length δ, of a cantilever beam loaded by its own weight varies as

$$\delta \sim L^2 \tag{2.11}$$

Meaning that a beam a thousand times smaller bends a million times less due to its own weight. Again, micro/nanomechanical systems are stiffer than macro-counterpart.

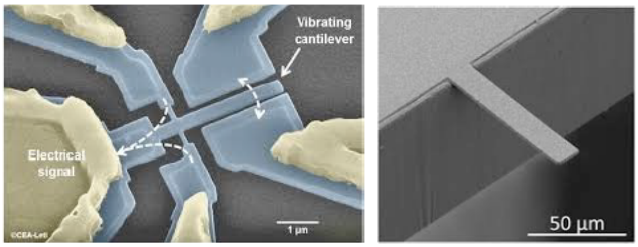


Figure 11: Microcantilever beam

2.3 Fluidics

2.3.1 Terminal velocity

When a body falls 'vertically' in a fluid (liquid or air), **viscous friction**[3] is such that it falls with a constant velocity v_{lim} after a transient time τ, that is proportional to v_{lim}.

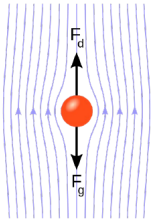


Figure 12: Solid sphere falling in a fluid

[3]this friction is surface dependant since there's an interaction between a solid and a flui, not solid/solid

Pressure driven flow results in a parabolic profile with no-slip boundary conditions (velocity at the interface is zero)

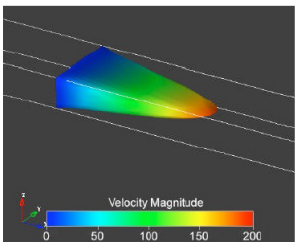


Figure 13: Parabolic profile of a pressure driven flow

2.3.3 Diffusion

Diffusion is the random motion of particles, so it is not deterministic and can be explained only with a stochastic approach.
A particle travels a distance L by diffusion during a diffusion time given by

$$\tau_{diff} = \frac{L^2}{\alpha D} \qquad \rightarrow \qquad \tau_{diff} \sim L^2 \tag{2.18}$$

where α is a geometrical constant and D is the diffusion coefficient. This is valid for particle and thermal diffusion and not effective at macroscale.
Diffusion times scales quickly with dimensions, thus being a relatively efficient way of fluid mixing and object moving at micro/nano-scale, without the need to increase temperature to increase D.

2.4 Electromagnetism

2.4.1 Ohm's law

We know that a conductor of length L and cross section S has an **electrical resistance** $R_{el} = \frac{\rho_{el} L}{S}$, with ρ_{el} the electrical resistivity. Thus

$$R_{el} \sim L^{-1} \tag{2.19}$$

When a constant voltage V_{el} is applied, the electric current I_{el} is given by the **Ohm's law** $I_{el} = V_{el}/R_{el}$, making it scaling proportionally with L

$$I_{el} \sim L \tag{2.20}$$

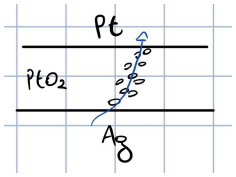
Consequently, the power dissipated by the element is given by **Joule's law** $W = R_{el} I_{el}^2$ and

$$W \sim L \tag{2.21}$$

So, smaller Ohmic conductors are less dissipative.
However, this W is just the **power dissipation per device**, which has a different behaviour w.r.t. the dissipation per unit of surface. Thus, at the end, the dissipation will increase in micro/nanoelectronics. because for the same area more devices are put.
As predicted by Moore's law, in micro/nanoelectronics the number of elements per unit area should scale like L^{-2}. So the electrical **power dissipated by unit area** varies like

$$W_{un} \sim L/L^2 = L^{-1} \tag{2.22}$$

- **Tunnel effect**, i.e. electrons are able to cross the potential barrier.
- **Resistive switching**[4], where the high E-field is able to move also ions together with electrons. An example is the extraction of silver (Ag) atoms buried in a Pt sample, forming an insulating layer in between.



2.4.2 Capacitor [OPTIONAL]

In a capacitor, capacitance, charge and energy stored decrease linearly for decreasing size. Indeed, a capacitor made of two parallel plates of area S, separated by a distance d has a capacitance $C = \epsilon_0 S/d$, so that

$$C \sim L \tag{2.27}$$

When one applies a voltage, the charge on each plate becomes $Q = CV$

$$Q \sim L \tag{2.28}$$

Under these conditions, the energy stored in the capacitor $E_{cap} = Q^2/2C$ scales as

$$E_{cap} \sim L \tag{2.29}$$

2.5 Magnetism

It's not totally true that we have only three families of electrical materials, i.e. conductors, semiconductors and insulators.In fact, an insulator like SiO_2 at nano level, due to the presence of tunnel effects, behave like a conductor, even if the same material.
The same concept can be applied to magnetism, where materials are classified as diamagnetic, paramagnetic, ferromagnetic and, at nanoscopic level, superparamagnetic.

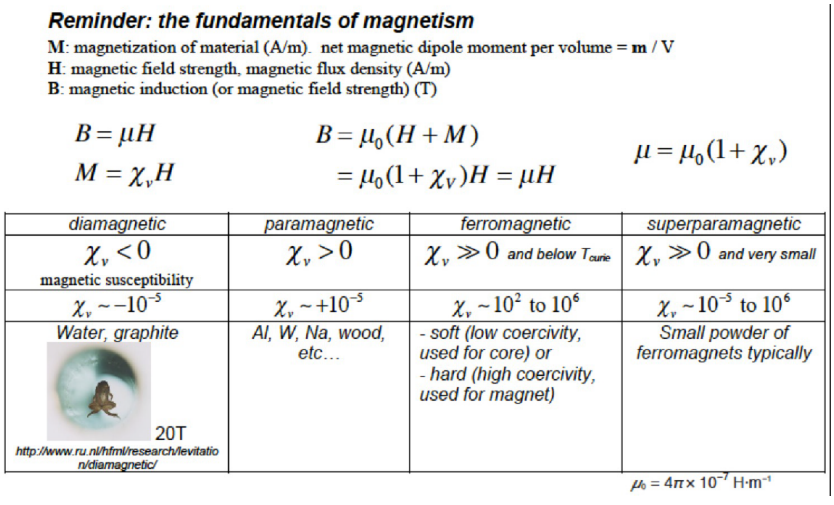


Figure 14: Reminder: the fundamental of magnetism

[4]this is important because in biology we deal always with ionic currents not electron ones

2.5.1 Optics

The limitations of optical methods typically arises from wave optics, in particular **diffraction**. The **divergence angle** is $\theta \approx \lambda/L$, so that, with scaling, it will increase as

$$\theta \sim L^{-1} \tag{2.30}$$

So in photolithography, when shrinking the minimum diameter of the irradiated zone (and thus the minimum feature of the system), the wavelength required to design an element of diameter scales like

$$\lambda \sim L \tag{2.31}$$

This is why the specialists of microelectronics go to ultraviolet lithography when they decide to diminish the size of the electronic components on chips.
The angular resolution $\Delta\theta$ of an element (like an eye) limited by a circular aperture of diameter L is also given by:

$$\Delta\theta \sim L^{-1} \tag{2.32}$$

Thus, one of the biggest problem for humans scaled to micro/nano-dimension would be the vision! To have the same resolution, we would need eyes thicker than our entire body in order to be able to absorb far smaller wavelenghts like x-rays or even gamma rays!

2.6 Themodynamics

2.6.1 Losses

We know, as rule of thumb, that the energy E_{th} to heat a system to a temperature T is proportional to the mass, so it scales with the volume

$$E_{th} \sim L^3 \tag{2.33}$$

On the opposite, the energy lost, in the form of **heat losses**, is depending on the surface and occur by conduction, convection and radiation.

$$P_{diss} \sim L^2 \tag{2.34}$$

This is the reason why, when shrinking the dimensions, since S/V increases (i.e. S more important than V) it's dissipated more than what is created.
The time τ_{th} needed to homogenize the temperature in a system of a given shape is proportional to the square of the linear dimensions of the system, i.e. in small systems the temperature can be changed very quickly.

$$\tau_{th} \sim L^2 \tag{2.35}$$

This is very nice when having, for example, a PCR test aimed to check the DNA of a virus. To do this, very quick step are needed, and the quicker they are, the higher is the final sensitivity.
[OPTIONAL]
If humans would be scaled to micro/nano-dimension, the heat power dissipation would be relatively much more rapid than at our actual size. That's why, coherently, allometry predict a roughly spherical behavior for small objects like cells.
In fact, on one side we would have to decrease the area of our body in order to maintain our temperature. On the other side, the amount of chemical energy necessary for compensating these losses is proportional to the volume. This implies that the more adequate shape for our body would be a sphere, i.e. the solid that minimizes surface with regards to fixed volume!
However, in order to maintain a constant temperature, the quantity of food to be eaten during a fixed

This explains why the number of surface atoms is different from the number of atoms in the borders.

2.6.4 Melting

The **melting temperature** is known to decreases when L decreases. Since the ratio S/V is not small in nanosystems, it is expected that the effects of the surface on the cohesive properties cannot be neglected.
In inorganic materials, the melting temperature, T_m , varies with the diameter of the particle, D , like ($T_{m,\infty}$ is the bulk melting point):

$$T_m = T_{m,\infty}[1 - \alpha/D] \qquad \Delta T/T = (T_{m,\infty} - T_m)/T_{m,\infty} = \alpha/D \tag{2.36}$$

where α depends on the material. It is between 0.4 and 3.3 nm. If we take $\alpha = 1nm$, it turns out that

$$\Delta T/T = 10^{-2},\ when\ D \approx 100nm$$

$$\Delta T/T = 10^{-1},\ when\ D \approx 10nm$$

2.7 Final considerations about scaling

Competing physical effects do not follow the same scaling.
At micro/nanoscale:

- Surface to volume ratio drastically increases
- Line and surface forces are increasingly more important
- Gravitation can be often neglected
- Adhesion, stiction and viscous forces are dominant
- Suspended structures are stiffer
- Diffiusion is more efficient, while no turbulence occurs
- Superparamagnetism arises
- Heat losses increase, melting temperature decreases

If we were isometrically shrinked to cellular dimension (μm):

- Feet and bones too large → no movement
- Surface to volume ratio too large → complete heat dissipation or continuous eating
- Angular resolution a million time worse → blindness

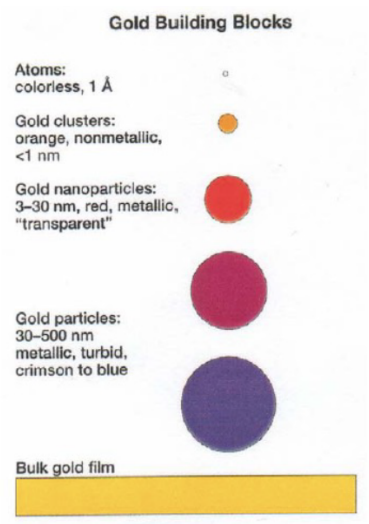


Figure 20: Gold building blocks, from the atomic to the mesoscopic, and their changing colors

The explanation to such effect is given by Quantum Confinement

3.2 3D Quantum Confinement (or 0D transport)

3.2.1 Optical Absorption in Semiconductors

Let's consider a photon with energy $h\nu$ impinging onto a semiconductor surface. If its energy is higher than the bandgap energy E_g of the semiconductor, this will cause the excitation of an electron from the valence band to the conduction band, creating a so called **electron-hole pair**. The energy gap is fixed, since it is an intrinsic property of the material.

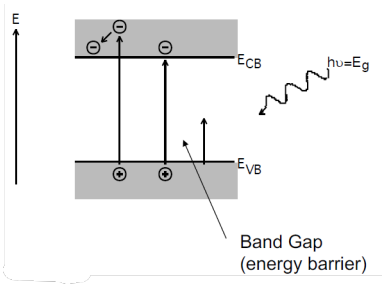


Figure 21: Photon absorption

This e/h pair originates a bound state, called **exciton**, in the solid semiconductor material. The concept of the Bohr radius r_B can be extended to describe the characteristic size of an exciton, since it represents the typical separation between the electron and the hole within an exciton.

$$r_B = \frac{\hbar^2 \epsilon}{e^2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) \tag{3.1}$$

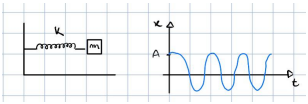
Typical values are:
- CdSe $\rightarrow r_B = 6nm$
- ZnS $\rightarrow r_B = 5nm$
- GaAs $\rightarrow r_B = 10nm$

This equation is very similar to the one of the **harmonic oscillator**, for which:

$$F = ma = m\frac{d^2x}{dt^2} = -kx \qquad \rightarrow \qquad \frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \tag{3.4}$$

From this we get

- Frequency $\omega = \sqrt{\frac{k}{m}}$
- Harmonic oscillation $x(t) = Acos(\omega t)$



Equivalently for our problem:

- $\alpha = \sqrt{\frac{2mE}{\hbar^2}}$
- $\Psi(x) = Acos(\alpha x)$

By applying the boundary conditions to the general solution (which up to now seems the classical one)

$$\Psi(x) = Ae^{i\alpha x} + Be^{-i\alpha x} \tag{3.5}$$

we are actually introducing the quantization, i.e. the squeezing of the electron in the box

$$\Psi(0) = A + B = 0 \qquad \rightarrow \qquad A = -B \qquad \rightarrow \qquad \Psi(x) = A(e^{i\alpha x} - e^{-i\alpha x})$$

$$\Psi(a) = 0 \qquad \rightarrow \qquad \Psi(a) = 2iAsin(\alpha a) \qquad \rightarrow \qquad \alpha a = n\pi$$

Thus

$$\alpha^2 a^2 = \frac{2mE}{\hbar^2}a^2 = n^2\pi^2$$

$$E_n = \frac{\hbar^2\pi^2}{2ma^2}n^2 = cn^2 \qquad n = 1, 2, 3.. \tag{3.6}$$

i.e. $E_n \sim a^{-2}$, the quantization of energy depends on dimensions, or in other words, the distance between two levels of energy depends on the size of the box (in our case the NP). As a consequence, the energy of photons emitted from electrons that are moving among the levels changes with dimensions too.

3.2.4 Quantum confinement effects

But, how much should be the dimension a in order to see quantum confinement?
To answer to this question let's start by rewriting from 3.6

$$c = \frac{\hbar^2\pi^2}{2ma^2} \tag{3.7}$$

Where
$\hbar = 10^{-34} J \cdot s$
$m = 9 \times 10^{-31} kg \approx 10^{-30} kg$ (free mass of electron in vacuum)
$1eV = 1.6 \times 10^{-19} J$
Thus

$$c = \frac{10^{-68} \cdot 9.86}{2 \times 10^{-30} a^2} \approx \frac{5 \cdot 10^{-38}}{a^2} J \approx \frac{5 \cdot 10^{-38}}{1.6 \times 10^{-19} a^2} eV \approx \frac{3 \times 10^{-19}}{a^2} eV \tag{3.8}$$

3.2.5 3D Confinement: Quantum Dot

Here, in the following, is shown a colloidal solution of nanoparticles.

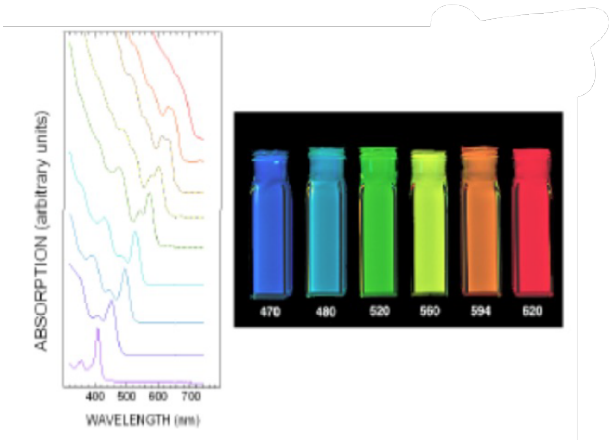


Figure 25: (left) absorption spectra of smc NPs of different diameter. (right) NPs suspended in solution

Going from the left (blue) to the right (red) suspension, the dimension of the NP and so also the wavelength are increasing. Viceversa, the frequency ν and the distance ΔE between two energy levels are decreasing.

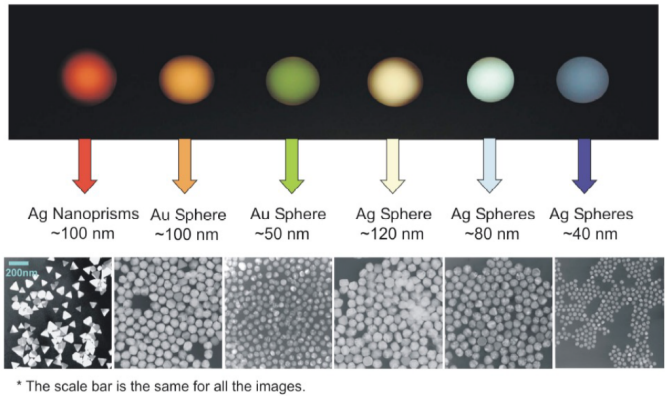


Figure 26: The smaller is the dimension of the NP, the more you go toward blue color

3.3 1D-2D Quantum Confinement (or 2D-1D transport)

[QUALITATIVE PART]
In this section we will explain the **electronic transport** in 1D and 2D nanostructures, the quantum wire and the quantum well respectively.

3.3.1 Free electron gas

In order to see what happens when 1 or 2 dimensions are squeezed the **Bethe-Sommerfeld model** is a good instrument to see what happens to an electron free to move in a solid.
The free electron gas is characterized by a periodic potential, illustrated at figure 27.

1D (unconfined) DOS

In this case, electrons are free to move only in one direction, because the other two dimensions are of nanometric size.
The final DOS is inversely proportionial to the square root of the energy.

$$D(E) = \frac{1}{\pi\hbar}\sqrt{\frac{2m}{E - E_0}} \tag{3.16}$$

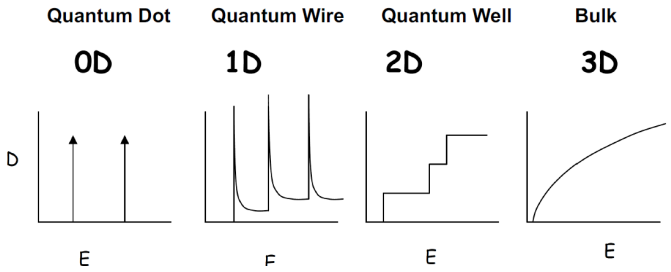


Figure 28: Density of available quantum states as a function of energy for systems of various dimensionalities

In conclusion, we can say that nano objects are not only size dependant but also shape dependant, in the sense that they depend on how many dimensions are nano. This concept is called **dimensionality**.

3.4 Quantum wire and Ballistic Transport

In this section we will make a distinction between quantum and ballistic behaviour. Indeed, they typically appear together, but they are two different things.

3.4.1 Mesoscopic and Ballistic Transport

First of all, three **characteristic lengths** must be defined:

- **Fermi wavelength $\lambda_F = 2\pi/k_F$**, it is the electron wavelength at the Fermi energy, and thus the smallest wavelength present in the system. As we said in 3.2.4, if the dimensions of our object are comparable with such wavelength, there's quantum confinement.
- **Mean free path $l_e = v_F\tau$**, it is the average distance that an electron travels before loosing its initial momentum (several collisions are usually needed).
  In a diffusive transport, the electrons experience both a drift[1] and a thermal velocity which may lead to collisions with phonons where they loose energy. The time between two successive collisions is τ.
  Thus, to have **diffusive charge transport**, characterized by the Ohm law, the not confined direction (e.g. of a nanowire) must be larger than the mean free path.
- **Phase coherence length L_ϕ**, it is the average distance over which an electron travels before loosing its initial phase (random phonon interaction).

[1] it is a collective velocity

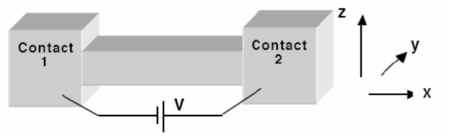


Figure 30: Ballistic wire

In a ohmic conductor, there are a lot of electrons free to move randomly, making the final conductance pretty high.
On the opposite, if electrons are sent one by one, like in the ballistic transport, each of them will experience less losses (but still not zero) but they are in low amount, so the final conductance is lower and the resistance is high, around $17k\Omega$ (due to the dissipation at the contact between ballistic sample and the rest of the circuit).

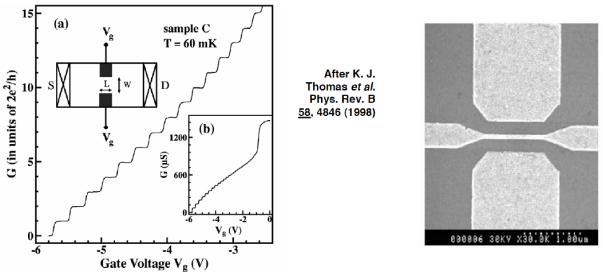


Figure 31: Staircase-like quantum conductance for a molecular transistor

Also here in the following is shown the conductance of a memristor, whose behaviour is attributed to the nanoscale filamentary structure of Cu conductance pathways formed inside SiO_2.

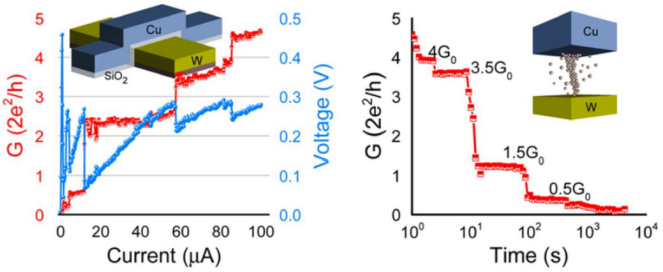


Figure 32: $Cu/SiO_w/W$ memristor with half-integer quantum conductance states

Nanoparticles, Nanowires and CNTs

In the previous chapter we've seen that a nanosystem is not always a quantum one, while a quantum system is typically nano.

4.1 Nanoparticles

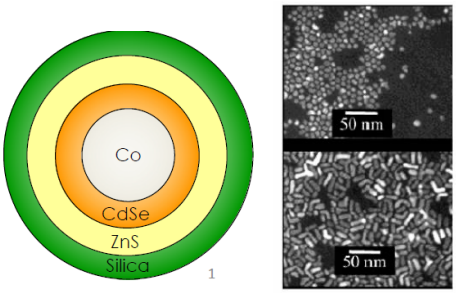


Figure 34: Quantum Dot Nanoparticle

Especially in BioSystems, when you talk about Nanoparticles, in the majority of cases you are talking about Quantum Dots.

- **Quantum Dot**: semiconductor nanoparticles that exhibit quantum confinement (typically less tahn 10nm in diameter)
- **Nanoparticle** (more general): a particle of inorganic material (e.g. Au, CdSe) or organic material (e.g. polymer, virus) with a diameter less than 100nm

4.1.1 Optical properties of Quantum Dots

In the majority of cases, QDs are used to shine light in *in-vivo* systems. Indeed, by exciting it with UV light, you get fluorescence in output.
QDs distinguishes from fluorophores in different ways.
An **organic fluorophore** is a dye able to absorb in the UV and emit in the visible range.

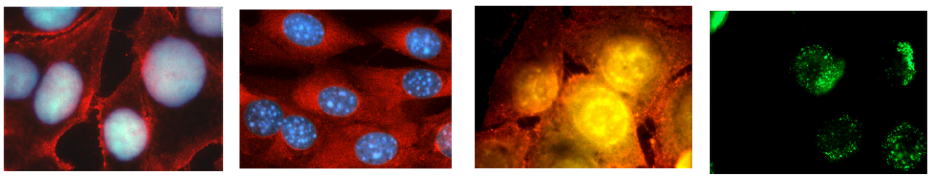


Figure 37: Some examples of in vitro imaging with QDs

QDs are also better than semiconductors because they are characterized just by sharp levels of energy, allowing a net simple transition at a certain frequency. Whereas semiconductors have a VB and a CB and an energy gap between them, without possibility of tuning.

4.1.2 Formation of Nanoparticles

Several methods for the **synthesis of NPs** are available. The right synthesis technique is chosen as a function of the material, desired size, quantity and quality of dispersion.
In the majority of cases, a **colloidal**[2] **solution of NPs** is created, coming from a rapid reduction of an organometallic (e.g. chloroauric acid $H[AuCl_4]$)in hot organics with surfactants.

Nucleation and Growth

Nucleation is the creation of nuclei upon which growth can occur.
Growth is the addition of more gold particles to the existing nuclei.
Finally, coagulation is the creation of the larger gold particles, such as 20nm, which requires a coagulation of multiple (smaller) twins of various shapes. A conglomeration of multiple nuclei into particles can be large enough to disturb the stability and fall out of the colloid, so it is very important to control the coagulation process.
This is a whole **bottom-up technology** characterized by NPs of different dimensions.

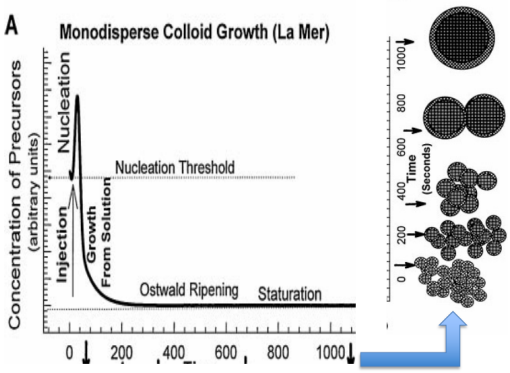


Figure 38: Stages of nucleation and growth

4.1.3 Synthesis of gold nanorods

Once that the seed, i.e. the NP, is created, it can be exploited to build a not isotropic object, the **nanorod**. There are basically two steps:

1. produce seed particles
2. seed grow into rod

[2]homogenous dispersion of particles in a solution

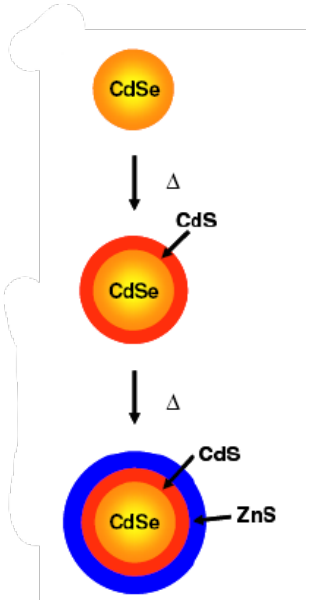


Figure 40: Capping of a CdSe QD with CdS/ZnS

The cap layer has an energy gap that is transparent in the frequency at which the QD emits. What you do in practice is to tune the photoluminescence (PL) to higher wavelengths, because the particle is larger.

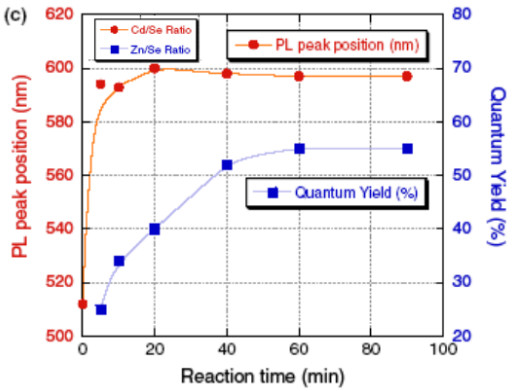


Figure 41: PL peak position and Quantum Yield variation during the clapping process

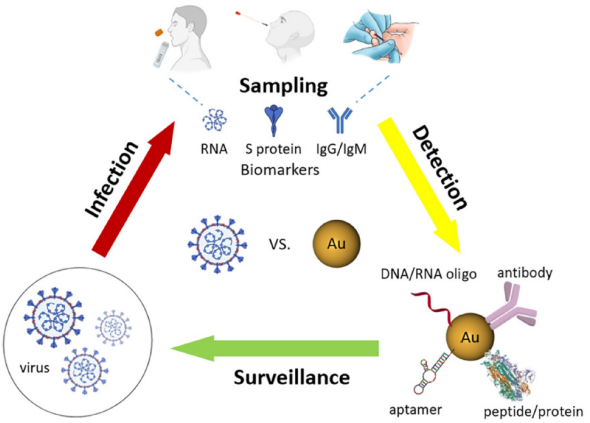


Figure 43: Working mechanism of the COVID antigenic test

Let's see how the **antigenic test** works.
Inside the test there's already the nanoparticle properly functionalized with several moieties, e.g. antibody, oligonucleotides, peptides, aptamers. The NP is dissolved in a liquid solution. Once that a drop is put inside the test, the liquid will move by means of capillarity dragging the NPs with itself.
In correspondence of the test (T-)line some antibodies are present to recognize the S-protein. Hence, if the virus is present, the Abs of the T-line will bind to the protein, the line will be colored red, resulting into a positive test.
The color intensity of the T-line is higher than the C-line because most of the nanoparticles stop there. Instead, on the control (C-) line other antibodies, devoted to target specifically the NP antibody, are present. Such binding occur always, and it is performed to check the validity of the test. If only the C-line will colour, the test is negative.
Finally, most of the time, once that the test is finished, the lines assume a black color. This is due to the fact that nanoparticles aggregate and they don't have quantum confinement anymore.

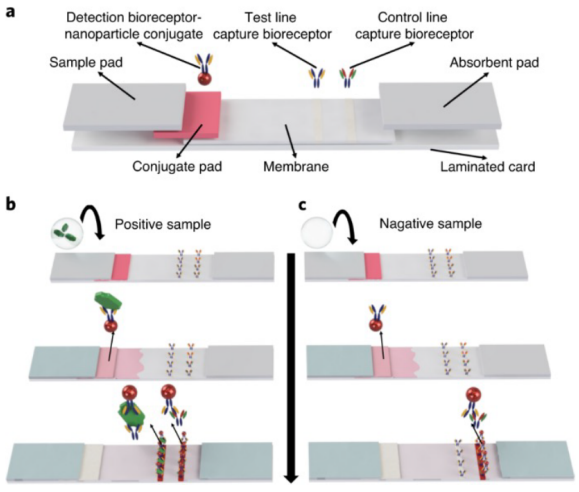


Figure 44: Working mechanism of the COVID antigenic test (continuum)

Moreover the drug concentration must be maintained as much as possible constant in the days, while with a conventional formulation (picked behaviour) the treatment should be repeated multiple times to be efficient.

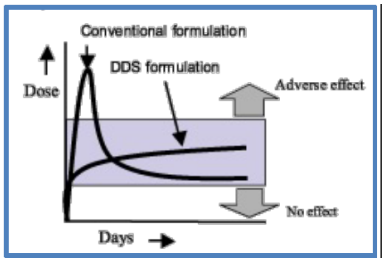


Figure 47: Drug Delivery System formulation vs conventional formulation

NPs for in-situ surgery

Once that NPs are inside the tumoral cell, they can perform a controlled physical destruction of tumoral cells by means of cascade signaling which will induce the cell to apoptosis. If heat is exploited, this treatment is called **local hyperthermia** and at the following link "Click here" is shown a nice example offered by the *MagForce company*.
Its cancer therapy is based onto an **iron-oxide NP**[4]. The particle is covered by a painted coating which ensures a good stability in the biological fluid. The diameter measures around 20nm, 500 times smaller than a red blood cell.
At the beginning of the therapy, the NP solution is injected directly into the tumor (e.g. a glioblastoma, brain tumor). After being injected, the NPs spread out in the space between the tumor cells. At this point, the patient enters the therapy device, in which an alternating (not dangerous) **magnetic field** is produced. The result is a thermal effect which leads particles to oscillate, causing the cancer cells to die either from active self-destruction or from swelling until they literally burst.
Tumor growth is stopped and destroyed cells are discharged by the body in a natural process. As a rule, the one hour minimally invasive treatment is repeated six times. However, the particles are only injected once, thus making the therapy especially gentle on the patient.

Another example about cancer therapy is offered by *Nanospectra Biosciences company* at the following link "Click here".
It is another way of treating solid tumors by means of **thermal ablation**. The **N-IR particles** are infused into the bloodstream and after a long circulation inside the body, they accumulate on the tumor site. Due to the rapid growth of most tumors, there, the blood supply is poor, resulting in holes large enough for particles to exit the bloodstream and accumulate in the tumor and not in the surrounding tissue (EPR effect). Particles which are not accumulated in the site of interest are eventually removed by liver and spleen.
At this point, a **laser fiber** is inserted through the skin near or into the tumor, and laser energy is applied. This energy is absorbed by particles which convert it to heat, resulting in the temperature rising up to a level sufficient to destroy the tumor.
As a result, the tumor will shrink.

4.2 Nanowires

[OPTIONAL, see slides 06_*addendum_NW_CNT_graphene*]

[4] magnetic nanoparticle

Memristors and Resistive Switching

The first NanoBioSystem to be discussed is the **artificial synapse**. The Artificial Intelligence is based on the idea of mimic how the brain works and , since the brain is made of neurons and synapses, is fundamental to look at synapses.

5.1 Memristor as fourth electrical element

The three well known circuital elements are resistor, capacitor and inductor. Their properties are not a combination of the other two, in the sense that, for example, we can't create a resistor by joining an inductor and a capacitor.
The idea of the professor Chua, however, is that a **fourth element** should exist and it is described starting from two definition, one for the current and one for the voltage, and two constitutive relations explicit in the following.

$$\begin{cases} q = \int i\ dt \rightarrow\ charge\ q\ as\ "current\ momentum" \\ \phi = \int v\ dt \rightarrow\ flux\ \phi\ as\ "voltage\ momentum" \end{cases} \tag{5.1}$$

Hence, from these we expect 4 independent relations between: v and i, v and q, ϕ and i and ϕ and q. $\rightarrow$ 4 circuital elements.

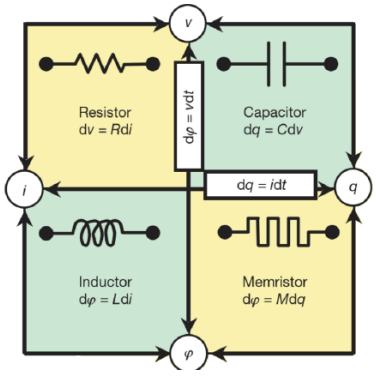


Figure 48: 4 circuital elements and their relations

1. **Resistor**

$$v = R \cdot i \quad or \quad dv = R \cdot di \tag{5.2}$$

has been switched off but the in output we obtain always something.
Again, the difference with memristor is that there's no hysteresis, because the loop is pinched in (0,0).

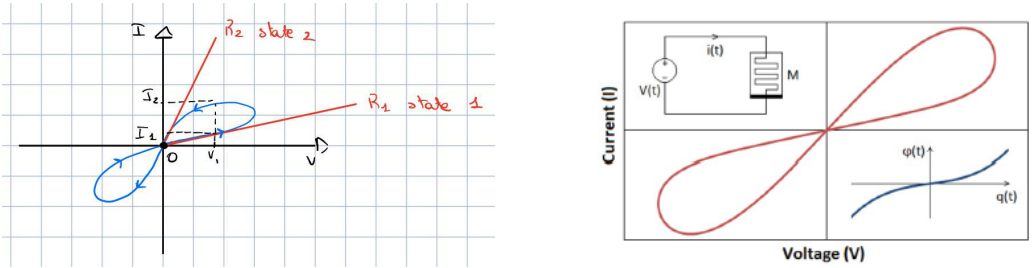


Figure 51: I-V curve of the memristor

However, also here, the same value of voltage v_1 corresponds to two different states of current I_1, I_2.

Another thing we have to highlight is that M depends- on q which, by definition, is the integral of the current in time, i.e. it is a picture of what is happening in time. Thus M is time dependant too or, in other words, **memresistance depends on history**.
This is a very powerful system because it has the ability of memorizing and adopting, useful for machine learning applications.

The reason why we have to introduce this new parameter $M(q)$ is that, as can be seen from the constitutive relation 5.5, if it was constant it would have been just a resistor.
Indeed, for the resistor

$$\underbrace{\int vdt}_{\phi} = \int Ridt = R\underbrace{\int idt}_{\text{q}} \tag{5.6}$$

And analogously, for the memristor:

$$\int vdt = \int Midt \neq M\int idt \tag{5.7}$$

reconfirming that the only possibility is that M should depend on q (or idt) and it has a memory of the charge.

In 1976 was discussed that the memory should be related to a **state variable ω** , which for the ferromagnetic material is the magnetic domain. Such variable is not necessarily electrical and modifies the constitutive relations of 5.5 as

$$d\phi = M(q(t),\omega)dq \quad or \quad \frac{d\omega}{dt} = f(q,\omega) \tag{5.8}$$

The outcomes of what we said up to now are that:

- The memristor is a **passive two terminal device** since, by definition, is a circuital element. Being two terminal, it is **low power**, i.e. it doesn't need a keeping voltage, you apply a voltage and read the current just when is needed. There are no capacitive effects.
  Thanks to this low power property and to the fact that it is highly scalable, the memristor is preferred to a three-terminal device (e.g. a transistor), to realize more complex systems like an artificial synapse, where high density of devices and low power are desired. Indeed, if you want to create an artificial intelligence, you don't want fixed behaviours, but devices able to learn.

5.2.3 I-V not suitable

A memristor should be defined in terms of current momentum q, voltage momentum ϕ and the function of state variable f.
Pinched hysteresis I-V curves are not suitable to define if a particular device is a memristor or not.Pinched hysteresis I-V curves are just the specific response to a given input.

5.3 The Resistive Switching

The term **resistive switching (RS)** includes any concept which relies on the change of the resistance at the nanoscale by changing the configuration of atoms to realize a binary (or multinary) switch.
Indeed, as we have already discussed, if we deal with a 1nm device on which we apply 0.1V, the final electric field is of the order of $10^8\ V/m$, higher than the saturation value. Thus, it is so large that we are able to move the atoms (ions) of the lattice, and not only the electrons (see figure 58).
Moreover, in some cases, like we will see in the ECM configuration (5.5), some Ag^+ ions are extracted from one electrode and move toward the insulator, creating a filament (fuse) attached to the other (Pt) electrode.

If such RS stimulus is so that it affects a state variable somehow related to the resistance, we talk about **Memory Resistive Swtiching (MEM-RS)**. At the end the modified resistance values are memorized in the element.
To resume, the memristor is an ideal circuital element and resistive switching is the solid state physical reason of the memristor.
Before 2008, people talked about **Nanoionics** as the RS-based memories.
This is a very powerful instrument, because in nature (biology), electricity is always mediated by ionic currents, not electronic. e.g. our synapses are ruled by ions.

The device resistance switches between

- **HRS: High Resistance State** or OFF state
- **LRS: Low Resistance State** or ON state

And two possibility of switching are possible

- **Bipolar Switching**, most frequent, where SET and RESET depend on voltage polarity (but at the same amplitude)
- **Unipolar Switching**, where SET and RESET depend on voltage amplitude

5.3.2 Unipolar Switching

- The starting point is the LRS condition
- Then, for $\boldsymbol{V = V_{RESET}}$ the state is changed from LRS to HRS, for which we obtain a very low current
- By further increasing the voltage up to $\boldsymbol{V_{SET}}$ we go back to LRS

Both V_{SET} and V_{RESET} are positive and they are different in magnitude, in particular

$$V_{RESET} < V_{SET}$$

making the curves evolving only in one quadrant.

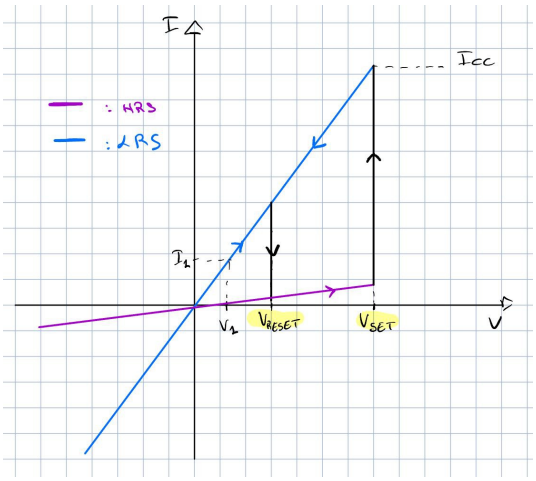


Figure 54: Unipolar Switching

5.3.3 Resistive Switching Classification

Different kind of RS are possible, we will see in detail *Valence change memory (VCM)*, *Electro chemical metallization (ECM)* and *Phase Change Memory (PCM)*.
Both ECM and VCM are typically bipolar, while PCM is unipolar.

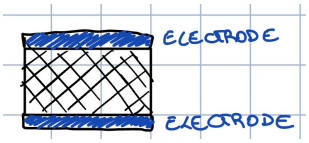


Figure 56: LRS, crystalline phase

2. Secondly, you apply a small voltage ($V_{RESET} = V_1$) for a long time. In this step, the material as enough time (higher than the relaxation time) to **become amorphous**.
   actually, the result is a hybrid material with a crystalline and an amorphous portion. Where it is amorphous it is less conductive and we say that we are in the **HRS**.

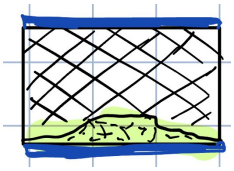


Figure 57: Amorphization. In green is represented the amorphous portion

3. And, then, viceversa: you apply a large voltage ($V_{SET} = V_2$) for a short time. Now a transition from an amorphous to a **crystalline (LRS)** state is registered. In this condition the conductivity is higher.

In general

$$V_2 >> V_1$$
$$\underbrace{t_2}_{\sim ns} << \underbrace{t_1}_{\sim 100\mu s}$$

Where V_2 and V_1 have the same polarity, thus we're dealing with an **unipolar device**. In other words, with these kind of devices negative weights cannot be assigned, which is good from an endurance point of view in neuromorphic computing.
This working mechanism is fully reversible.

(C) Under the high field[4] the ion is accelerated in the electrolyte, towards the cathode (Pt) side, and at the same time the voltage is increased up to V_{SET} ($\approx 0.3V$).
Since Platinum is inert, there's no reaction with silver ions, thus the only possibility is a reduction at the cathode side.

$$Ag^+ + e^- \rightarrow Ag \tag{5.13}$$

(D) A filament is created (**forming**[5]) and a jump in the current is registered, because electrons now can move across this fuse. Now the insulator basically exhibits a metallic behaviour, meaning that the resistance changes from HRS to LRS.

(E) This system is totally reversible. Indeed, once that the current reaches I_{cc}, which is a reference current aimed to tune the diameter of the filament, the voltage is lowered back and when $V = 0$ also $I = 0$, because the output current is given by electrons.
Then, for reversed polarity (i.e. anode negative and cathode positive), from the Pt electrode Ag atoms are extracted and ionized, i.e. Ag^+ moves toward the negative side, which now is the anode made of silver. This reaction is called Ag filament rupture.

$$Ag \rightarrow Ag^+ + e^- \tag{5.14}$$

Thus, RESET occurs when the fuse is broken

To resume, the dynamic of the system, in terms of Set and Reset, is defined by the ions, while the output signal, i.e. the readout 0 or 1, is determined by the electrons.
This explains also why $\boldsymbol{V_{SET} > V_{RESET}}$, because the set voltage is the potential to create the filament, while to break the filament (reset), the extraction of one atom is enough. Thus, overall, we can say that, from the second cycle onwards, in order to change the state is enough to move just one atom.
Finally, for what regards the curve of figure 58, we observe a non linear behaviour, starting at V_{SET}, typical of the insulator, and a linear one, from D to E, of a metallic behaviour.

The main **limitation** of an ECM device is that it tends to break even alone with time, because it relies on just few metallic (so highly energetic) atoms. Indeed, if you switch off the voltage, the filament tends to become a sphere, in order maximize its surface energy. → The result is a **volatile memory**. However, on the other side, this principle is very similar to our synapses, which don't work if no stimulus is applied.

5.5.1 ECM mechanism: filament growth dynamics

Depending on the ion mobility μ (i.e. how much ions are moving) and on the redox rates Γ^i (i.e. how much the ions tend to collaps together) we can observe different conditions for the ion filament creation.

[4] which goes from left (+) to right (-)
[5] actually the forming is the creation of the filament for the first time

5.6 Valence Change Mechanism (VCM)

VCM is able to overcome the ECM limitation and, for this reason, it is the elective device for RAM memories.

- As ECM, VCMs are **bipolar** devices.
- Both the **electrodes are not electromigrating**, usually inert.
- The **insulator** in between is **nanoscale, amorphous** and typically a **transition metal oxide** like TiO_x, HfO_x, Ta_yO_x. Transition means more than 1 valence state.
- Also in this case the dynamic of the system is regulated by **ionic current**, but this time, the ions don't come from metal electrode, they are **oxygen vacancies**.
  And since the oxide is amorphous it is not perfectly stoichiometric, i.e. usually the molar fraction x of the oxygen is less than 2. This explains the oxygen vacancies, which are positively charged, because oxygen is an electronegative element.

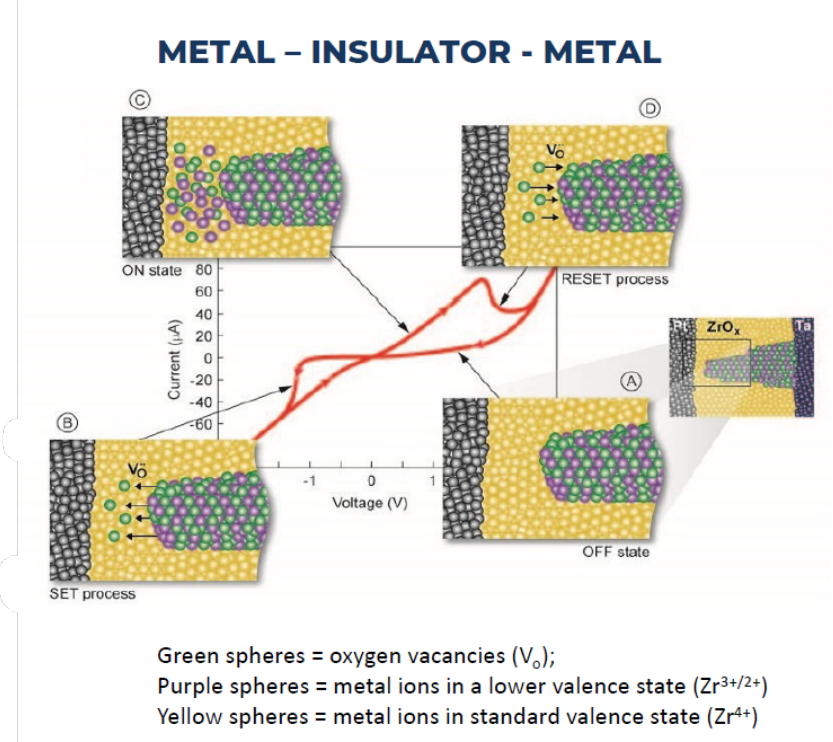


Figure 61: Valence Change Mechanism. The two electrodes are made of Pt and Ta, with ZrO_x in between

In particular, in the example of fig 61 the stoichiometric oxide is zirconia ZrO_2 (yellow spheres) originated from:

$$ZrO_2 \rightarrow Zr^{4+} + 2O^{2-}$$

while non stoichiometric form is ZrO_x (purple spheres) obtained from:

$$Zr^{2+} + xO^{2-} = ZrO_x$$

Exactly as in ECM, a **forming** process is always needed. This is, basically, a first set to introduce a permanent "soft" breakdown of the symmetry. Indeed, if the structure remains perfectly simmetric,

Pay attention that the thickness of the buffer layer is given by the initial conditions of the interface and by the diffusion of atoms through it, not by the voltage.
Referring to the example reported in figure 61, let's see how VCM works step by step.

(A) **OFF (HRS) state**

(B) **SET process**. A negative voltage is applied so that oxygen vacancies, positively charged, are attracted at the Pt anode side (negative biased). This will result into a less and less stoichiometric oxide where the valence state of Zr has been lowered from 4+ to 2+, which means that 2 electrons has been released.
When the relative concentrations of vacancies s enough a jump in the current is registered, going from HRS to **LRS**.

(C) **ON (LRS) state**. A positive voltage is applied, a further current increase is registered and the resulting i-v plot is non linear, indicative of the fact that a semiconductor, not metallic, behaviour occurs.

(D) **RESET process**, where oxygen vacancies return back to the forming part at the Ta side.

A proof that oxygen vacancies are created, is given by the picture below, which is the result of a Cyclic Voltammetry (CV) measure performed at 80% relative humidity for increasing voltage. Indeed, for large voltages, oxygen flows outside of the device and the energy is so large that a gas of oxygen is formed.
We can see the gas evolution starting at the edges of the cell, with bubbles evolving over time.

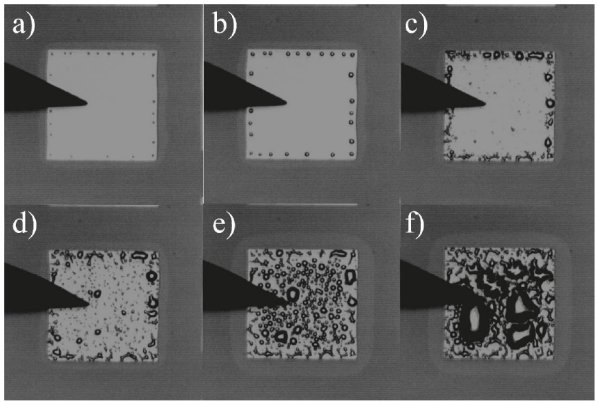


Figure 64: Gas evolution during CV measurement

5.6.1 Differences between ECM and VCM

- The two mechanisms have different **resistivity window**. For the ECM one, the resistivity changes of several order of magnitude, due to the fact that there's a passage from an insulator-like behaviour to a metallic one. While, for the VCM the resistivity fluctuates much less, because it's like only the doping is changing, i.e. you change only the conductivity of the forming region, by increasing oxygen vacancies concentration.

- **VCM** is completely **reversible**, because holes (absence of atoms) move through the oxide, without stressing the device. → a memory based on VCM will last nearly forever.
Instead **ECM** has a **lower durability** because atoms belonging to one of the electrodes are extracted and reinserted.

- Direct tunneling
- Fowler Nordheim tunneling
- Quantum point resistance, due to quantum confinement effects

In these kind of devices, **quantization of current** arises and the reason is the quantum confinement in the filament, limited to few atoms.

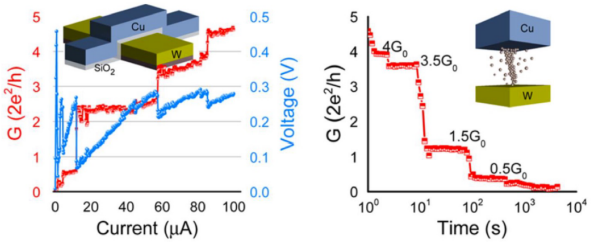


Figure 66: Nonlinear conductance quantization effects

Then, **Joule heating** is determinant in these kind of mechanisms, becuase it is obtained from electronic current and also make possible the ionic current. Thus, at the end, the temperature is our state variable.

See slides 07_08_*Resistive switching* for further details about the simulation methods and some experimental observations.

6.1.1 Nanowire-based Memristors

1D Nanowires benefit of several advantages. In particular, thanks to a bottom-up synthesis process, very small ($\sim 1nm$) NW made of few atoms can be built.
Moreover this realization is low cost and NWs ensure a good interface with biological neural networks.
Each NW is a single crystal, has a high chemical purity and shows a clean surface.

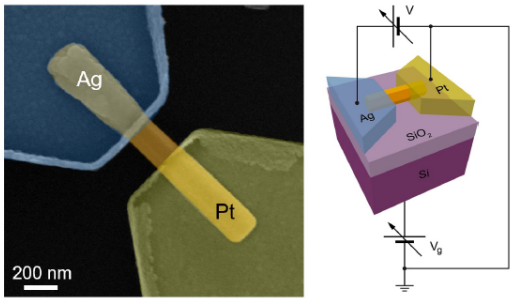


Figure 68: Nanowire Memristor, the active electrode is Ag, the inert one is Pt

In a **NW-based Memristor**, the nanowire is put in between the two electrode (Ag, Pt) in a sort of FET configuration.
Once that a voltage has been applied, the resistive switching mechanism may start.
1. At the anode side, the **dissolution of Ag** is registered and some Ag^+ are extracted, according to the reaction

$$Ag \rightarrow Ag^+ + e^-$$

2. Going on with time and with voltage supply, the **Ag ions drift** under the action of the (huge) electric field, moving along the nanowire.
3. Finally, at the cathod side, **silver reduces and crystallizes**, in line with:

$$Ag^+ + e^- \rightarrow Ag$$

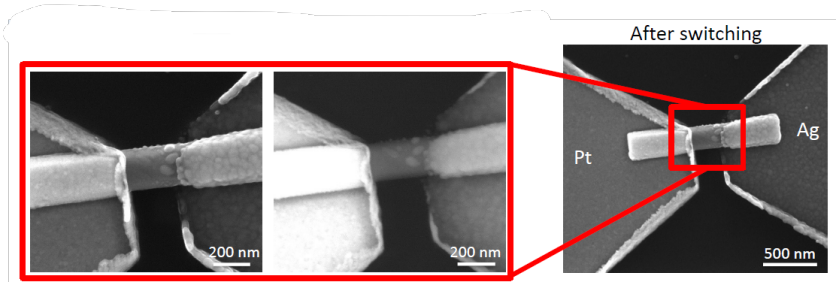


Figure 69: NW-memristor after switching

By directly observating the SEM picture of the Ag conductive path, we can see that it may happen that if the ion mobility is too low in single crystalline bulk, surface mobility occurs, i.e. the ions position at the surface.
This is confirmed also by the X-ray fluorescence images of the section of the Zn-wire, for which the silver ions put all around the edge.

bit 0, while to LRS corresponds high current and so bit 1.

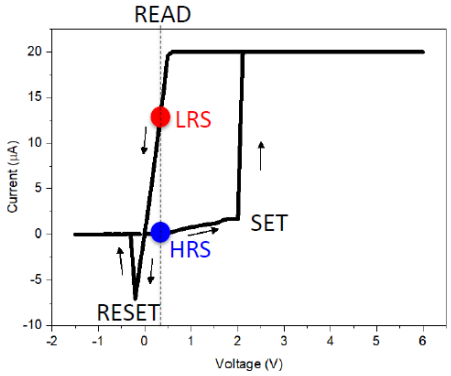


Figure 72: I-V characterization

The passage from one state to the other is, basically, accomplished by just apply voltage pulses.

- **READ**: pulse of small voltage for small time
- **SET**: pulse of high voltage for small time

Pay attention that V_{READ} is not equal to zero, so the device is still perturbed, but the voltage is so small with respect to V_{SET}, that the state is not changed.
During **reading**

- If the output current I is small we are in the HRS (bit 0)
- If the output current I is large we are in the LRS (bit 1)

During **writing** we wanna set the new state, and this is realized by applying a voltage much larger than the reading one.

$$V_{READ} << V_{SET}, V_{RESET}$$

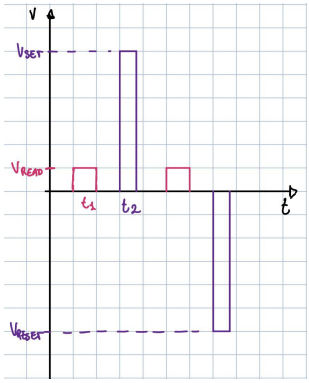


Figure 73: reading and writing mechanism in terms of applied voltage

Thus, at the end, we can say that in order to move from one state to the other, after the forming cycle, we have to just apply large pulses at V_{SET} or V_{RESET}, without stressing the device.
From a neuromorphic point of view this mechanism is similar to the one of the axon potential.

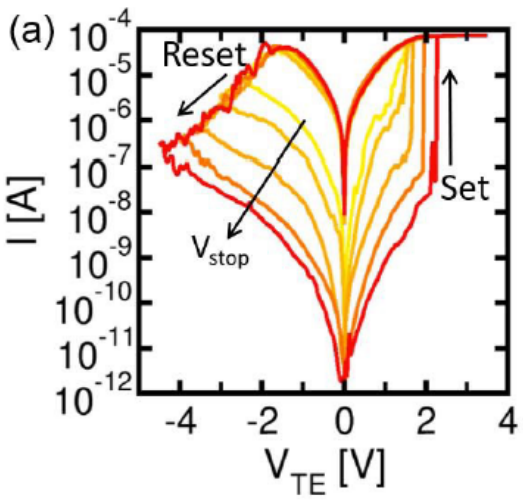


Figure 75: I-V curve for different V_{stop} values. 8 curves correspond to 8 possible different bits

This last two examples represent a step toward an **analog behaviour**, not digital. Because if the cell resistance can be tuned in a continuous manner, infinite states can be reached.

Example of AC memory: ECM on ZNO NW memristor

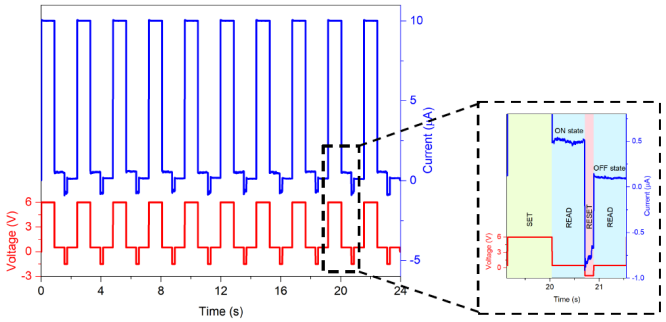


Figure 76: I-V curve for different V_{stop} values. 8 curves correspond to 8 possible different bits

The positive pulses represent the writing, the negative pulses represent the reading, which is performed at low voltages, around 0.1V. Then, higher current values correspond to LRS (ON) state, and smaller current correspond to HRS (OFF) state.
Long (800ms) writing pulses are necessary to preserve the ON state in non-volatile resistive switching.

The brain has a very complex **hierarchy**.
Starting from neurons, they are large cells ($\sim 100\mu m$) that when interacting with the others, form local circuits. The latters are the responsible of our memory, i.e. when we remember about something, we typically "turn on" the local circuit.
More local circuits form subcortical nuclei, then cortical regions ($\sim mm^2$), then systems ($\sim cm^2$) and finally the brain.
The **typical keywords** when talking about brain are:

- complexity
- hierarchy
- connectivity
- redundancy
- self-organization, based on experience; it needs hierarchy and redundancy
- **plasticity**, the main property of brain, it is the ability to adapt to experience, i.e. intelligence

7.1.1 Neurons

The **neuron** is a single, relatively large ($\sim 10 - 100\mu m$), nerve cell with a nucleus. Neurons are the fundamental units of the nervous system responsible for transmitting and processing information in the form of electrical signals.
Neurites ($10^3 - 10^4$)are specialized projections or extensions of neurons. They can be categorized into two main types: axons and dendrites.

- **Dendrites** are branched projections that receive electrical signals from other neurons and transmit them towards the cell body. They have a tree-like appearance, with numerous smaller branches called dendritic spines. Each neuron is made of many dendrites.
- **Axons** are elongated projections that transmit electrical/chemical signals away from the cell body of a neuron in an unidirectional way. They are typically longer than dendrites and can extend over long distances, enabling communication between different regions of the nervous system. Axons are covered by a fatty insulating layer called the myelin sheath, which helps to speed up signal conduction. For each neuron we have only one axon.

Both axons and dendrites play crucial roles in neuronal communication and information processing. Axons transmit signals from the neuron's cell body to other neurons, muscles, or glands, allowing for the relay of information across the nervous system. Dendrites receive signals from axons of other neurons, collecting and integrating incoming information, which ultimately determines whether the neuron will generate an electrical impulse or action potential.
All of this makes the neuron as an integrator able to compress data going from $10^3 - 10^4$ inputs to just a single output.
The intricate network of interconnected axons and dendrites forms the basis of neural circuits, enabling complex communication and coordination within the nervous system.

7.2 Information processing in brain

To understand better how the information is processed in the brain we have to consider four fundamental aspects.

1. The **neuron** can be seen as a **voting system**. When the sum of (electrical) inputs is larger than a certain threshold, an **action potential**[2] is sent through the axon. A sum needs always a voting system, i.e. a consensus.
   This means that the neuron behaves like an integrator from a circuital point of view, that from multi-uncorrelated events (inputs) it gives a single output.
   Such mechanism allows:
   - reduction of the information complexity
   - data compression
   - pattern recognition. The pattern extracted in output is the majority of the inputs.
2. The neuron is not democratic at all, you need a **weighted vote**. All the connections have different "importance", i.e. the inputs have different conductance.
   The final consensus is not just given by the number of connections giving the signal, but it is mediated by their importance.
   This ability to give different importance to different connections (from a chemical point of view) is called **synaptic plasticity** and it is shaped by experience.
   Without hierarchy, the weighted mechanism doesn't work.

   EXAMPLE

   ▷ Consider the sentence *"The apple is good"* → This is level 4, also called **phrase level** and it is the highest one, since the major tool of our brain is language. ChatGPT works at this level.
   The phrase is understoo only when the neuron **A4** is active.

   ▷ When listening to this sentence, our brain, as first thing, recognized the word *"apple"* → This is level 3, the world, or better, the **idea level**. When our brain thinks, it thinks by means of images, ideas. Thus, the first thing we think when we hear the word "apple", is the image of an apple.
   To identify the word *apple* the neuron **A3** is adopted.

   ▷ First the letter *"A"* is understood → This is level 2, the **letter level**.
   To recognize the letter A the neuron **A2** turns on.

   ▷ Thinks how the letter A is composed $/ \oplus \setminus \oplus -$ → This is level 1, the **sign letter level**.
   To each sign of the letter is assigned one neuron of level 1: **A1, B1** and **C1**, and just when they fire together, the letter is recognized.

   The fascinating thing of what we say up to now is that either if we write the word *Apple* or we sketch an apple or even if we smell the flavour of an apple, the same neurons are fired. This means that a part of our brain is shaped, is reactive to the idea of an apple, then it can be recognized by different inputs like: reading, looking at an image, smelling, touching, or hearing (e.g. the cutting of an apple)
   The more we learn, the more we are able to shape connections and have ideas, in other words we are building an **intelligent system**.

[2] it is just a spike or a pulse

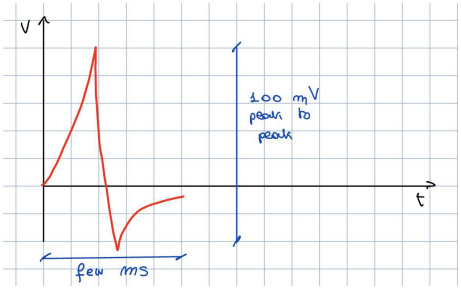


Figure 81: Single action potential

Each neuron sends an **action potential** (81), which is always the same.
The spike arrives at the axon terminal, it travels across it and, in correspondence of the synaptic cleft, it polarizes the membrane. The Ca^{2+} channels will open leaving the vesicles with the neurotransmitters inside to break at that point. Not only Ca, but also other ions are free to move in the cleft, like Na^{+} or K^{+}. [3]
Then, the neurotransmitters reach the postsynaptic cleft and bind to it thanks to some bioreceptors (present only on the dendrites not the axon[4]). There, it changes the pH, which polarizes the membrane that in output produces an action potential.

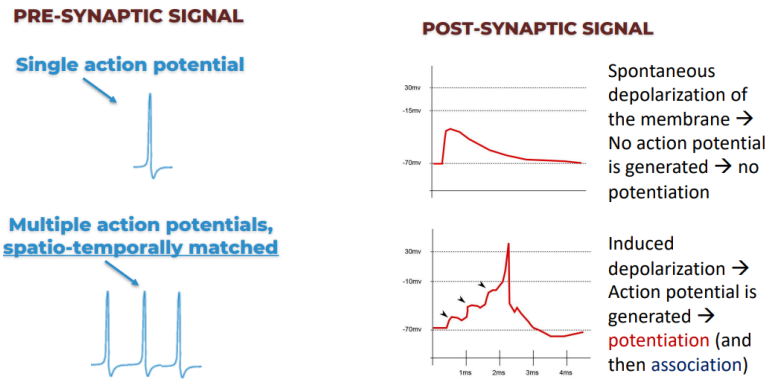


Figure 82: (left) Pre-synaptic signal (right) Post-synaptic signal

The conductance is tuned by changing the number of neurotransmitters that go from one side to the other.
The amplitude of the post-synaptic signal is first small, because just one action potential has been generated in principle, and then the magnitude increase, if the frequency of spikes gets higher.
This is the principle of association explained from a biochemical point of view.
However, he **movement of ions** mediating the connection between one neuron and another will not last forever and this determines the death of the brain.
Furthermore, the synapse can be considered a short memory, i.e. if no more spikes arrive, the post-synaptic signal just last the time of polarization and depolarization of the membrane. This is a very important aspect, since today most of the systems which are trying to emulate the synapses are based on non-volatile (long term) memories, but on the other side, the way we shape the connection is shorter (short potentiation).

[3] a lot of diseases are vaused by a too high (or too low) concentration of Ca^{2+}
[4] unidirectionality

- The switch is activated if the input stimuli from other neurons are spatio-temporally matched.
- When a certain threshold is reached, a pulse (called action potential or just spike) is sent from the information output to other neurons.
- Incoming signals from other neurons or cells are transferred to a neuron by special connections: the synapses. An electrical signal received by the synapse, i.e. coming from the presynaptic side (axon end), is directly transferred to the postsynaptic sites of the cell (dendrites).
- Small molecules called neurotransmitters are released and degradated in the synaptic cleft. There are neurotransmitters that stimulate (excitatory) the postsynaptic cell nucleus, and others that slow down such stimulation (inhibitory).
- Synapses can form stronger (potentiation) and weaker (depression) connections. This adjustability of a synaptic connection is called synaptic strenght and defines the weight each input signal incoming in a neuron has in reaching the threshold.
- Such analogue plasticity (chemical, electrical and physical) is known to be the basis of higher brain functions like memory and learning (Hebbian rule: neurons that fire together, wire together).

7.3 Artificial Synapses

Now, we are ready to move to the **neuromorphic part**, focused on the way neurons can be modelized and on how information can be processed.
In order to build an **artificial synapse**, we can exploit a **memristive device** where ions move in between the two electrodes, where instead of action potential, **pulses**, between 0 and 1, are sent. Each pulse will create a movement of ions and, then, depending on the operating condition (i.e. on the history), the dimension of the pulse can be tuned so that we can analogously change the conductance.

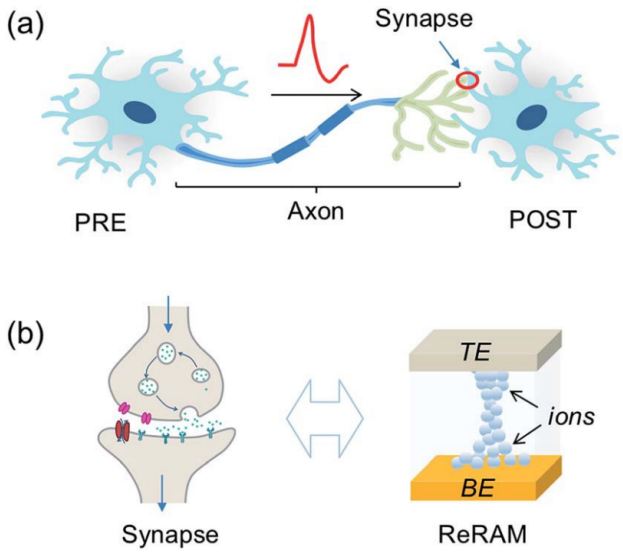


Figure 85: (a)Illustration of a building bloc of the biological neural systems, consisting of a pre- and a post-synaptic neuron and a synapse between the PRE axon and POST dendrite. (b) Comparison between the biological synapse and electronic synapse (ReRAM)

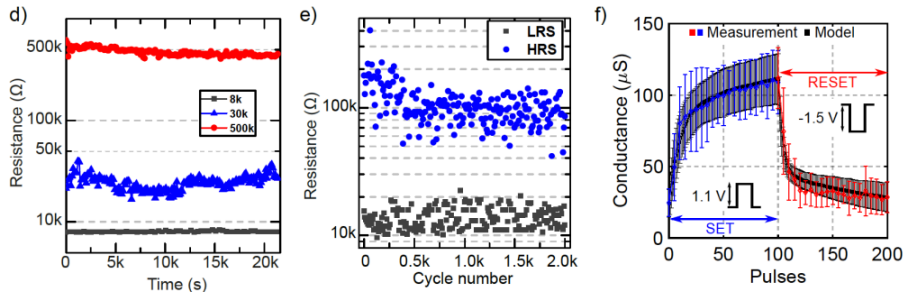


Figure 87: (d) Retention, in black the LRS, in blue an intermediate RS and in red the HRS (e) Endurance (f) Depression Potentiation cycles (DPC)

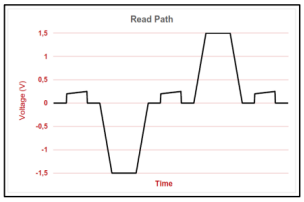


Figure 88: Voltage pulses

In graph (f) of figure 87, we can see the potentiation phenomenon once that a setting pulse voltage of 1.1V has been applied. After 100 pulses, a substantial increase of the conductance has been registered.

Let's see now the structure of a VCM device exploited for long term potentiation **LTP**, already discussed in section 5.6 (*Example case 2*)

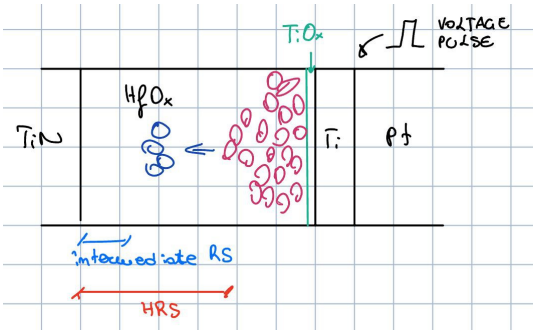


Figure 89: VCM structure

The asymmetry is ensured by the presence of a non stoichiometric layer of TiO_x which takes oxygen from hafnium oxide.
Then, if a voltage pulse is applied on the Pt electrode, **oxygen vacancies** will move from right (+) to left (-). The region is populated until the intermediate resistance state is reached. Indeed, we don't wanna reach the LRS, i.e. the SET condition, otherwise we will reach the habituation state. Instead, we want to change the weight, potentiating the synapse, and this is accomplished by properly working in between the HRS and the LRS.

Because we want an analog response, otherwise habituation occurs.
Moreover, knowing that the retention measures how much time a certain state is maintained, to make the device working, without going to the HRS, we have to ensure that

$$t_4 - t_2 < t_{retention} \tag{7.2}$$

Spike-timing dependent plasticity (STDP)

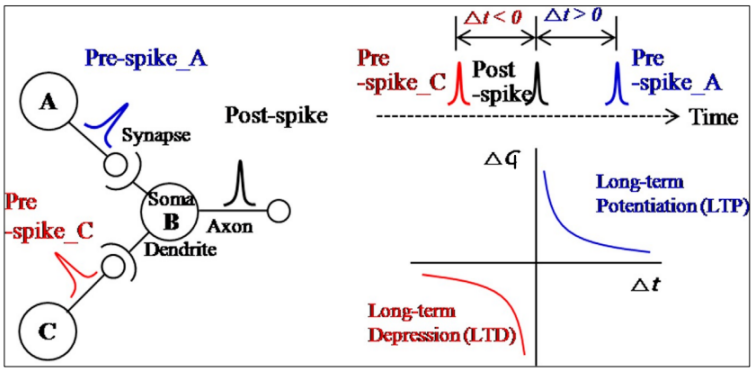


Figure 92: Spike-timing dependent plasticity

Referring to the upper figure, first a spike, labeled as A, is sent in input and in output a post-spike is generated. Then, after a certain time, another potential coming from another dendrite and labeled as C is sent.
In a normal electronic device, both A and C contribute in the same way to the final current. While, for a device whose aim is to emulate a neuron, spike A is more important since it occurs before the output, and C should be neglected.
In fact, to do **potentiation**, the temporal match is fundamental. In other words, the distance in time between pre and post spike must be really close to make the conductance to be high.
The smaller is the distance in time, the larger will result the potentiation.
If $t_{post} - t_{pre}$ is large, this means that PRE is not the cause of POST, i.e. they are uncorrelated events.
For what regard, instead, the signal coming from neuron C, it has to be **depotentiated**, because it arrives after than the consensus has already been reached.
In general:

- Correlation → Potentiation → Current increase
- Uncorrelation → Depression → Current decrease

Volatile threshold switching with NW memristor

In this case, a very short (10ms) voltage pulse is sent and after that a relaxation process due to the spontaneous dissolution of conductive path into spherical Ag nanoclusters is observe. In terms of current this is reflected into the presence of multiple plateau and so **quantized conductance levels**. The threshold switching is repeatable over cycling, with similar relaxation times (retention).

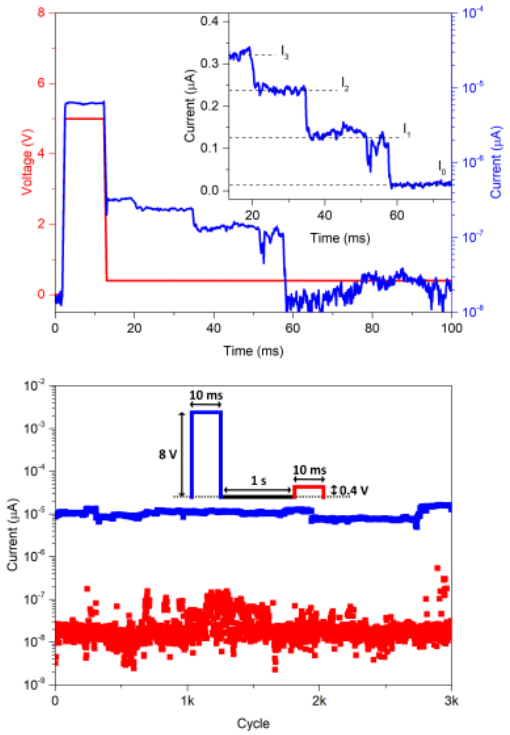


Figure 95: Voltage current behaviour and endurance plot

Neuromorphic behaviour: train pulses

Both the **output current** and **incubation time** can be modulated by **train amplitude**. Taking as a reference the picture below, in red is shown the voltage, i.e. the signal sent to the memristive device, while in blue is represented the output current.

Pay attention that we are always in **short term plasticity**, because if the train of pulses is stopped the current relaxes down to zero.

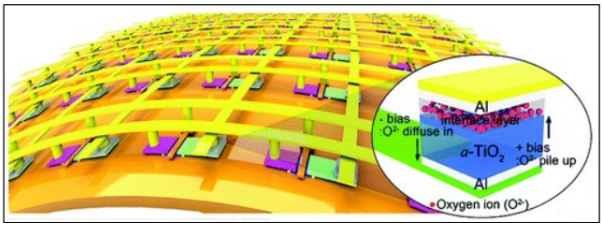


Figure 99: ReRAM crossbar architecture, the vertical wires are the inputs and the horizontal ones are the output

To emulate the neurons network a **RRAM or ReRAM** resistive random access memory in **crossbar architecture** is realized. It could be passive or 1T 1R, which means that every memristor has a transistor, acting as a selector, on top.
However, neuromorphic (brain inspired) computing is a more general term w.r.t. in memory computing, because it can be an architecture also not based in crossbar.
Pay attention that an analogue traditional CMOS circuit needs tens of transistors and capacitors to simulate a single neuron!

With a **memristive device** we are dealing of endurance values around 10^{12} cycles and retention of 10^9 seconds. They offer new computing paradigms which are:

- Better memory storage
- Bioinspired computing
- In-memory computing

Emergency of new computing paradigms

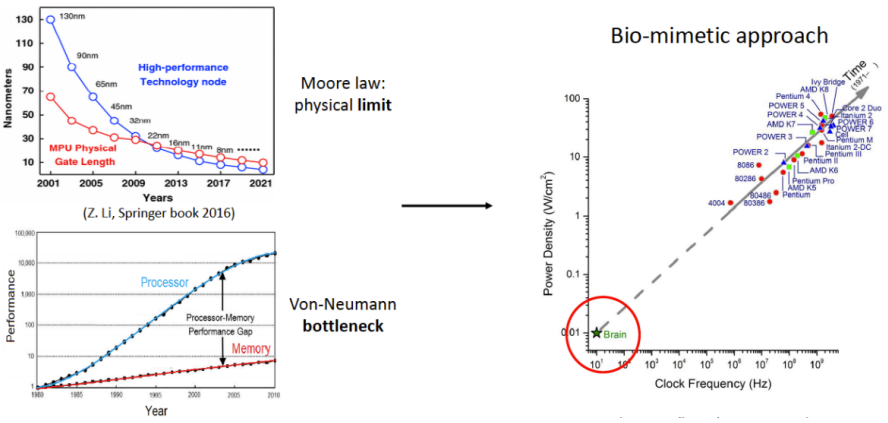


Figure 100: Moore law physical limit, Von-Neumann bottleneck and bio-mimetic approach

If on one side in the last decades a huge increase of the processor performance has been registered, the same can not be said from the memory point of view.
However higher performance means higher clock frequency and, at the end, higher power density, which is an undesired effect.
On the other side, our brain works at very low frequency, in the range of hundreds of ms ($\sim 10Hz$),