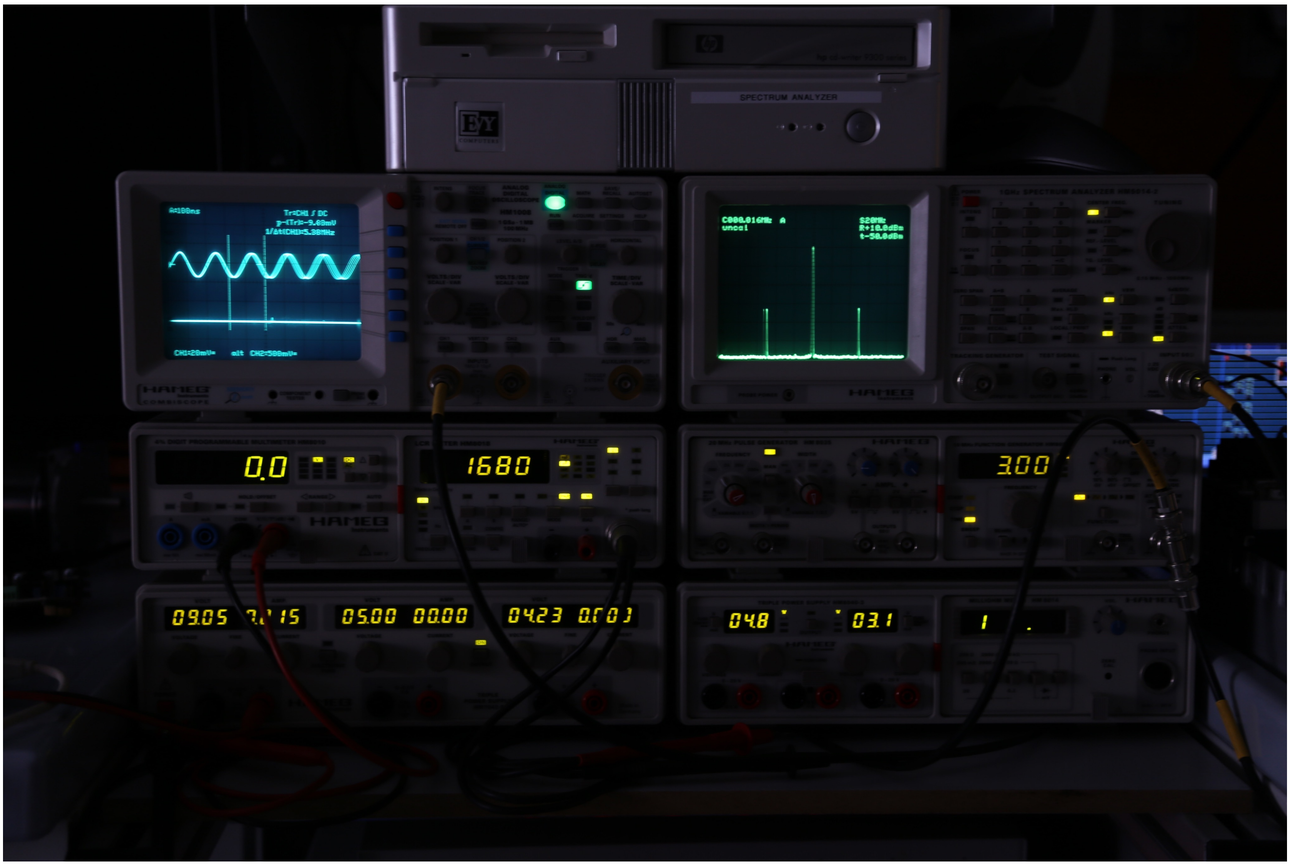
ELECTRONICS FOR NEUROSCIENCE



Erik ZORZIN

INDEX

CHAPTER 1: VECTORS

CHAPTER 2: COMPLEX NUMBERS

CHAPTER 3: SIGNALS

CHAPTER 4: FOURIER SIGNAL ANALYSIS

CHAPTER 5: ELECTROMAGNETISM

CHAPTER 6: BASIC ELECTRONIC COMPONENTS

CHAPTER 7: FROM PHYSICS TO ELECTRONICS

CHAPTER 8: MAIN INSTRUMENTS AND TOOLS

CHAPTER 9: COMMERCIAL COMPONENTS

CHAPTER 10: DC CIRCUITS

CHAPTER 11: AC CIRCUITS

CHAPTER 12: FILTERS

CHAPTER 13: SEMICONDUCTORS

CHAPTER 14: AMPLIFIERS

CHAPTER 15: BASIC DIGITAL CIRCUITS

CHAPTER 16: ADVANCED DIGITAL CIRCUITS

CHAPTER 17: NEURAL NETWORKS

CHAPTER 18: DIGITAL SIGNAL PROCESSING

CHAPTER 19: ELECTRODES

CHAPTER 20: NEURAL RECORDINGS

1.1 Vectors - Column and row vectors

A vector is as a list of elements, which can be ordered in a column or in a row.

COLUMN VECTOR:

$$|a\rangle = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_n \end{pmatrix}$$

(Eq 1.1.1)

ROW VECTOR:

$$\langle a| = \left(a_1, a_2, a_3, \dots, a_n\right)$$

(Eq 1.1.2)

The number of elements in a vector, namely its length, is called its dimension. Such kind of vectors are said to be transposed one respect each other.

The angular bracket symbols usage, as a concise representation of multidimensional vectors and matrices too, is usually called the "Dirac's notation" in honour of the English physicist P.A.M. Dirac who invented it.

(Fig. 1.1.1: P.A.M. Dirac, 1902 - 1984)

1.2 Vectors - Addition

Vectors having the same dimension can be added together element by element.

ADDITION OF COLUMN VECTORS:

$$|c\rangle = |a\rangle + |b\rangle$$

(Eq 1.2.1)

$$|c_1, c_2, c_3, ..., c_n\rangle = |a_1, +b_1, a_2 + b_2, a_3 + b_3, ...a_n + b_n\rangle$$

(Eq 1.2.2)

ADDITION OF ROW VECTORS:

$$\langle c| = \langle a| + \langle b|$$

(Eq 1.2.3)

$$\langle c_1, c_2, c_3, ..., c_n| = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3, ..., a_n + b_n|$$

(Eq 1.2.4)

(Fig. 1.2.1: the inventors of vectors, some thousands of years B.C.)

1.3 Vectors - Inner product

The inner product between two vectors having the same dimension is one scalar number defined as follows:

$$\langle a|b\rangle = \sum_k a_k b_k = a_1 b_1 + a_2 b_2 + a_3 b_3 + \ldots + a_n b_n$$

(Eq 1.3.1)

The inner product of a vector with itself is the norm of that vector, which is the square of its magnitude:

$$\langle a|a\rangle = |a|^2$$

(Eq 1.3.2)

More generally:

$$\langle a|b\rangle = |a||b|\cos(\varphi)$$

(Eq 1.3.3)

where ϕ is the angle between the two vectors.

Notice that for orthogonal vectors their inner product is zero:

$$\langle a|b\rangle = 0 \Leftrightarrow \varphi = \frac{\pi}{2}$$

(Eq 1.3.4)

(Fig. 1.3.1: two orthogonal directions)

1.4 Vectors - Projection

A particular case is the inner product between a vector and another one having unitary magnitude:

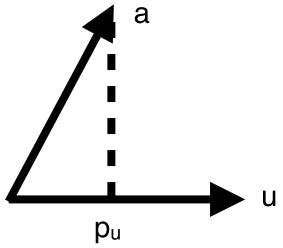
$$|u| = 1$$

(Eq 1.4.1)

The result of this operation represents the projection of the first vector onto the second one:

$$p_u = \langle a | u \rangle$$

(Eq 1.4.2)



(Fig 1.4.1: vector projection)

For any given vector it is always possible to obtain a corresponding vector which is normalised to a unitary magnitude by dividing each of its elements by its original magnitude:

$$|a| = \sqrt{\left\langle \frac{a_1}{|a|}, \frac{a_2}{|a|}, \frac{a_3}{|a|}, \ldots, \frac{a_n}{|a|} \middle| \frac{a_1}{|a|}, \frac{a_2}{|a|}, \frac{a_3}{|a|}, \ldots, \frac{a_n}{|a|} \right\rangle} = \sqrt{\frac{1}{|a|} \cdot \frac{1}{|a|} \cdot \langle a | a \rangle} = 1$$

(Eq 1.4.3)

(Fig. 1.4.1: Leaning tower of Pisa - Italy)

1.5 Vectors - Vector space

A "*N-dimensional vector space*" is the set of all possible vectors having a defined dimension N. Each vector of that space can be represented by a unique linear combination of a vector basis made of N orthonormal vectors:

$$\{|u_1\rangle, |u_2\rangle, |u_3\rangle, \ldots, |u_n\rangle\} \Leftrightarrow \langle u_i | u_j \rangle = \delta_{ij}$$

(Eq 1.5.1: orthonormal basis)

where the greek delta is called the "*Kronecker's delta*" and it is equal to one when "*i*" is equal to "*j*"; it is zero otherwise.

The linear combination is:

$$|a\rangle = a_1|u_1\rangle + a_2|u_2\rangle + a_3|u_3\rangle + \ldots + a_n|u_n\rangle$$

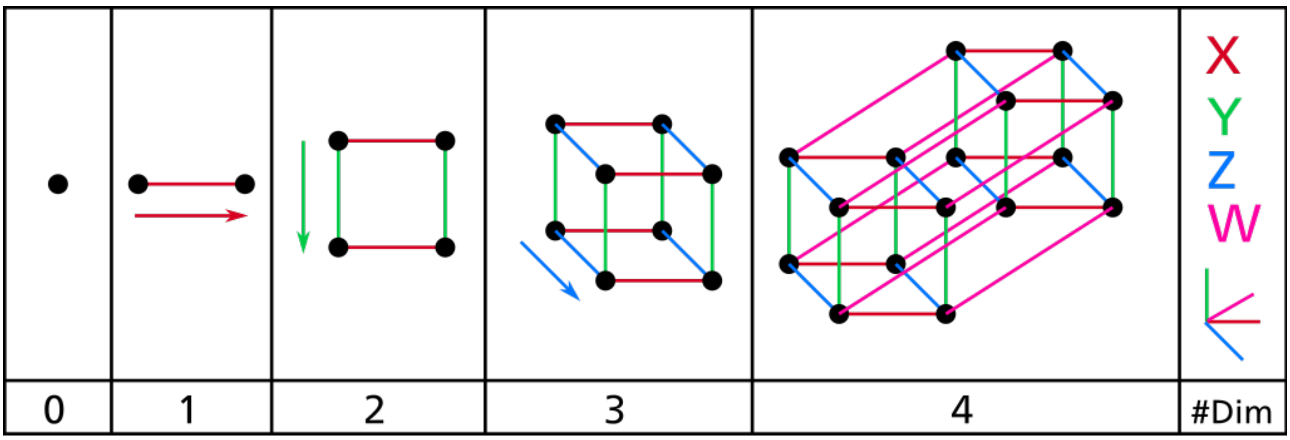
(Eq 1.5.2)

The a_k coefficients of the linear combination are scalar numbers corresponding to the individual projections of the vector:

$$a_1 = \langle a|u_1\rangle, a_2 = \langle a|u_2\rangle, a_3 = \langle a|u_3\rangle, \ldots, a_n = \langle a|u_n\rangle$$

(Eq 1.5.3)

With respect to the given vector space and basis, those coefficients are called the "*coordinates of the vector*" while the single addenda are its "*components*".



(Fig. 1.5.1: various dimension vector spaces)

2.1 Complex numbers - Cartesian form

The "*complex numbers*" satisfy the "*fundamental theorem of algebra*", which states that each polynomial equation:

$$c_n z^n + c_{n-1} z^{n-1} + \ldots + c_2 z^2 + c_1 z + c_0 = 0$$

(Eq 2.1.1)

has at least one complex number as solution. With respect to this property, which does not generally hold for the real numbers, they represent an extension of the latter ones.

To do so, complex numbers need to be written as the addition of a "*real part*" "*a*" and an "*imaginary part*" "*b*" (both of them real numbers):

$$z = a + ib$$

(Eq 2.1.2)

where the coefficient "i", called "*imaginary unit*", is defined in order to have the following property:

$$i^2 = -1$$

(Eq 2.1.3)

which, in a real number context, would have no meaning: this is the reason the name "imaginary" has been given to it.

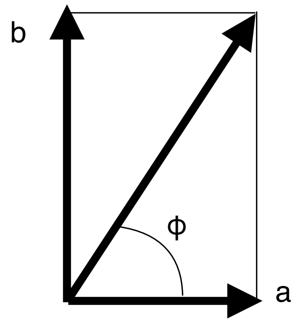
This representation of a complex number is called the "cartesian form".

L'ALGEBRA
OPERA
Di Rafael Bombelli da Bologna
Diuiſa in tre Libri.
Con la quale ciaſcuno da ſe potrà venire in perfetta cognitione della teorica dell'Arimetica.
Con vna Tauola copioſa delle materie, che in eſſa ſi contengono.
Poſta hora in luce à beneficio delli Studioſi di detta profeſſione.
IN BOLOGNA,
Per Giouanni Roſsi. MDLXXIX.
Con licenza de' Superiori

(Fig. 2.1.1: first publication regarding the complex numbers, by Rafael Bombelli, 1526 - 1572)

2.2 Complex numbers - Magnitude and phase

The complex numbers can be represented as 2D-vectors in a cartesian reference having a real axis and an imaginary axis frame:



(Fig 2.2.1)

According to this vector representation, a complex number also has a norm corresponding to the square of its magnitude:

$$|\rho|^2 = a^2 + b^2$$

(Eq 2.2.1)

and a phase so that:

$$\tan(\varphi) = \frac{b}{a}$$

(Eq 2.2.2)

(Fig. 2.2.1: R. Descartes, 1596 - 1650)

2.3 Complex numbers - Complex exponential

The exponential function can be extended to have a complex number argument:

$$e^{z} = e^{y+ix} = e^{y} e^{ix}$$

(Eq 2.3.1)

The second factor can be evaluated by a *Taylor's series* expansion around $x_0 = 0$ as follows:

$$e^{ix} = \sum_{k=0}^{+\infty} \frac{\left.\frac{d^k}{dx^k} e^{ix}\right|_{x=x_0} \left(x - x_0\right)^k}{k!} = \sum_{k=0}^{+\infty} \frac{i^k x^k}{k!} = 1 + ix - \frac{x^2}{2} - \frac{ix^3}{6} + \frac{x^4}{24} + \frac{ix^5}{120} - \ldots$$

(Eq 2.3.2)

The addenda of the Taylor series can be rearranged into two groups:

$$e^{ix} = \left(1 - \frac{x^2}{2} + \frac{x^4}{24} - \ldots\right) + i\left(x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots\right)$$

(Eq 2.3.3)

where the individual Taylor expansion of the "*sin*" and "*cos*" functions can be recognised:

$$e^{ix} = \cos(x) + i\sin(x)$$

(Eq 2.3.4)

The latter result is commonly known as "*the Euler's formula*".

(Fig. 2.3.1: L. Euler, 1707 - 1783)

2.4 Complex numbers - Polar form

The complex numbers can be also represented in terms of complex exponentials, recognising from their cartesian representation that:

$$z = a + ib \Leftrightarrow \begin{cases} a = \rho\cos(\varphi) \\ b = \rho\sin(\varphi) \\ \rho = \sqrt{a^2 + b^2} \\ \tan(\varphi) = \dfrac{b}{a} \end{cases}$$

(Eq 2.4.1)

Hence:

$$z = a + ib = \rho\left[\cos(\varphi) + i\sin(\varphi)\right] = \rho e^{i\varphi}$$

(Eq 2.4.2)

(Fig. 2.4.1: the hands of a clock as an example of the polar form representation of vectors)

2.5 Complex numbers - Complex conjugate

Two complex numbers are defined to be "*conjugated*" each other when they have the same real and opposite imaginary part. In the cartesian representation that is:

$$z = a + ib \Leftrightarrow z^* = a - ib$$

(Eq 2.5.1)

while in the polar representation is:

$$z = \rho e^{i\varphi} \Leftrightarrow z^* = \rho e^{-i\varphi}$$

(Eq 2.5.2)

This way the norm of a complex number can be also represented as:

$$|z|^2 = z^* z$$

(Eq 2.5.3)

Recalling the vector notation, for complex numbers vector entries this becomes:

$$|z|^2 = \langle z^* | z \rangle$$

(Eq 2.5.4)

and more generally:

$$\langle z_a{}^* | z_b \rangle = |z_a||z_b|\cos(\varphi)$$

(Eq 2.5.5)

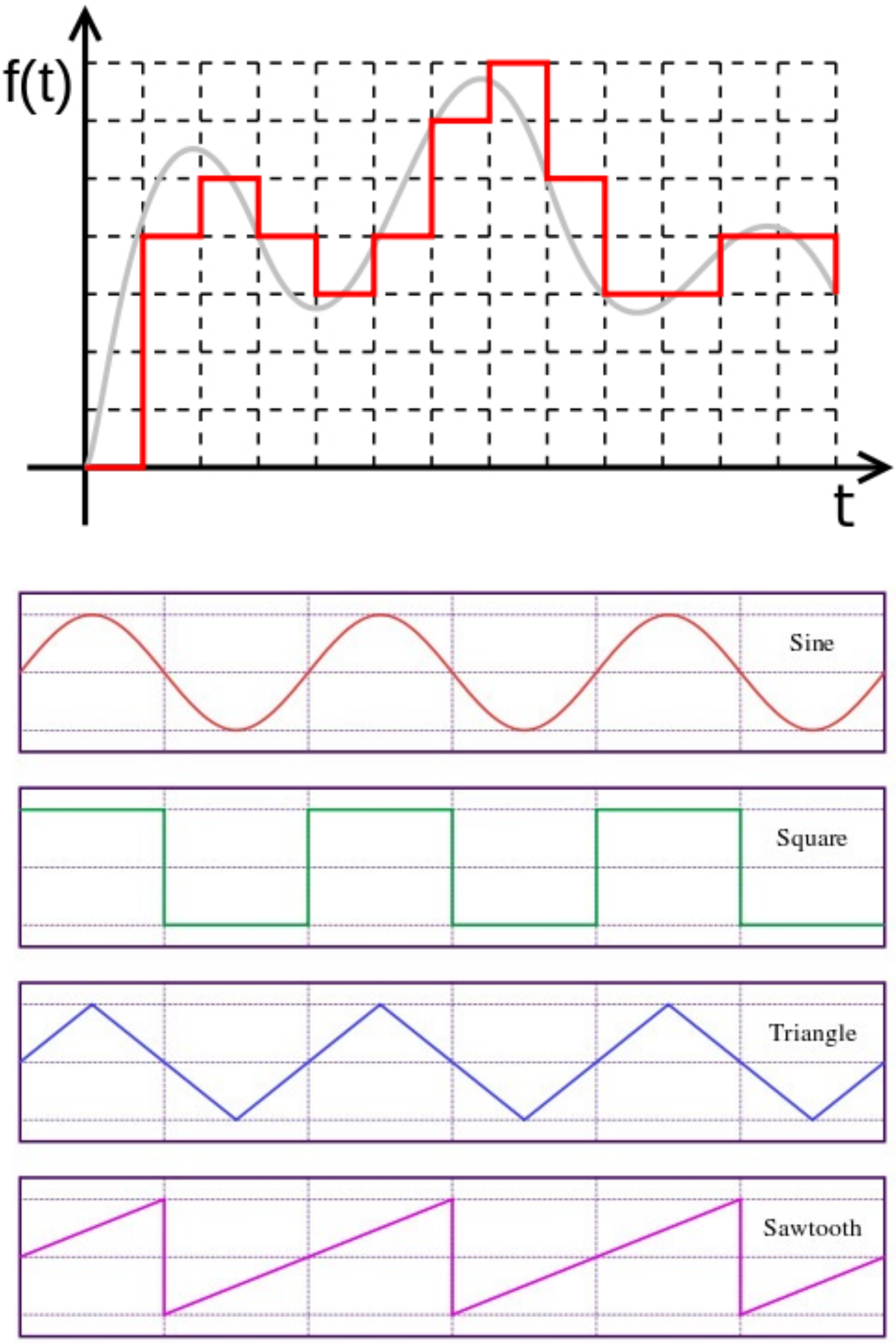
represents the inner product between any two vectors having complex entries. Notice that even in a N-D vector space there is always a 2-D plane where two N-D vector can lie and on which it is possible to consider an angle between them.

(Fig. 2.5.1: Narciso, by Caravaggio, 1571 - 1610)

3.1 Signals - Analog and digital signals

In electronics analog signals are "*continuous*", "*real single-valued*", "*time-varying*" signals. Digital signals are a representation of analog signals after a sampling and discretisation process.

EXAMPLES:



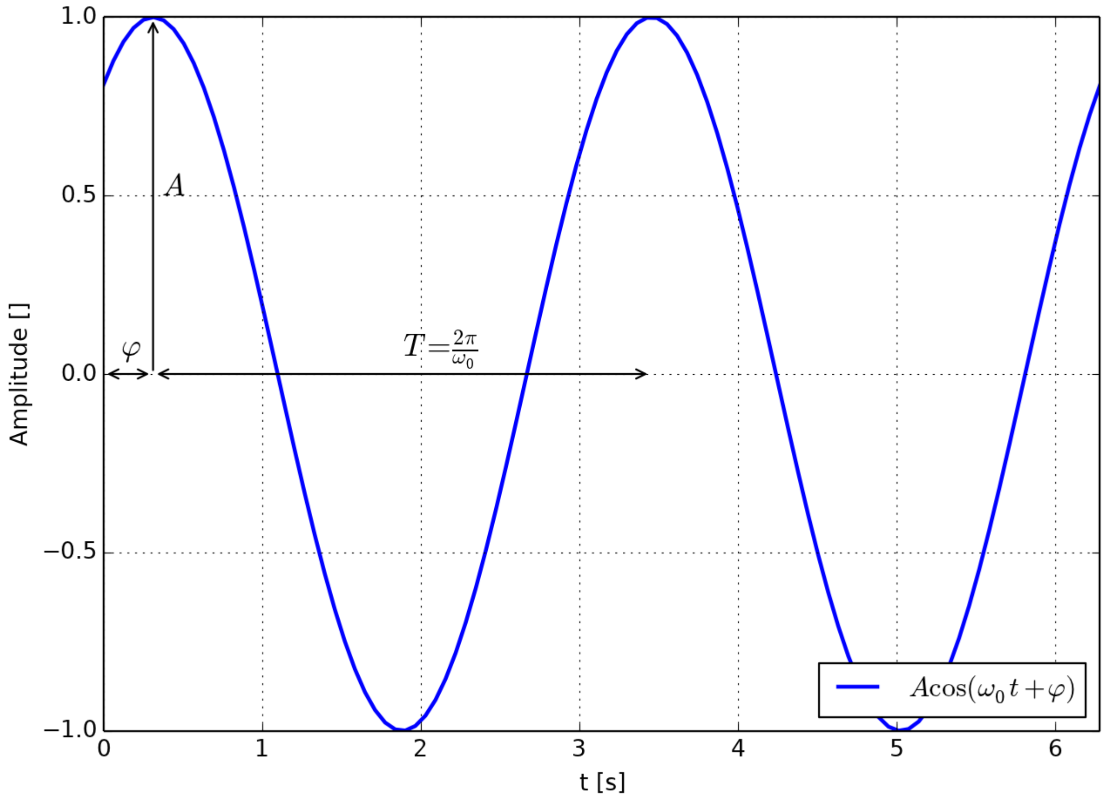
(Fig 3.1.1: various type of signals)

3.2 Signals - Sinusoidal signals

The sinusoidal signal is probably the most important type of signal in electronics. In its most plain form it can be represented as:

$$f(t) = A\cos(\omega_0 t + \varphi)$$

(Eq 3.2.1)



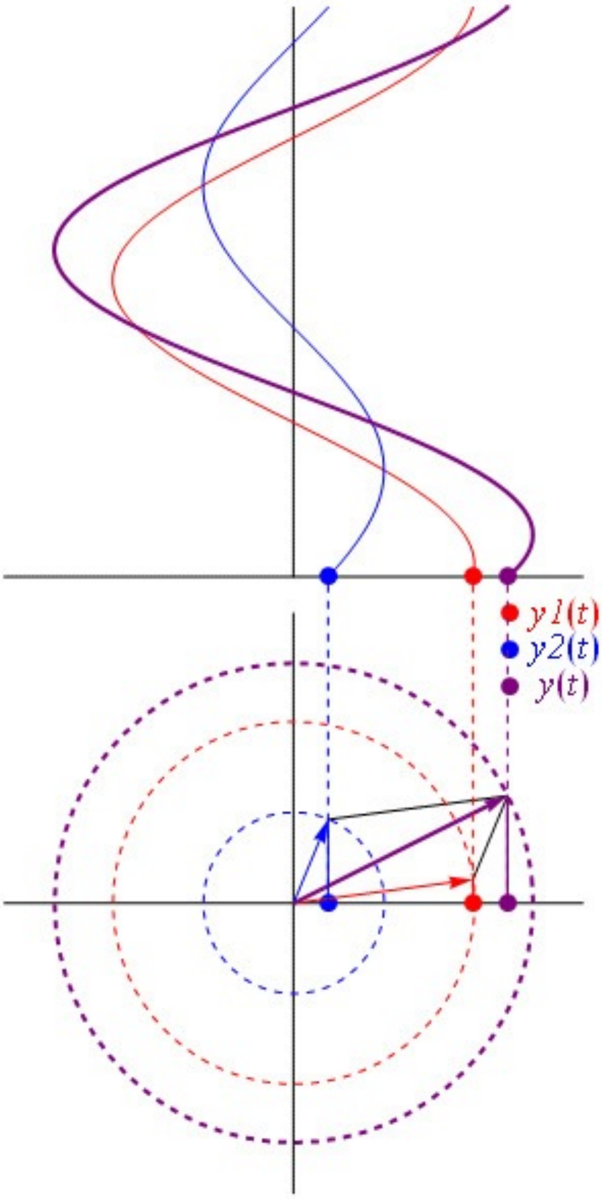
(Fig 3.2.1: generic sinusoidal signal)

3.3 Signals - Phasors

A sinusoidal signal can be also represented by a complex number in its polar form:

$$f(t) = \mathrm{Re}\left[Ae^{i(\omega_0 t + \varphi)} \right]$$

(Eq 3.3.1)



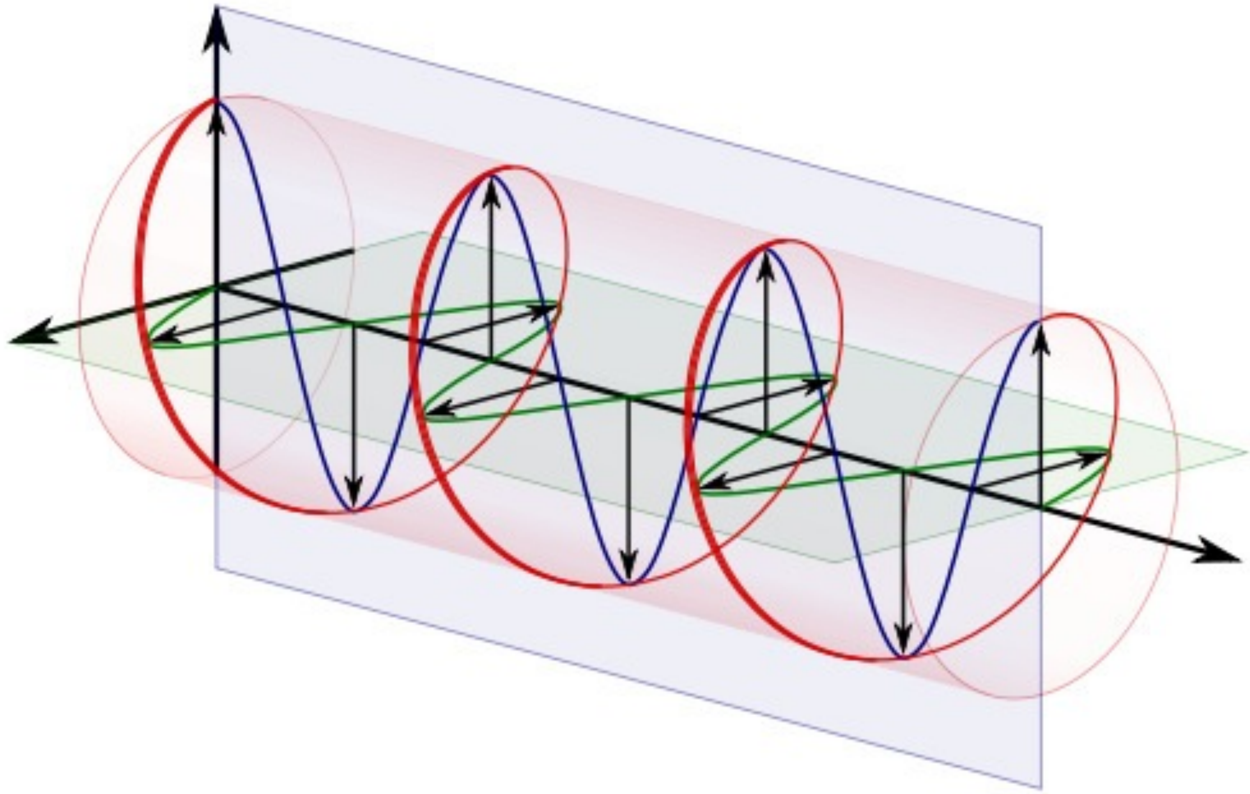
(Fig 3.3.1: phasors)

In this picture, phasors are vectors rotating *counter clockwise* if the frequency is a positive value.

Notice that *negative frequencies* also have a meaning: they are represented by phasor vectors rotating clockwise.

3.4 Signals - Complex exponential signals

A more complete visualisation of a sinusoidal signal represented by a phasor is the following graph:



(Fig 3.4.1: complex signal)

Notice the sinusoidal and co-sinusoidal projections onto the real and imaginary axes.

3.5 Signals - Non-sinusoidal signals

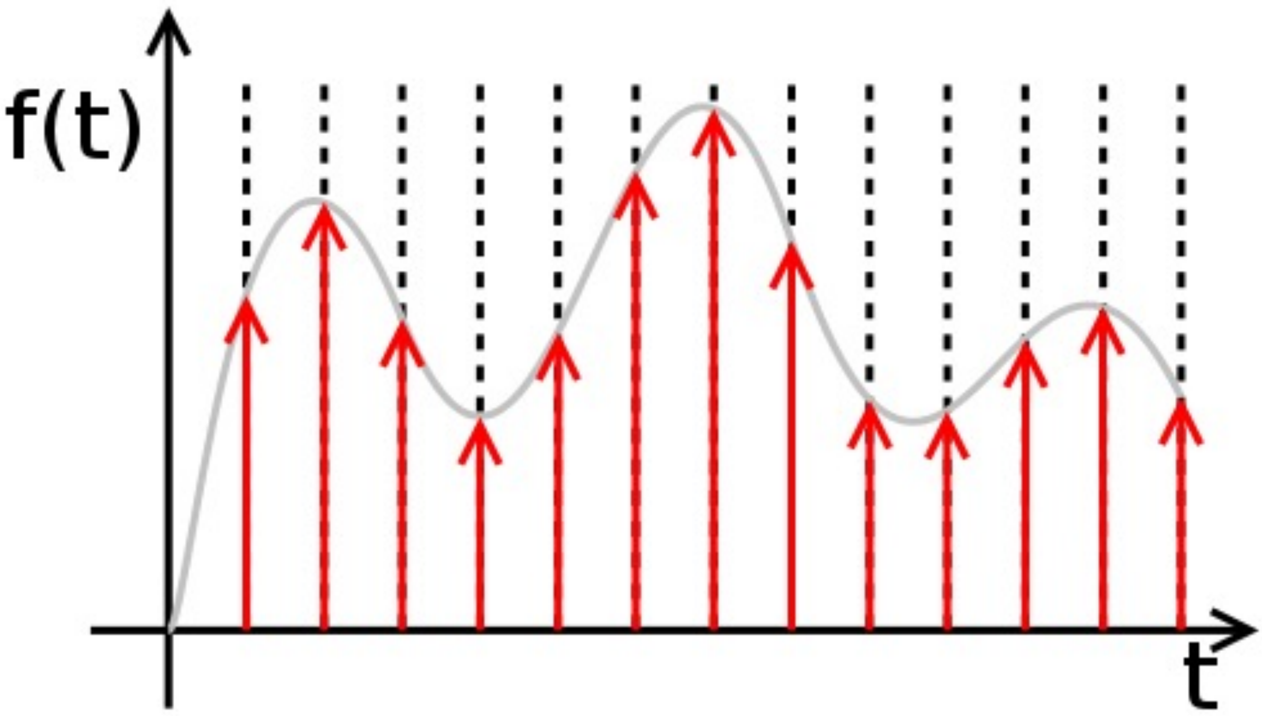
Non-sinusoidal signals can be represented, under certain hypothesis, as a summation of many sinusoidal signals, each of them having a certain magnitude, phase and frequency.

This method is called "Fourier's signal analysis".

(Fig 3.5.1: J.B.J. Fourier, 1768 - 1830)

4.1 Fourier signal analysis - Signals as vectors

In the digital approximation an analog signal can be thought as a vector made of elements having values corresponding to the samples and dimension corresponding to the total number of samples (possibly also infinitive):



(Fig 4.1.1: time sampled signal)

$$f(t) = \left| f \right\rangle = \left| .., s_{-2\Delta t}, s_{-\Delta t}, s_0, s_{\Delta t}, s_{2\Delta t}, ... \right\rangle$$

(Eq 4.1.1)

meaning that:

$$s_{-2\Delta t} = f(-2\Delta t), s_{-\Delta t} = f(-\Delta t), s_0 = f(0), ...etc.$$

(Eq 4.1.2)

4.2 Fourier signal analysis - Continuous representation limit

As an extension to a vector of complex numbers, considering its vector representation, the norm of the function is:

$$\left|f(t)\right|^2 = \left\langle f(t)^* \middle| f(t) \right\rangle$$

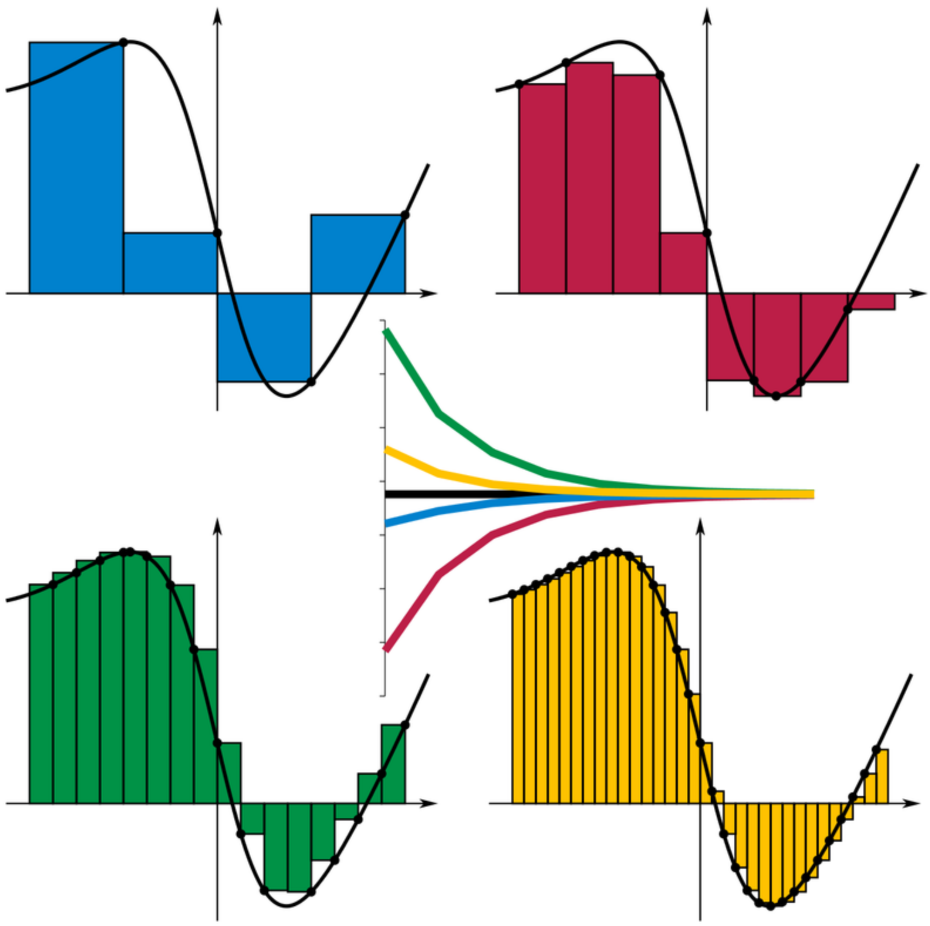
(Eq 4.2.1)

Ideally, with an infinitesimal sampling time, the discrete vector representation of a function becomes continuous.

Then the inner product summation becomes then an integral:

$$\left\langle f(t)^* \middle| f(t) \right\rangle = \sum_{k=-\infty}^{k=+\infty} s_{k\Delta t} * s_{k\Delta t} \Delta t \rightarrow \int_{-\infty}^{+\infty} f(t)^* f(t) dt$$

(Eq 4.2.2)



(Fig. 4.2.1: continuos limit representation of a summation as an integral)

4.3 Fourier signal analysis - Orthonormal function basis

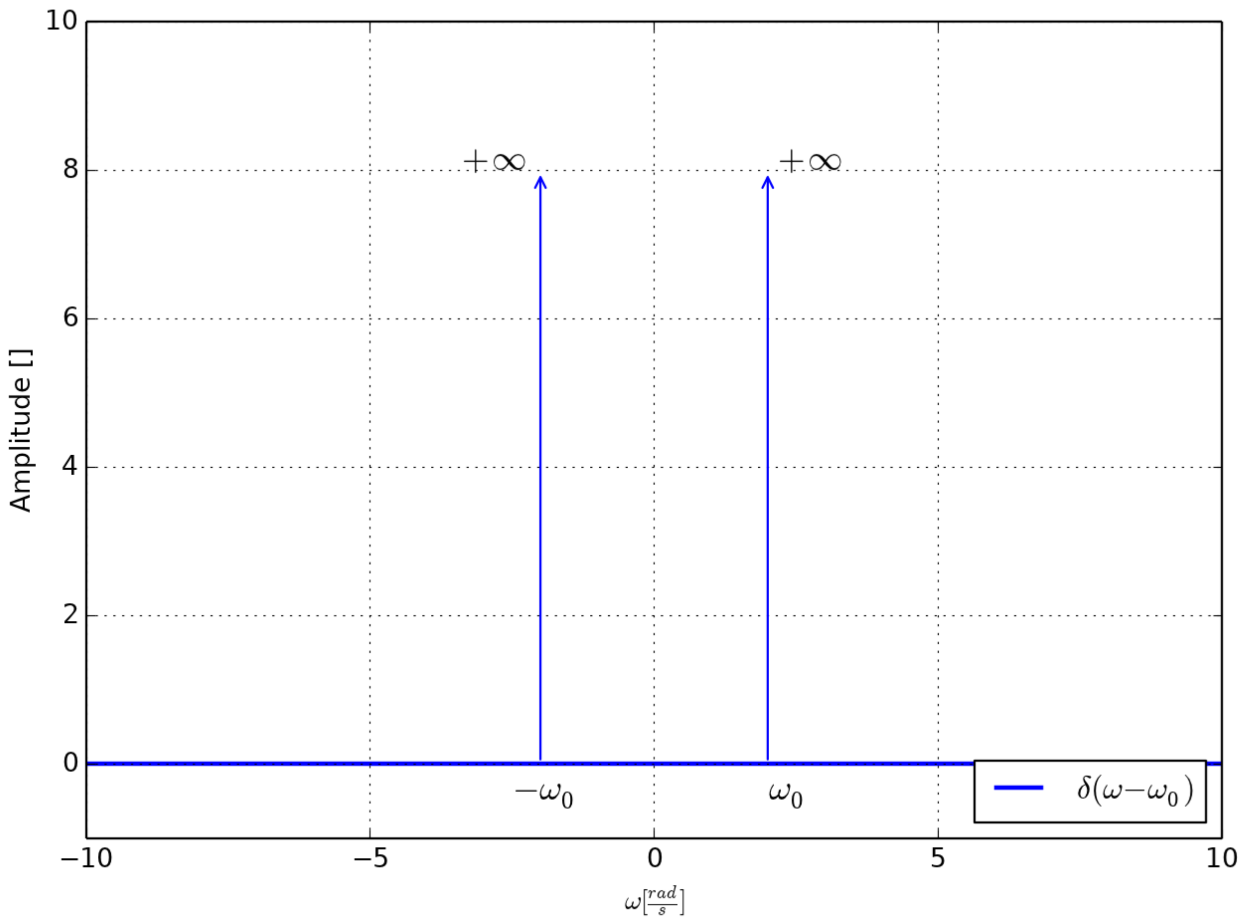
In order to define the reference frame of such vector space one convenient possibility is to choose a set of functions corresponding to orthonormal vectors.

As a generalisation of a vector base, complex exponential functions characterised by having different frequencies are orthogonal each other:

$$\left\langle e^{-i\omega_0 t} \middle| e^{i\omega_0 t} \right\rangle \sqrt{2\pi}\delta\left(\omega - \omega_0\right)$$

(Eq 4.3.1)

where:



(Fig 4.3.1: the Dirac's delta distribution representation)

is the *Dirac's delta distribution*. It is not, strictly mathematically speaking, a function.

The Dirac's delta distribution is a generalisation of the Kronecker's delta function.

4.4 Fourier signal analysis - Fourier's transform

The "*Fourier's transform*" a function f(t) corresponds to the projection of the function onto the orthonormal basis made of sinusoidal functions:

$$\left\langle f(t)^{*} \middle| e^{i\omega t} \right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)^{*}\, e^{i\omega t}\, dt = F(\omega)$$

(Eq 4.4.1)

In fact the formula above represents a number of projections, one for each value ω which identifies the frequency axis onto which the projection occurs.

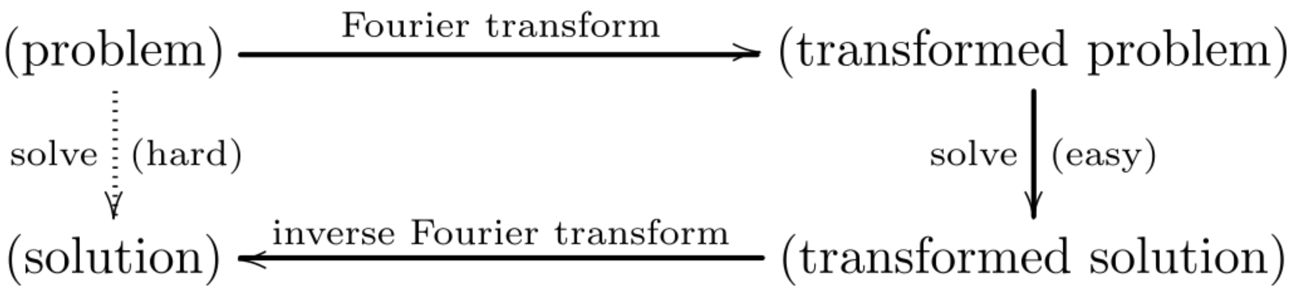
Notice that a physical signal is always a real-valued function, hence its complex conjugate is equal to the signal itself, but its Fourier transform is generally complex-valued.

According to this, the signal can be represented as the summation (integration) of all its sinusoidal frequency components:

$$\left\langle e^{i\omega t} \middle| F(\omega) \right\rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(\omega)\, e^{-i\omega t}\, d\omega = f(t)$$

(Eq 4.4.2)

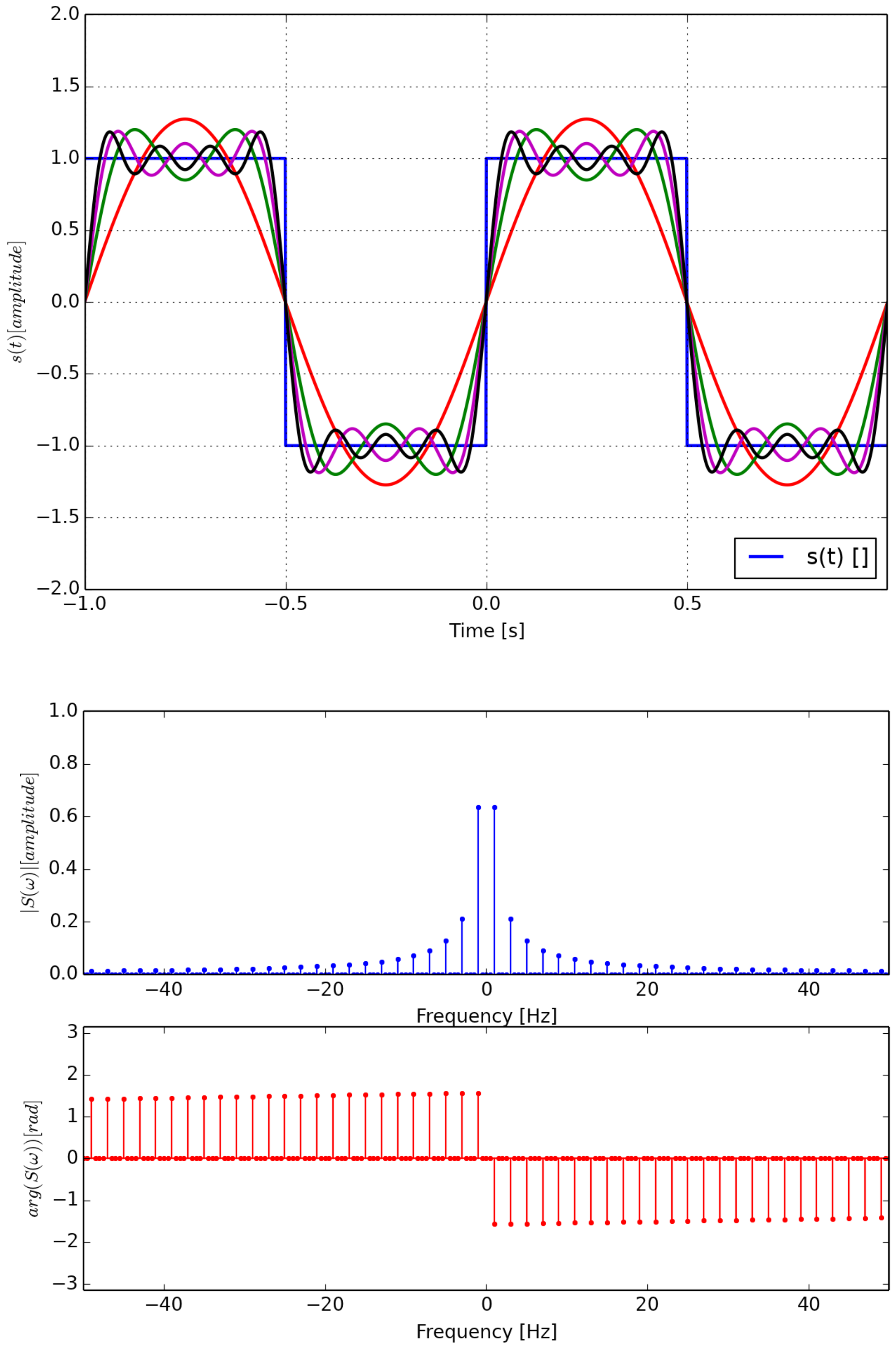
The latter equation corresponds to the "*inverse Fourier's transform*".



(Fig. 4.2.1: Fourier's transform paradigm)

4.5 Fourier signal analysis - Frequency and phase spectra

The Fourier's transform of a signal is a complex-valued function. The chart of its magnitude as a function of the frequency is called the "*amplitude spectrum*", while the one relative to its phase is the "*phase spectrum*".



(Fig 4.5.1: amplitude and phase spectra of a square wave)

5.1 Electromagnetism - Atoms

Atoms are small structures composed of a central nucleus, made of protons and neutrons, surrounded by a cloud of electrons described by a "*kinetic energy operator K*" and bounded to it due to its "*potential energy V*", both combined together to form the "*Hamiltonian operator H*":

$$H = K + V$$

(Eq 5.1.1)

The electrons are distributed according to their *wave function*, which describes the probability of finding an electron at a given location and time according to the "*Schroedinger's time-dependent equation*":

$$i\hbar \frac{\partial}{\partial t}\psi = H\psi$$

(Eq 5.1.2)

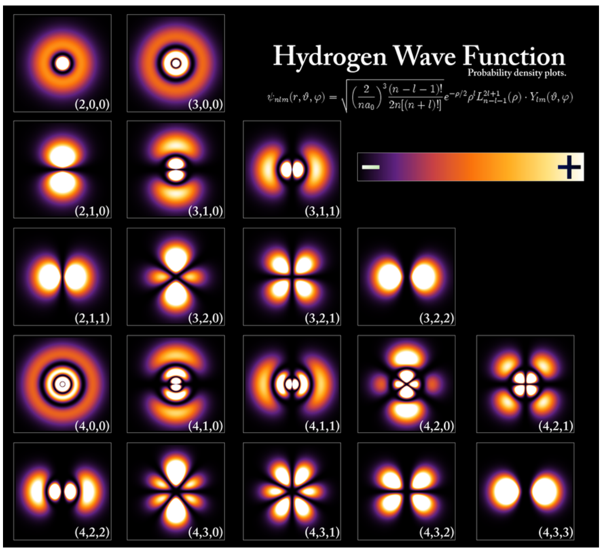
The boundary conditions set by the potential energy due the nucleus make the wave functions of the surrounding electrons having the form of "*standing waves*":

$$E\psi = H\psi$$

(Eq 5.1.3)

The solutions of this equation are called "*eigen-states of H*", and are represented by complex-valued functions ψ called "*orbitals*" associated to discrete - *quantised* - real-valued "*energy levels E*".

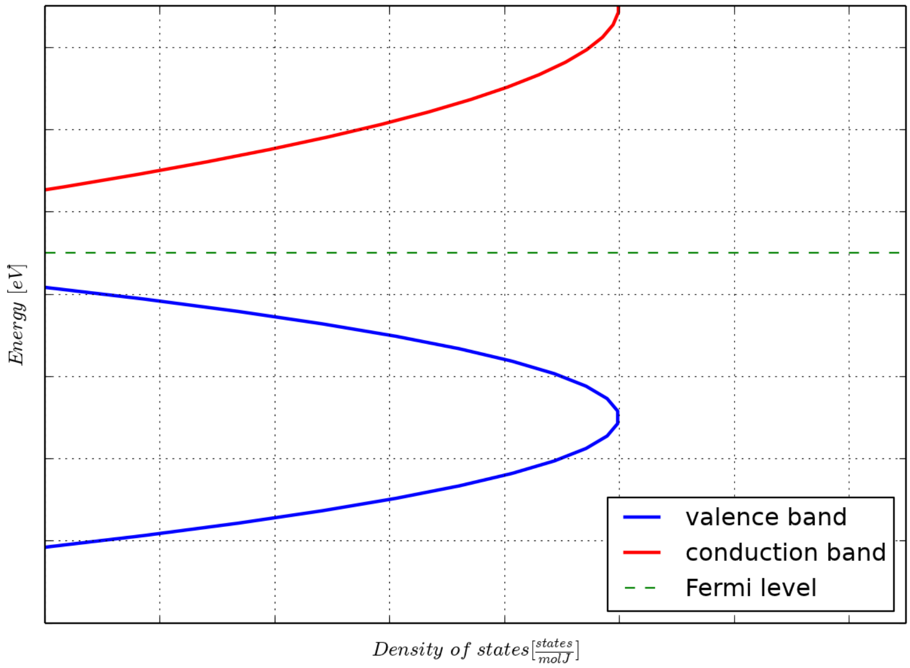
The state of an electron is a time-variable combination of its eigen-states, but when "*observed*" its state immediately changes to the closest eigen-state at the moment of the osbervation.



(Fig. 5.1.1: hydrogen atom eigenstates)

5.2 Electromagnetism - Band gap

Electrons in solid materials exhibit different structures of states according to the type of nuclei and nuclear arrangements they are bonded to:



(Fig 5.2.1: bandgap)

Some electrons, the ones in the valence band, always stay close to the nuclei while others may be able to freely move through the material when they have enough energy to be in the conduction band. To do so they have to have enough energy to cross a *band gap*, which depends on the properties of the nuclei composing the solid material in which they are. Materials having a very high band gap are called "insulators", as opposed to materials having a very little or no band gap, which are called "*conductors*". Materials having "*not so small but not so high*" band gaps are called "*semiconductors*". The value of energy "E_f" is the so called "*Fermi level*", also called "*total electrochemical potential*", which is a property of the body considered alone (atom, molecule, cell membrane, piece of material, etc.) and corresponding to the amount of energy which is necessary to add or remove an electron from that body. Notice that when attaching the two probes of a voltmeter to two different places A and B of a body (or a series chain of bodies) the measured voltage exactly corresponds, by definition of voltage, to the difference of the Fermi levels of the two places divided by the elementary charge of one electron (q_e = -1.602*10^-19 C):

$$\Delta V = V_A - V_B = -\frac{E_A - E_B}{q_e}$$

(Eq. 5.2.1)

When in contact, different bodies change their original Fermi level in order to establish a new energy equilibrium. **In neuronal cell membranes, perturbations of the electrochemical boundary conditions is the key point for understanding the nature of the "*action potential*".**

5.3 Electromagnetism - Electric field

The nuclei and the electrons are bonded together forming solid materials because of an attractive force among them. In the simplest, not fully exhaustive, picture the nuclei exert an attractive force on the electrons, which are subjected to.

(Fig 5.3.1: Coulomb's force)

This force which binds one electron to one nucleus can be described in terms of a point-like particle having *charge* "*q*" being subjected to the influence a an *electric field* "*E*" generated by the other one, according to the *Coulomb's law*:

$$F = qE$$

(Eq 5.3.1)

where "*E*" is the electric field generated by the nucleus:

$$E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}$$

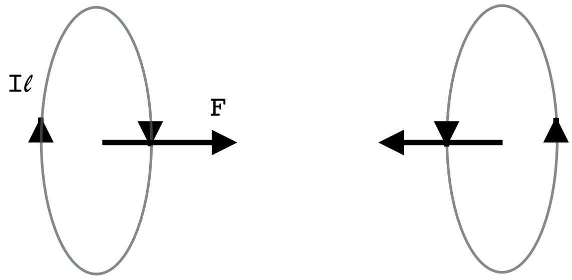
(Eq 5.3.2)

The charge "Q" is the charge of the nucleus, "r" is the distance between the two particles and "ε_0" is the *dielectric constant of the vacuum*.

NOTE: do not make confusion between the *electric field* and the *energy*, since they usually have the same letter *E*.

5.4 Electromagnetism - Magnetic field

A continuous flux of free electrons somehow moving along a closed path having length "l", possibly inside a conductor, makes another similar circuit being attracted and aligned to its axis:



(Fig 5.4.1: Laplace's force)

This interaction can be mathematically described by a current loop subjected to a force due to a *magnetic field* "B" generated by the other one. In case of a circular loop having radius "r", it is possible to write the *Laplace's force*:

$$F = Il \times B = 2\pi r I \times B$$

(Eq 5.4.1)

Notice the "outer product":

$$\mathrm{R} \times \mathrm{P} = |R\rangle\langle P| - |P\rangle\langle R|$$

(Eq 5.4.2)

The expression of the magnetic field "B" takes a simple form when the loop generating it is a circle and when it is evaluated at its centre:

$$B = \frac{\mu_0}{2}\frac{I}{r}$$

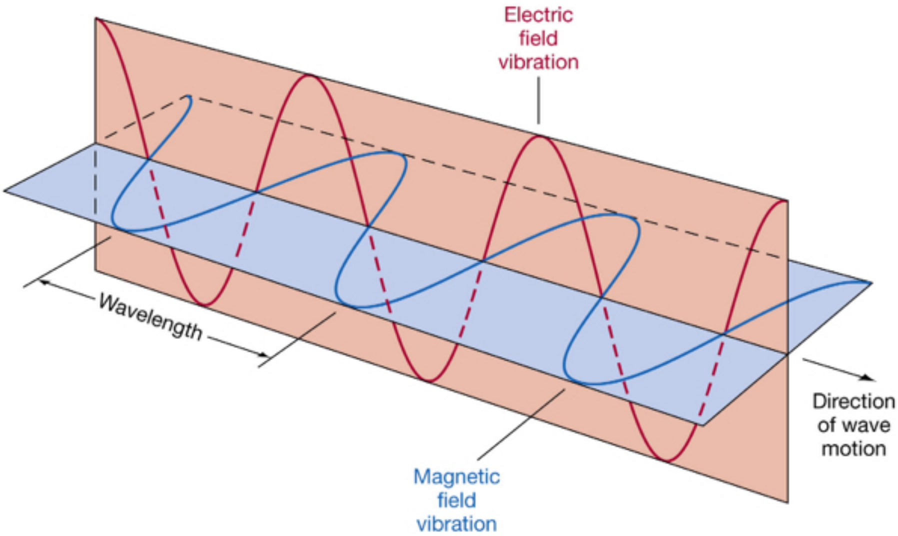
(Eq. 5.4.3)

where "I" is the current, as a quantity describing the flux of moving charge, "r" is the radius of the loop and "μ_0" is a constant called "*magnetic permeability of vacuum*". Those simplifications makes

the Laplace's force to be valid, in this case, only if used to computed the force of the adjacent loops. The analysis of this case turns useful to understand that in the case of two adjacent loops having infinitesimal radius the force between them can still be not null (because "r" cancels you in the outer product). **This suggests a concept of *infinitesimal current loop* to be the magnetic analogous of the *point-like charge*.** However, permanent magnets does not have any current loop being responsible of their magnetic properties: their magnetism is in fact due to a the alignment of the "*spin*" of some of their electrons. The spin, historically named as suggesting a kind of rotary motion, is in fact a sort of intrinsic "*angular momentum*" particles may have which can explained by the "*Einstein's relativity*" laws. It is as linked to the relativistic analysis of the "*rotations*" as like as the concept of "*rest mass*" is to the relativistic analysis of the "*translations*".

5.5 Electromagnetism - Electromagnetic waves

When varying in time, the electric and magnetic fields are not separate but are always associated either with *travelling* or *stationary electromagnetic waves* (EM waves) describing either the propagation or the storage of energy in space or materials:



(Fig 5.5.1: representation of a electromagnetic plane wave)

The simplest type of EM wave is the plane wave in the free space, where the electric and the magnetic fields are orthogonal each other and have the following representation:

$$E = E_{\max} e^{i(kx-\omega t)},\ B = \frac{E}{c},\ c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

(Eq 5.5.1)

where k is $2\pi/\lambda$, λ is the *wavelength*, ω is its *frequency* and *c* its *speed of propagation* which, in *free space*, corresponds to the *speed of light*.

Notice the link existing between the concept of speed of light as coming from the electromagnetic framework and the concepts of dielectric constant and the magnetic permeability separately rising from, respectively, the electric and the magnetic field theory.

6.1 Basic electronic components - Battery

The electric field can be indirectly measured by knowing its voltage potential ϕ:

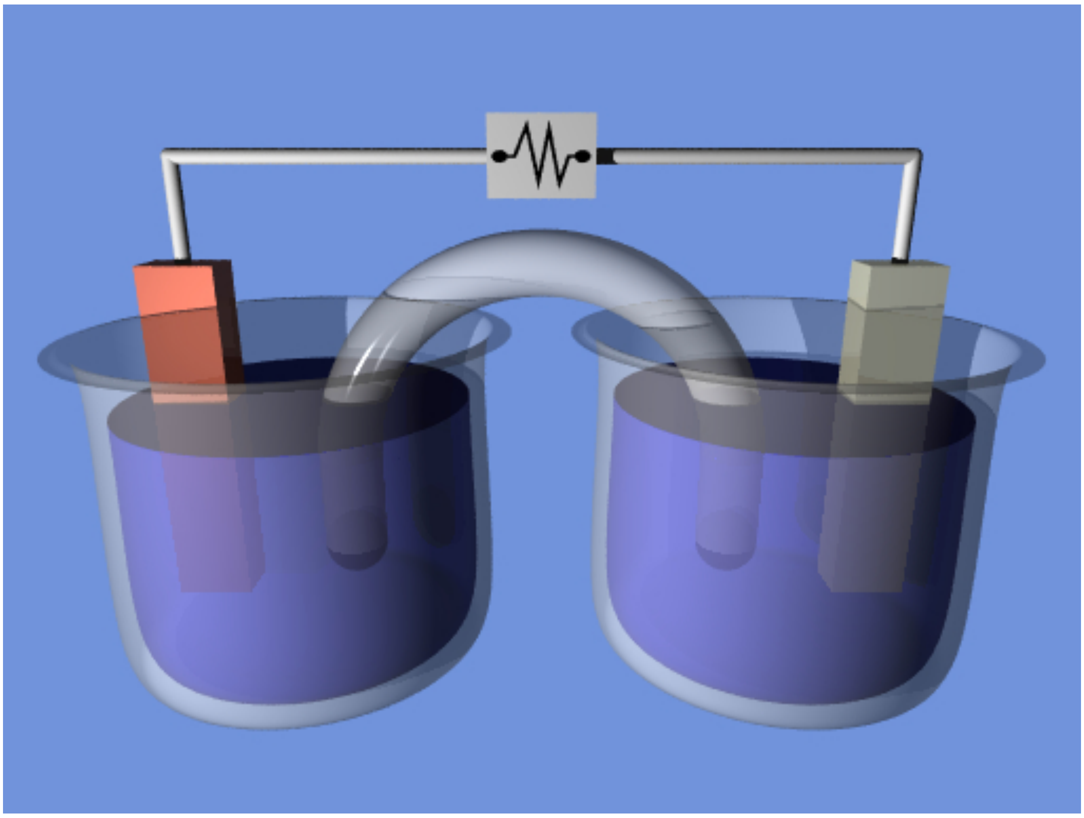
$$E = -\nabla\varphi \overset{1D}{\rightarrow} = -\frac{d\varphi}{dx}$$

(Eq 5.4.1)

The battery is a source of constant voltage *ΔV*, being the difference between the *redox potentials* of the half chemical reactions occurring at the cathode and anode:

$$\Delta V = \varphi_{cathode} - \varphi_{anode}$$

(Eq 5.4.2)



(Fig 5.4.1: the battery)

When the circuit is closed, electrical current can flow as the result of an electric field inside the battery which is sustained by the chemical reactions in it.

Notice that, by definition, the cathode is the electrode from which the current leaves a polarised electrical device, while the anode is the electrode from which the current flows into. According to this definition, the cathode and the anode can be either positive and negative charged or vice versa: in a discharging battery the cathode is the positive terminal and the anode is the negative, while in a device which consumes power its the opposite.

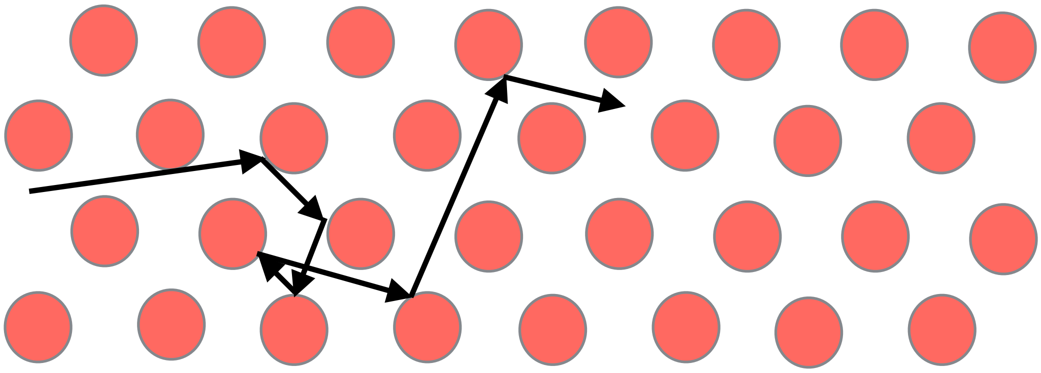
6.2 Basic electronic components - Ohmic conductor

Applying a constant voltage difference ΔV to a piece of conductive material having length Δx, the resulting electric field inside the conductor is:

$$E = -\frac{\Delta V}{\Delta x}$$

(Eq 5.5.1)

This makes the free electrons in the conductor material moving pushed by the Coulomb force (Eq 5.3.1). According to the *Drude-Lorentz model*, electrons are supposed to be constantly accelerated by that force but also continuously stopped and re-accelerated by hitting the nuclei they meet along their free path like in a pinball game, hence reaching a steady velocity:



(Fig 5.5.1: Drude & Lorentz's "pinball" model)

This model gives a first, classical, interpretation of the *current density j* inside a conductor:

$$j = \left(\frac{nq^2\tau}{m} \right) E$$

(Eq 5.5.2)

where *n, q, τ* and *m* are respectively the electron's density, charge, free path and mass.

6.3 Basic electronics components - Resistor

According to the Drude-Lorentz's model the current density *j* flowing inside an ohmic conductor having length l is proportional (here by a constant $1/\varrho$) to the applied electric field *E*:

$$j = \frac{1}{\rho} E = \frac{1}{\rho} \frac{V}{l}$$

(Eq 6.3.1)

The current is hence the flux of the current density *j* calculated through the cross section *S* of the conductor:

$$I = jS = \frac{S}{\rho} \frac{V}{l} \Leftrightarrow R = \frac{V}{I} = \rho \frac{l}{S}$$

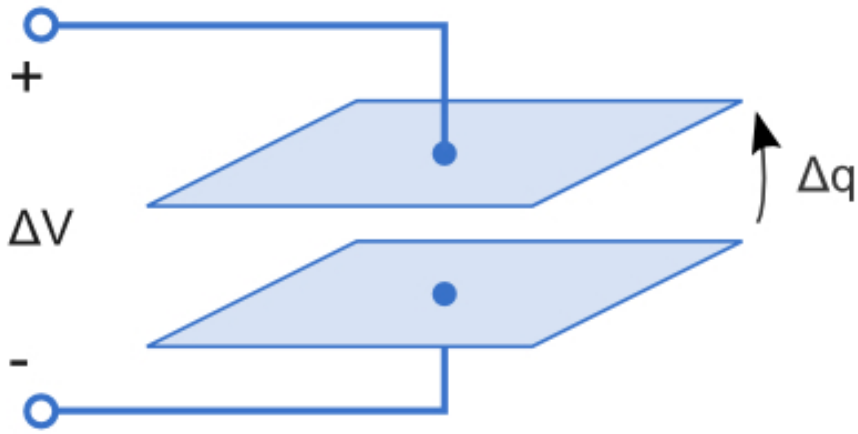
(Eq 6.3.2)

This result is known as the *Ohm's law*, where *R* is the *resistance* of the conductor and ϱ is the *resistivity*, a property that is connected to the nature of the material of which the conductor is made of.

(Fig 6.3.1: a commercial axial resistor)

6.4 Basic electronic components - Capacitor

Applying battery having a voltage *V* across two parallel conductive plates having a surface *S* and facing each other at a distance d causes the electrons deployed at the anode being attracted by the vacancy of electrons depleted from the cathode:



(Fig 6.4.1: capacitor)

The phenomenon is facilitated by making the plates more closed each other, because this way the attraction is stronger, and also by increasing the surface, because by doing so there is more place for the charge. This ability is called *capacitance C*:

$$Q = CV = \varepsilon_0 \frac{S}{d} V$$

(Eq 6.4.1)

The capacitor is a storage of energy in form of electrical field:

$$E = \frac{1}{2} CV^2$$

(Eq 6.4.2)

6.5 Basic electronic components - Inductor

It has been observed that a variable current applied to coil made of a good conductor material leads to have a certain time-dependent voltage across it which is much bigger than the one predicted by the Ohm law:

(Fig. 6.5.1: inductor)

$$V = L\frac{dI}{dt}$$

(Eq 6.5.1)

This result is known as the *Faraday's law of induction*, and the coefficient of proportionality *L* is called the *inductance* of the circuit, which depends on its area and shape. For a long and thin coil it can be calculate as:

(Eq 6.5.2)

$$L = \mu_0 \frac{N^2 A}{l}$$

The inductor is a storage of energy in form of magnetic field:

$$E = \frac{1}{2} L I^2$$

(Eq 6.5.3)

More precisely speaking, the voltage *V* across an inductor is due to the variation of a magnetic field coupled to it. This magnetic field variation can be generated either by a separate second external coil driven by a time-variable current or by the first coil itself. In the first case the voltage is due to the interaction of the magnetic field generated by the coil with itself, hence *L* is properly called *auto inductance*. In the second case the magneto-electric circuit is made of two coils, one exchanging energy with the other one: in this case *L* is called *mutual inductance*.

7.1 From physics to electronics - Datasheets

Besides the basic electronics components, the real components used in everyday's life are much more sophisticated and need more information to be described.
This information is written and made available by the manufacturer in form of "datasheets":

STP3NK90Z - ST
STD3NK90Z - ST

N-CHANNEL 900V - 4.1Ω - 3A TO-220/TO-220
Zener-Protected SuperMESH™P

TYPE	V_{DSS}	$R_{DS(on)}$	I_D	Pw
STP3NK90Z	900 V	< 4.8 Ω	3 A	90 W
STP3NK90ZFP	900 V	< 4.8 Ω	3 A	25 W
STD3NK90Z	900 V	< 4.8 Ω	3 A	90 W
STD3NK90Z-1	900 V	< 4.8 Ω	3 A	90 W

- TYPICAL R_{DS}(on) = 4.1 Ω
- EXTREMELY HIGH dv/dt CAPABILITY
- 100% AVALANCHE TESTED

TO-220

Characteristics	Symbol	Rating	Unit
Collector-base voltage	V_{CBO}	230	V
Collector-emitter voltage	V_{CEO}	230	V
Emitter-base voltage	V_{EBO}	5	V
Collector current	I_C	15	A
Base current	I_B	1.5	A
Collector power dissipation (Tc = 25°C)	P_C	150	W
Junction temperature	T_j	150	°C
Storage temperature range	T_{stg}	−55 to 150	°C

(Fig 7.1.1: example of datasheet)

An important remark: **DO NOT INFER ANY PROPERTY WHICH IS NOT SPECIFICALLY WRITTEN IN THE DATASHEET!**

7.2 From physics to electronics. - SPICE models

Furthermore, nowadays almost every component comes from the manufacturer with a precise mathematical model in form of a text file written in standard language called SPICE - Simulation Program with Integrated Circuit Emphasis.

An example of a SPICE model code for a NMOS-transistor:

```
.MODEL CMOSN NMOS LEVEL=3 PHI=0.600000 TOX=2.1200E-08 XJ=0.20000

+TPG=1 VTO=0.7860 DELTA=6.9670E-01 LD=1.6470E-07 KP=9.6379E-05

+UO=591.7 THETA=8.1220E-02 RSH=8.5450E+01 GAMMA=0.5863

+NSUB=2.7470E+16 NFS=1.98E+12 VMAX=1.7330E+05 ETA=4.3680E-02

+KAPPA=1.3960E-01 CGDO=4.0241E-10 CGSO=4.0241E-10

+CGBO=3.6144E-10 CJ=3.8541E-04 MJ=1.1854 CJSW=1.3940E-10

+MJSW=0.125195 PB=0.800000
```

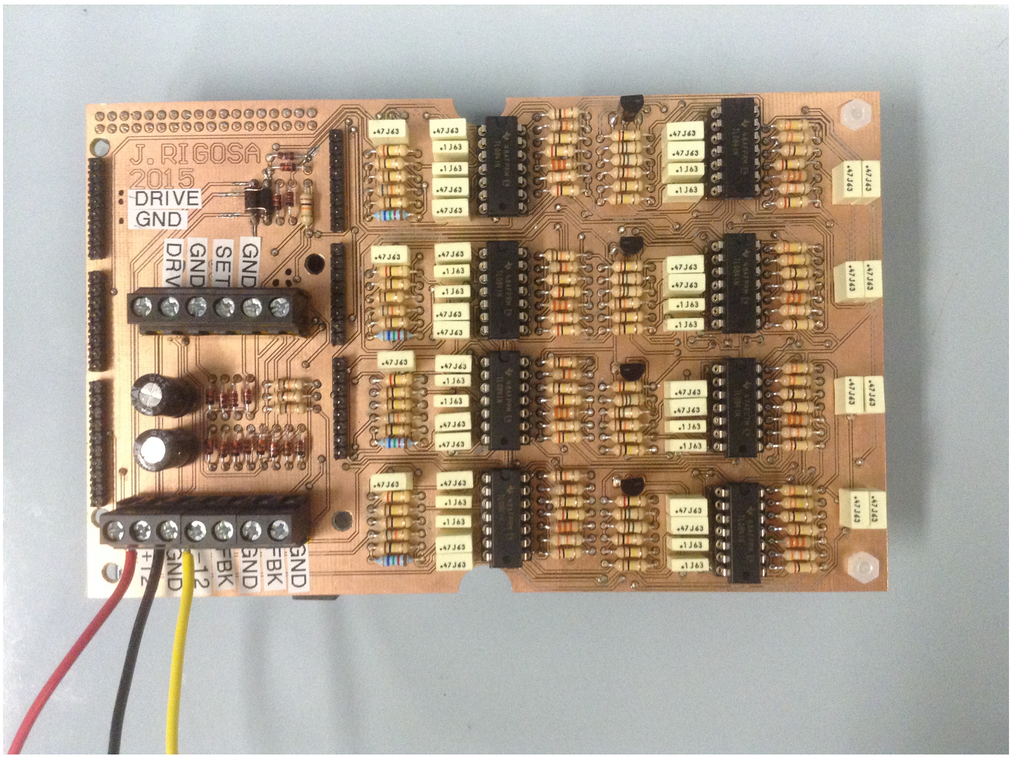
(Fig 7.2.1: example of SPICE code)

7.3 From physics to electronics - CAD and simulators

Besides simple circuits, whose behaviour can be calculated by analytically and even by hand, all the rest have so many interactions which makes computation so extensive to be practically viable only by means of a computer. Circuits are represented by standard symbols identifying each component and drawn with a CAD - Computer Aided Design system, which is also able to simulate them:

(Fig. 7.3.1: a CAD workstation)

The final result of the process of making a circuit is usually a PCB - Printed Circuit Board.



(Fig 7.3.2: a Printed Circuit Board)

7.4 From physics to electronics - Workflow

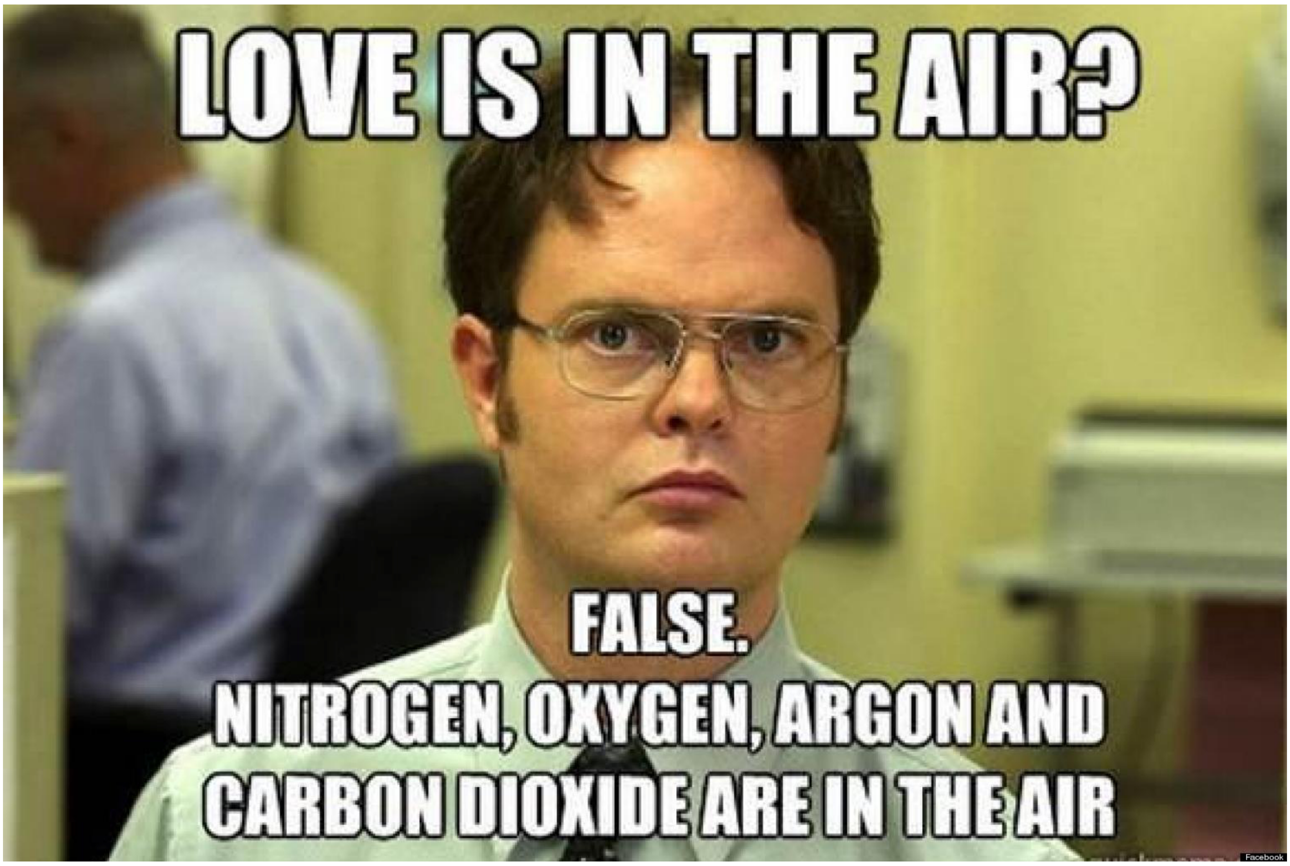
Electronics is made of a theoretical background, knowledge about the materials and components and about instruments and tools.

All this will transform circuit drawings into a real working devices operating the desired functions.

In order to be able achieve this result it is recommended to follow a workflow:

1. Define what you want → SPECIFICATION
2. Abstract the properties into functional blocks → SYNTHESIS
3. Build the blocks → IMPLEMENTATION
4. Test each block → VALIDATION
5. Take tracks of what you did → DOCUMENTATION

and always: **DESIGN FOR TESTABILITY, WHICH MEANS SELECT THE TOOLS WHICH ARE APPROPRIATE TO THE INVESTIGATION OF YOUR QUESTIONS!**



(Fig. 7.4.1: common fallacy of the design-for-testability paradigm due to the usage of tools which are inappropriate to the question-under-investigation)

Eventually, all this stuff will help you during the debugging!

7.5 From physics to electronics - Instruments, tools and materials

Common things that you need in order to start doing some electronics are:

INSTRUMENTS:

1. a multimeter
2. a variable regulated power supply
3. an oscilloscope
4. a PC with some free PCB CAD

TOOLS:

1. a soldering iron
2. a solder sucker
3. a pair of tweezers
4. a cutter
5. a pair of clippers
6. a pair of small pliers
7. some small and medium screwdrivers

MATERIALS:

1. a thread of solder
2. a fluxant pen dispenser
3. a breadboard
4. some prototyping boards
5. some "banana" and BNC cables, some "alligator" clamps and some BNC adaptors

(Fig. 7.5.1: banana and alligator --> see later in this textbook)

8.1 Main instruments and tools - Multimeter

The multimeter is for an electronic engineer what probably is the stethoscope for a physician:

(Fig 8.1.1: multimeter)

The multimeter can perform a number of different measurements, usually at least:

1. voltage
2. current
3. resistance

Modern multimeters are usually also "autoranging", which means they can understand themselves the magnitude of the variable they are measuring and select input range of the measuring instruments where this is best represented accordingly, without reaching an overflow or underflow.
They, nowadays, also usually have a digital display, as opposed to a moving needle indicator present in the legacy instruments.

8.2 Main instruments and tools - Variable regulated power supply

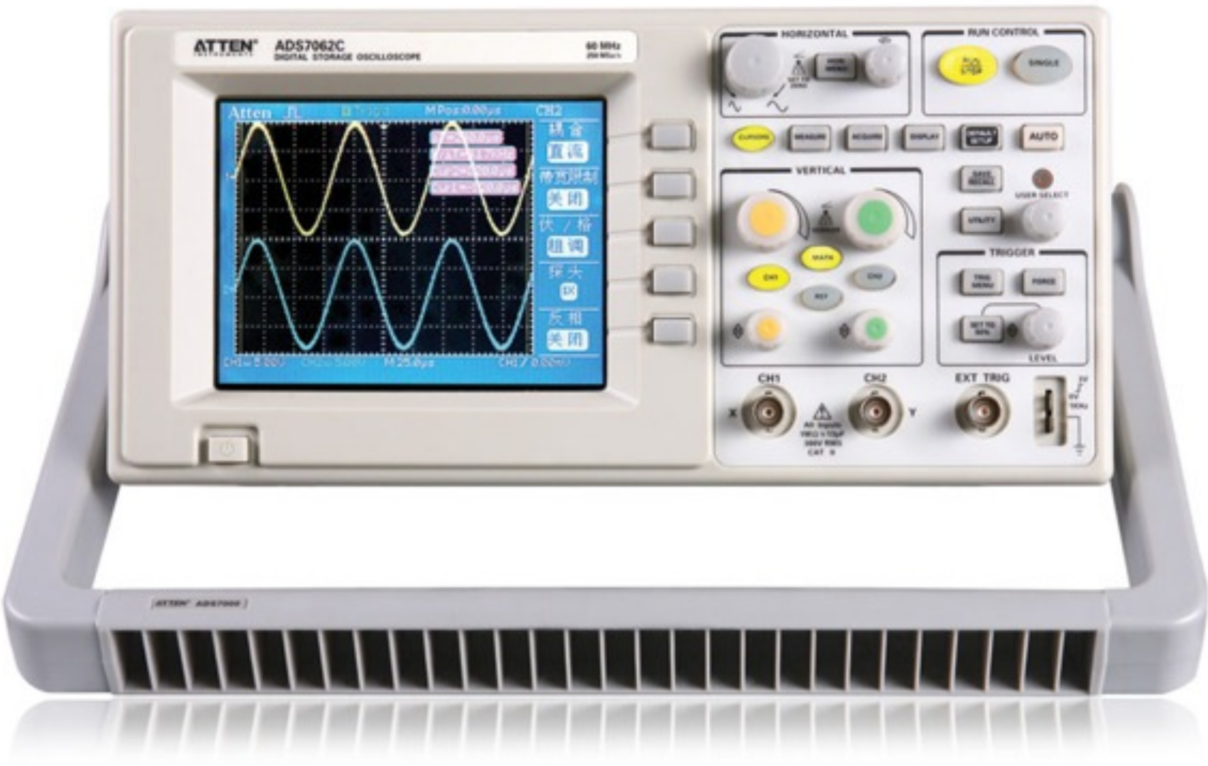
The variable regulated power supply is a device which implement an ideal constant voltage source, and also some others features:

(Fig 8.2.1: variable regulated power bench supply)

It can have multiple independent outputs, each of them regulating the desired voltage and having a variable current limit protection.
They also usually have a short-circuit protection.

8.3 Main instruments and tools - Oscilloscope

The oscilloscope allows you to "see" signals which are varying in the time domain:



(Fig 8.3.1: oscilloscope)

Very important remark: **OSCILLOSCOPES CAN MEASURE ONLY VOLTAGES!**

NOTE:
Nowadays many, even low cost, oscilloscopes can also show signals in the frequency domain.

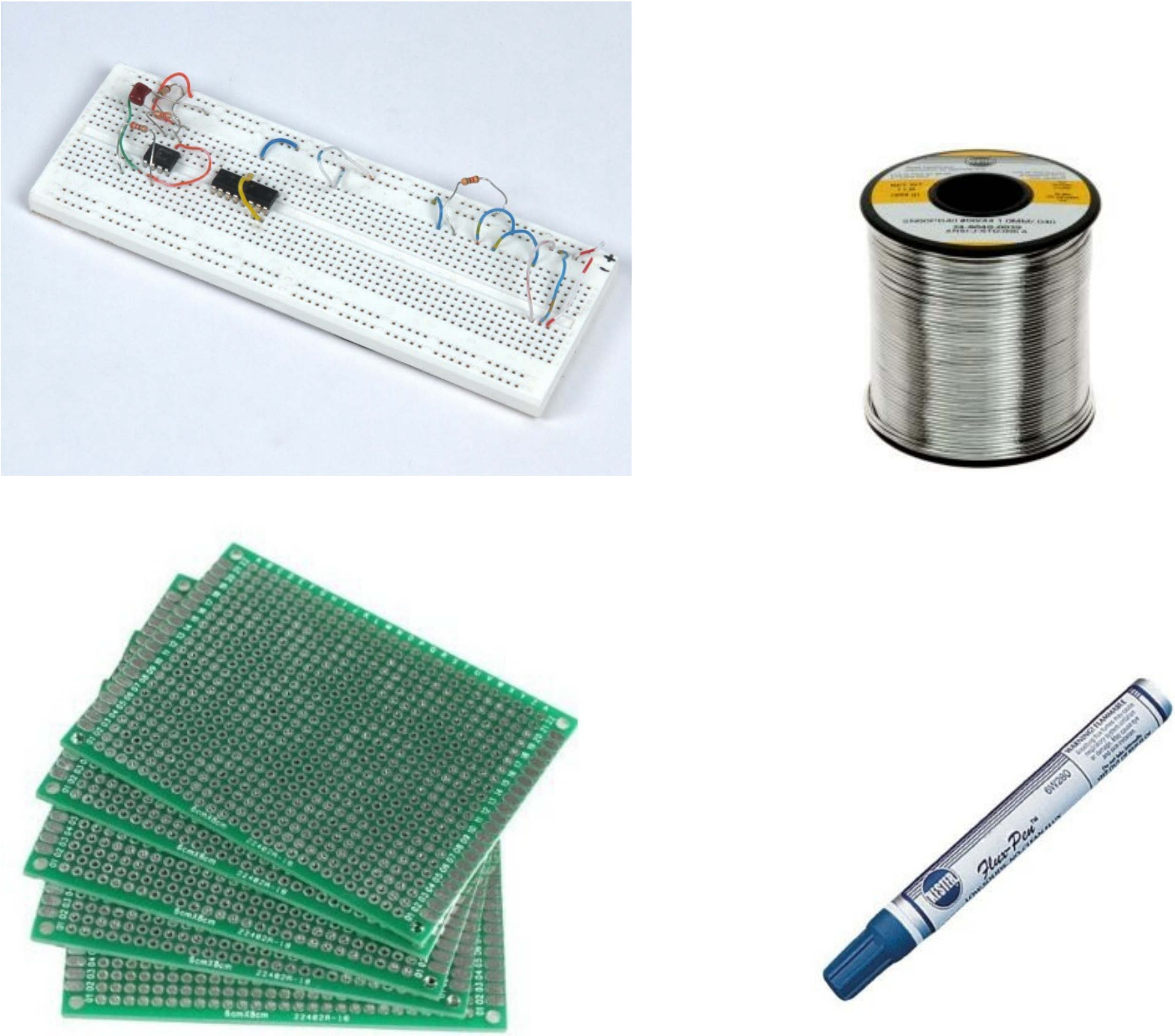
8.4 Main instruments and tools - Soldering iron and sucker

The soldering iron and the solder sucker pump are two complementary instruments: the first one is used to create solder joints between electronic components/wires while the second one, in combination with the first one, is used to remove a solder joint:

(Fig 8.4.1: soldering iron and solder sucker pump)

8.5 Main instruments and tools - Common materials

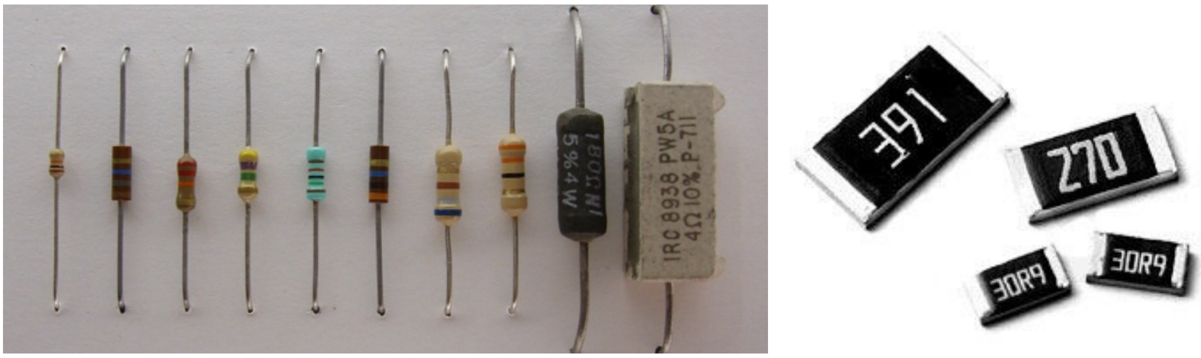
The breadboard, solder and flux pen are common disposable materials in electronics. The "breadboards" can come in different flavours, usually at least two: the one where you can plug-in components and wires as many time as necessary (the white one below in Fig. 8.5.1), and the one on which components and wires are supposed to be firmly soldered on (the green one below in Fig. 8.5.1).



(Fig 8.5.1: breadboards, solder and flux pen)

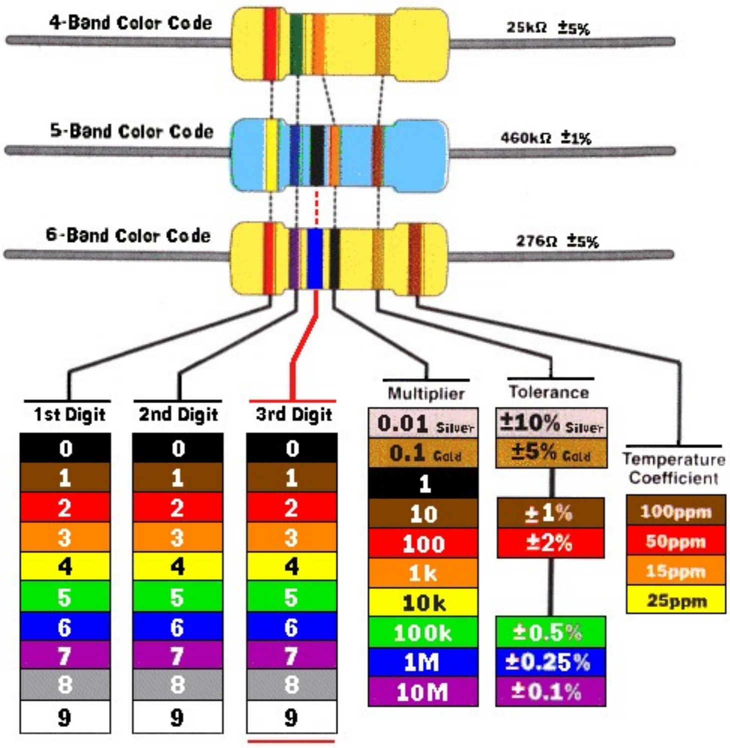
9.1 Commercial components - Resistors

There are different types of resistors. Mainly electronics components are divided in two categories: pass-through hole (PTH) and surface-mount devices (SMD):



(Fig. 9.1.1: PTH resistor -on left- and SMD resistors -on right-)

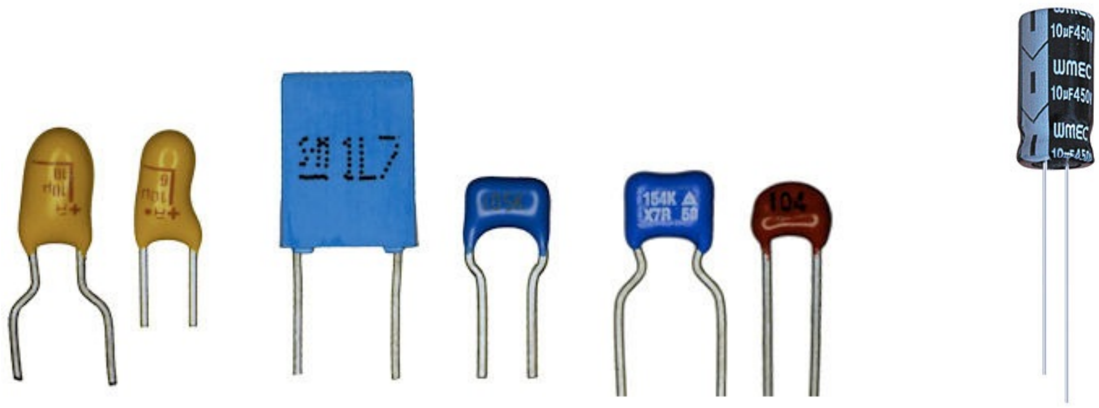
The PTH resistors have a color code which defined the nominal value:



(Fig 9.1.2: the color code)

9.2 Commercial components - Capacitors

There any many types of capacitors, the main two categories are the "non polarised" and the "polarised":



(Fig 9.2.1: PTH capacitors)

(Fig. 9.2.2: SMD capacitors)

9.3 Commercial components - Inductors

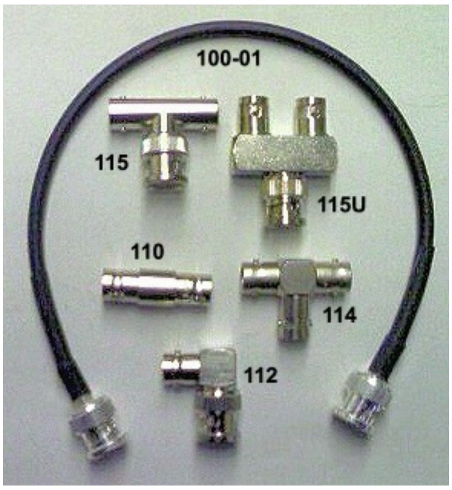
There are many types of inductors, most of them also have a so called magnetic "core":

(Fig 9.3.1: PTH and SMD inductors)

9.4 Commercial components - Cables and connectors

Here there are some examples of cables which are commonly used in laboratory:

(Fig 9.4.1: banana and alligator clamp cable; yes, the fruit and beast were definitely out-of-scope)



(Fig 9.4.2: BNC cable and connectors)

(Fig 9.4.3: IDC -on left- and DB9 -on right- connectors)

9.5 Commercial components - Others

Various electronic components:

(Fig 9.5.1: various electronic components)

and various electronic modules:

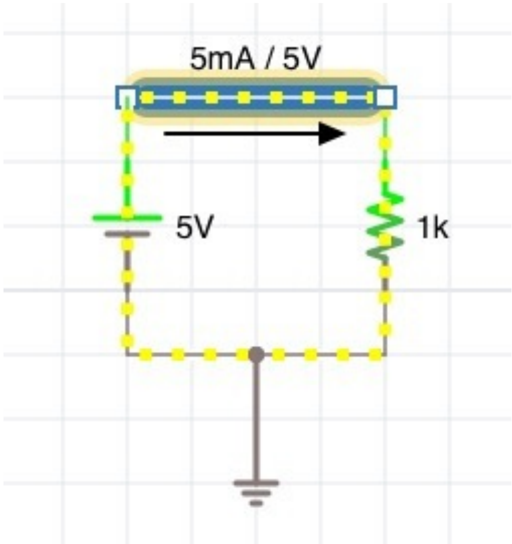
(Fig 9.5.2: various electronic modules)

10.1 DC circuits - Voltage and current source

The term DC stands for "direct current". It used to identify situations where either the voltages or the currents involved in a circuit do not depend on time, in other words when they are constant.

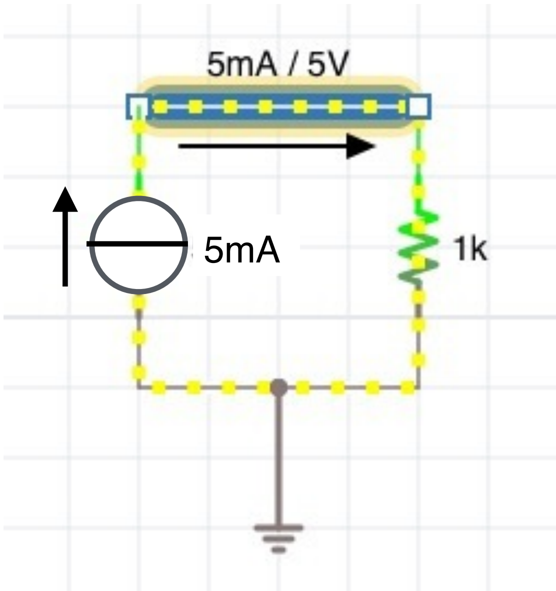
The simplest examples are the following circuits:

- Voltage generator attached to a resistor



(Fig 10.1.1: voltage generator)

- Current generator attached to a resistor



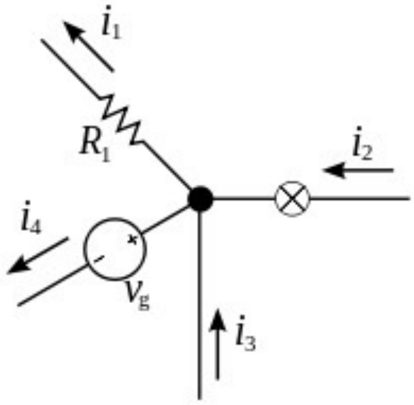
(Fig 10.1.2: current generator)

10.2 DC circuits - Kirchoff's laws

The Kirchoff's laws describe how current and voltage behave in circuits. They have general validity in either stationary or quasi-stationary time-variable condition (until radiative effects become important, otherwise the relativistic electromagnetism formulation must be used):

- 1ST KIRCHOFF'S LAW

The algebraic sum of all currents converging to a node is zero.

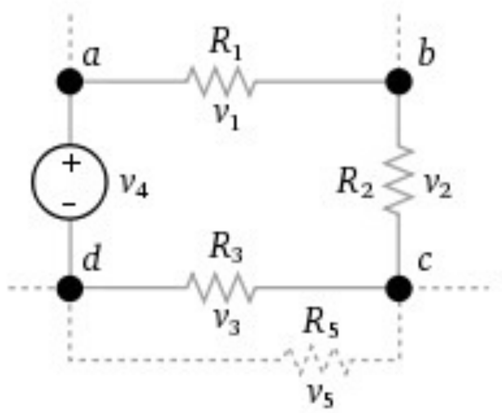


(Fig 10.2.1: first Kirchoff's law)

Currents entering a node are considered to be positive, while the ones exiting the node are considered to be negative. In other words, in a node the sum of all entering currents must equal the sum of all exiting currents. This basically corresponds to conservation of the charge for electrostatic fields.

- 2ND KIRCHOFF'S LAW

The algebraic sum of all voltages along a mesh is zero.

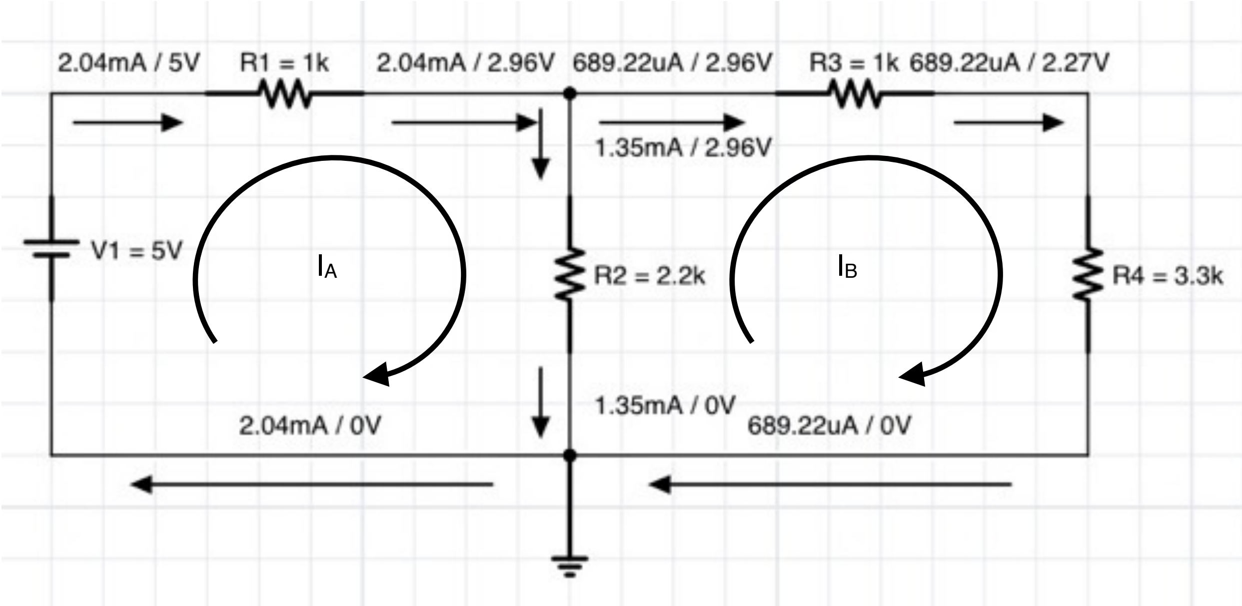


(Fig 10.2.2: second Kirchoff's law)

Currents entering passive components make voltage drops with positive pole at the entering side, currents flowing inside voltage generators go from the negative pole to the positive pole. In other words, in a mesh the sum of all voltage drops must equal the sum of all the voltage generators. This is because the electrostatic field is a conservative field.

10.3 DC circuits - Node analysis

This is an example of a DC circuit:



(Fig 10.3.1: example of DC circuit)

which gives the following simultaneous equations:

$$\begin{cases} R_1 I_A + R_2\left(I_A - I_B\right) = V_1 \\ R_2\left(I_B - I_A\right) + R_3 I_B + R_4 I_B = 0 \end{cases}$$

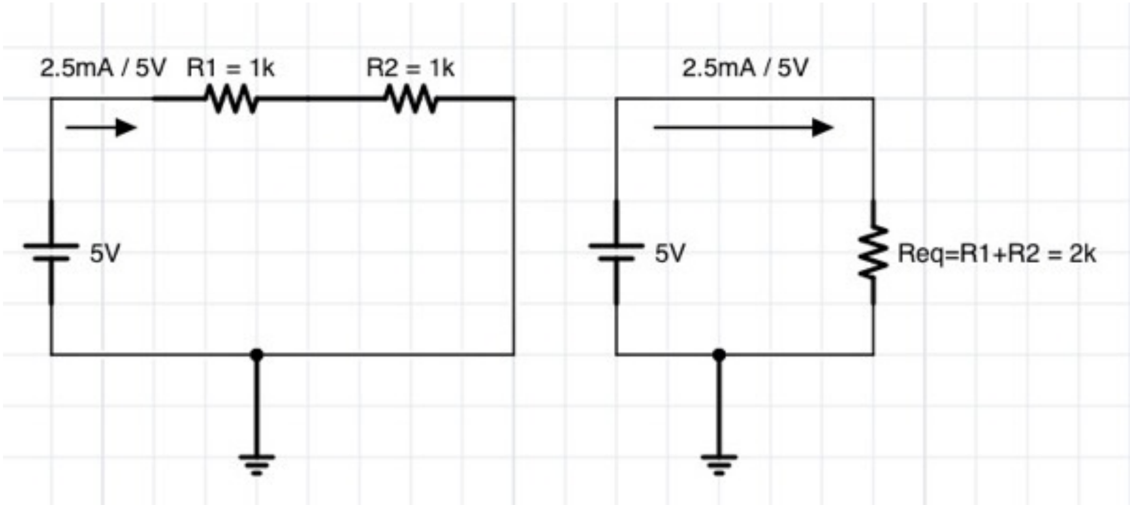
(Eq 10.3.1)

from which the mesh currents I_A and I_B can be calculated and hence the voltage of each node can be obtained.

10.4 DC circuits - Series and parallel circuits

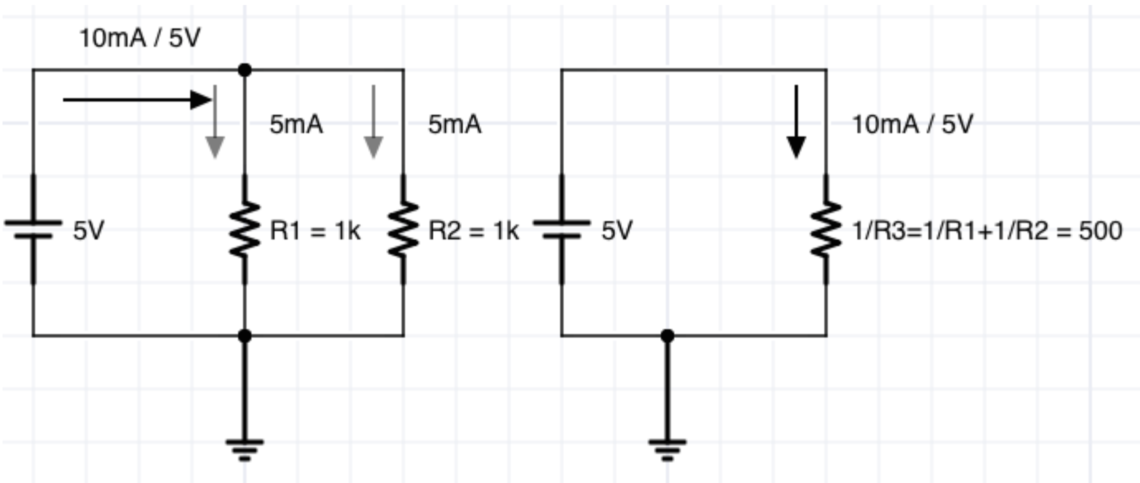
There are two very small circuit topologies which are very important to know:

- SERIES (resistors)



(Fig 10.4.1: series circuit)

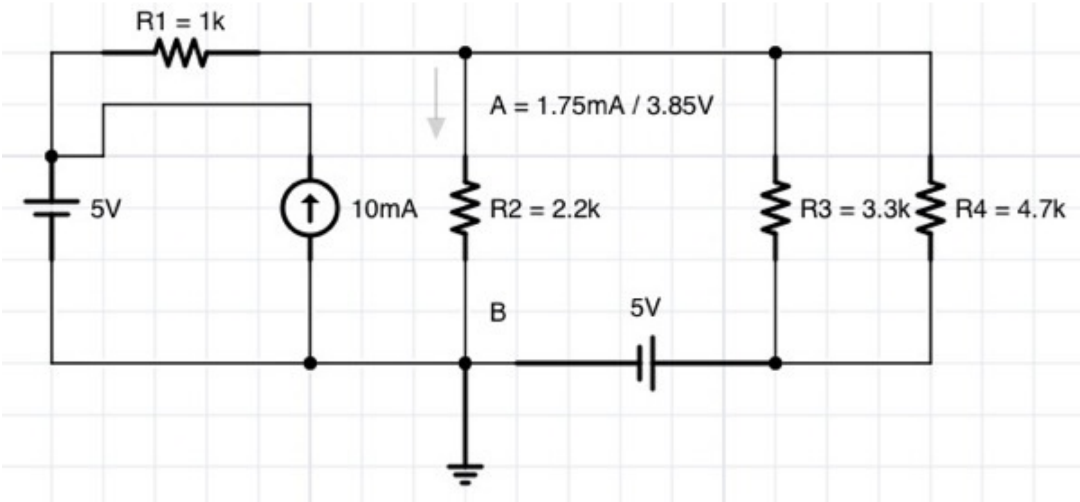
- PARALLEL (resistors)



(Fig 10.4.2: parallel circuit)

10.5 DC circuits - Thevenin's theorem

The Thevenin's theorem shows how to replace a portion of circuit with an equivalent circuit made of only one voltage source in series with only one circuit element:

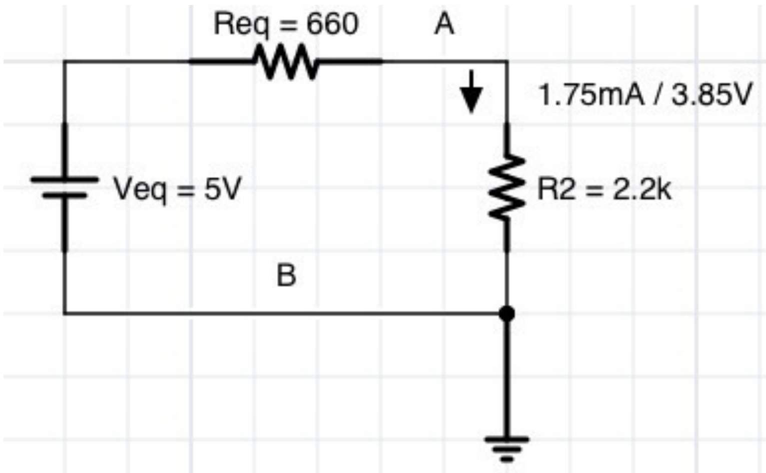


(Fig 10.5.1: circuit)

The method works as follows:

1. Leave alone the portion of circuit of interest (here, remove R_2)
2. Open all the current sources
3. Calculate the equivalent voltage between A and B
4. Short all the voltage sources
5. Calculate the equivalent resistance between A and B

The resulting equivalent circuit is then:



(Fig 10.5.2: equivalent circuit)

The latter circuit is "equivalent" of the original one in the sense that, from the point of view of the nodes A and B around which the equivalent circuit has been developed, the voltage between A and B and the current from A to B remain the same.

11.1 AC circuits - Linear circuits

The term AC stands for alternate circuit and is used when circuits are driven by sinusoidal voltage or current sources. The case of electromagnetic radiation is still not considered: this analysis is valid when the radiative effects are negligible.

Among all the possible circuits, an important category is the one called "linear circuits". Linear circuits are circuits for which all waveforms in it, voltage or current, are linear combinations of sinusoids having different amplitude or phase but maintaining the same frequency.

In other words, the presence of the circuit elements does not alter the frequency of the sources, but in case only the amplitude or the phase.

There are only some circuit elements having this property, they are:

1. Resistors
2. Capacitors
3. Inductors
4. Ideal amplifiers

All the theorems of DC circuits, build from the same elements in the list, also hold for the linear circuits: in fact DC linear circuits are a special case of AC linear circuits having frequency equals to zero.

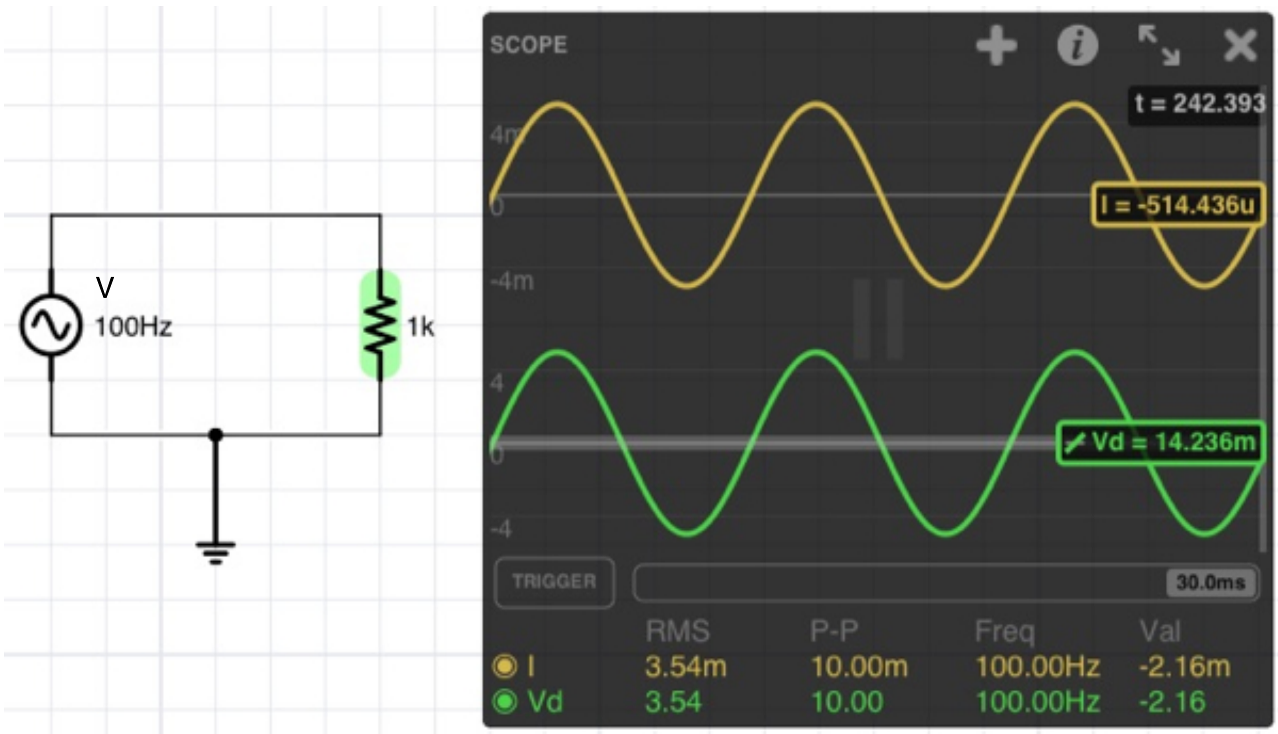
NOTICE: once again, the assumption is that radiative effects are negligible otherwise the relativistic equations of the electromagnetism should be taken in account. This assumption is generally valid when the electromagnetic energy due to the radiative phenomena is small compared to the electromagnetic energy present within the circuit elements. Just as a rule of thumb, in range of voltage magnitude up to about 10V, current magnitude up to about 1A, frequency up to about 10 MHz and circuit size up to about 1 cubic meter the assumption is usually valid.

11. 2 AC circuits - Impedance of a resistor

The main concept in AC circuit analysis is the "impedance".
The impedance is the extension of the concept of resistance for AC circuits:

$$Z_R(\omega) = \frac{V(\omega)}{I(\omega)} = \frac{Ve^{i\omega t}}{\frac{Ve^{i\omega t}}{R}} = R$$

(Eq 11.2.1)



(Fig 11.2.1: impedance of a resistor)

where it has been made use of the Ohm's law.

Notice that:

$$\omega = 2\pi f$$

(Eq. 11.1.2)

11.3 AC circuits - Impedance of a capacitor

The impedance of a capacitor comes from the Coulomb's law and the definition of current:

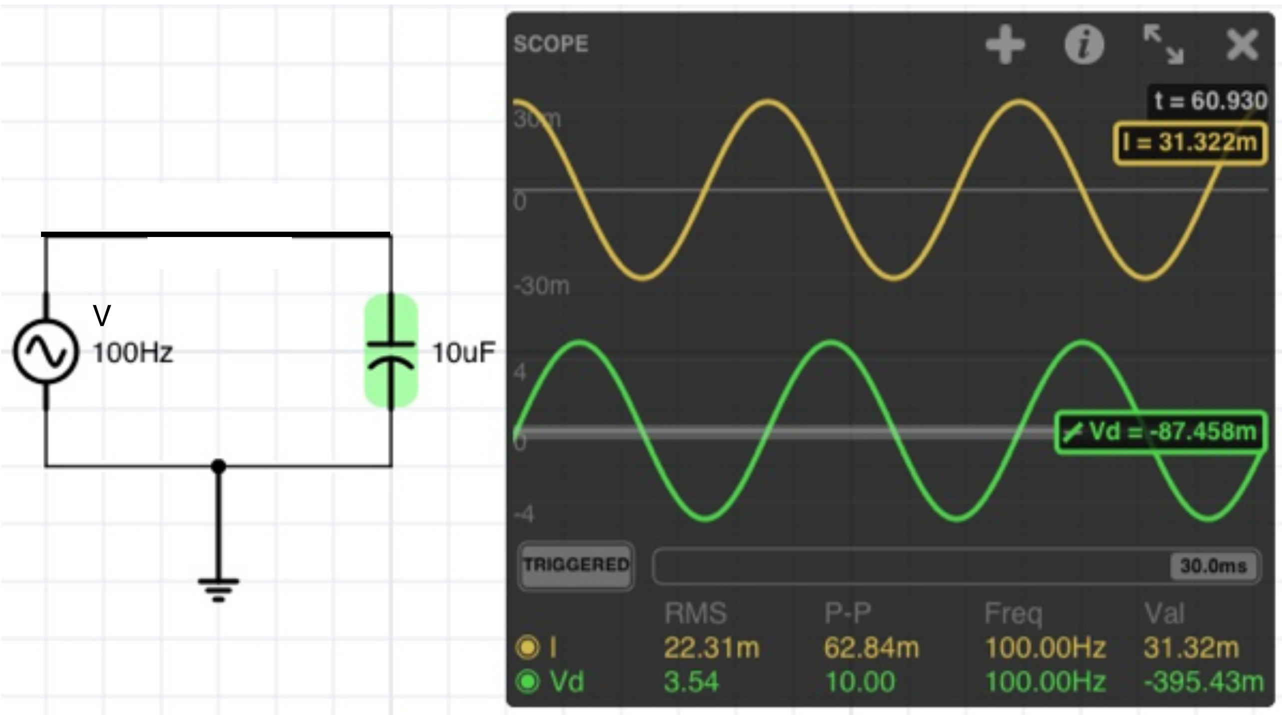
$$\begin{cases} Q(t) = CV(t) \\ I(t) = \dfrac{dQ}{dt} \end{cases}$$

(Eq 11.3.1)

Using those equations, the impedance of a capacitor is:

$$Z_C(\omega) = \frac{V(\omega)}{I(\omega)} = \frac{1}{C}\frac{Ve^{i\omega t}}{i\omega Ve^{i\omega t}} = \frac{1}{i\omega C}$$

(Eq 11.3.2)



(Fig 11.3.1: impedance of a capacitor)

Notice the AC generator is a voltage generator.

11.4 AC circuits - Impedance of an inductor

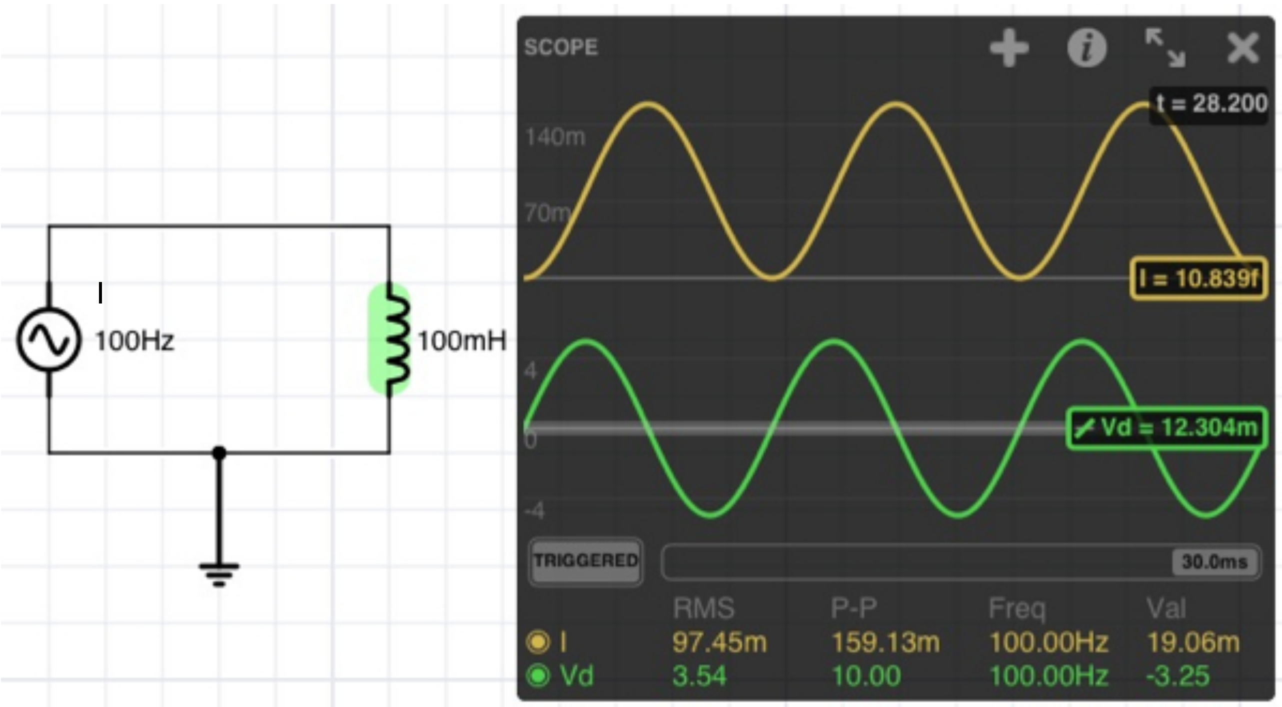
The impedance of an inductor comes from the Faraday's law:

$$V(t) = L\frac{dI}{dt}$$

(Eq 11.4.1)

$$Z_L(\omega) = \frac{V(\omega)}{I(\omega)} = L\frac{i\omega I e^{i\omega t}}{I e^{i\omega t}} = i\omega L$$

(Eq 11.4.2)

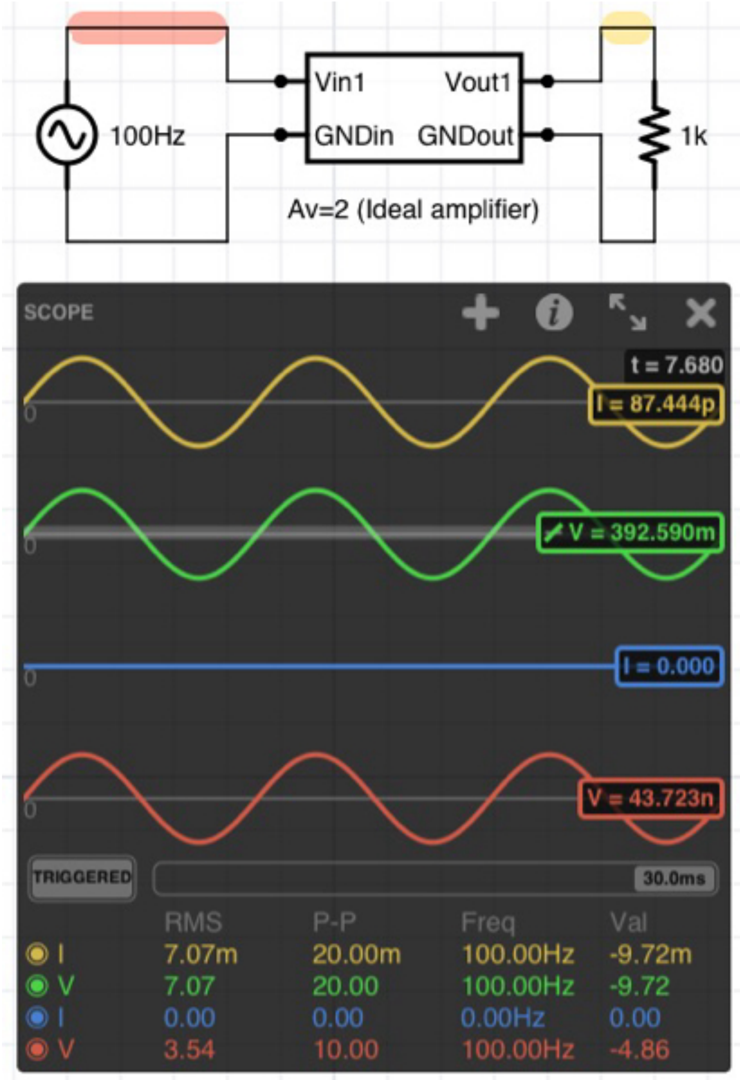


(Fig 11.4.1: impedance of an inductor)

Notice this time the AC generator is a current generator.

11.5 AC circuits - Ideal amplifiers

The ideal amplifier is a device having an input impedance equals to infinity (no current is drawn from the source attached to its input) and output impedance equals to zero (it can drive any load holding its output voltage unchanged). The ratio between the output and the input voltage is called "voltage gain" or simply "gain", while the ratio between the output and the input current is the "current gain":



(Fig 11.5.1: ideal amplifier)

The voltage gain is hence:

$$G = \frac{V_{out}}{V_{in}}$$

(Eq. 11.5.1)

Notice that the output and the input voltage are in phase.

12.1 Filters - Transfer function

In electronics, filters are 2-port network circuits able to transform a signal applied to their input into an output signal having a different frequency spectrum respect to the input signal.
Important filters are the "linear filters", which can only change the amplitude and the phase of each frequency component present in the input signal. They cannot create in the output frequency components not present in the input. More precisely, linear filters are "linear time-invariant systems".

The mathematical description of this transformation in the frequency domain is:

$$V_{out}(\omega) = A(\omega) V_{in}(\omega)$$

(Eq 12.1.1)
where A(ω) is called "transfer function" of the filter and:

$$\begin{cases} V_{in}(\omega) = \left\langle V_{in}^{*}(t) \middle| e^{i\omega t} \right\rangle \\ V_{out}(\omega) = \left\langle V_{out}^{*}(t) \middle| e^{i\omega t} \right\rangle \end{cases}$$

(Eq 12.1.2)

are the Fourier transforms of their associated time-varying signals.

Transfer functions which can be written on a rational polynomial form are important because they can be implement by lumped element circuits:

$$\begin{cases} A(s) = \dfrac{N(s)}{D(s)} = \dfrac{a_n s^n + \ldots + a_2 s^2 + a_1 s + a_0}{b_m s^m + \ldots + b_2 s^2 + b_1 s + b_0} \\ s = i\omega \end{cases}$$

(Eq 12.1.3)

The roots of N(s) are called "zeros", while the ones of D(s) are called "poles". The degree of the rational polynomial (which is the maximum between the degree of the numerator and the denominator) is called "order" of the filter.

12.2 Filters - Kernel

The "kernel" is the associated representation of the transfer function in the time domain:

$$V_{out}(t) = k * V_{in}(t) = \int_{-\infty}^{+\infty} k(\tau) V_{in}(t-\tau) d\tau$$

(Eq 12.2.1)

where the kernel k(t) is the inverse Fourier transform of the transfer function A(ω) of the filter:

$$k(t) = \left\langle e^{-i\omega t} \middle| A(\omega) \right\rangle$$

(Eq 12.2.2)

This is the result of the so called "convolution theorem".

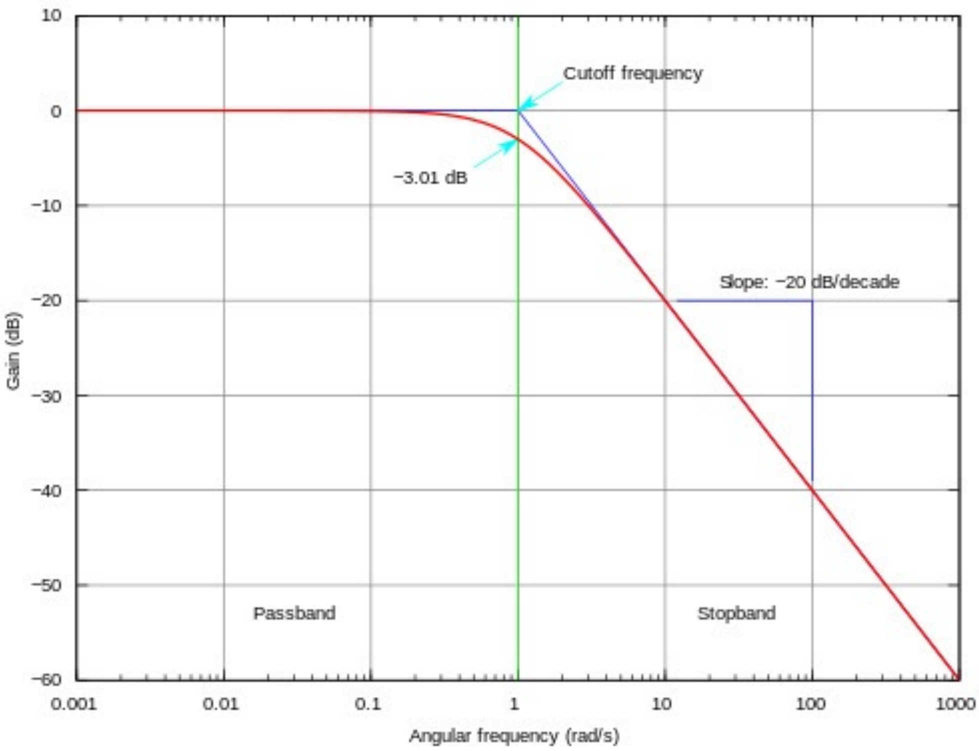
12.3 Filters - Low pass filter

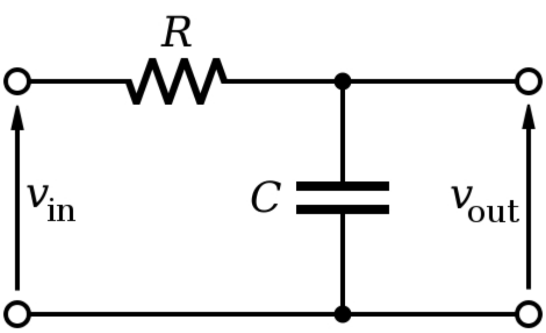
The low pass filter rejects frequencies above its cut-off frequency. It can be implemented as a polynomial filter:

$$\begin{cases} A(\omega) = \dfrac{1}{1+s\tau} \\ \omega_0 = \dfrac{1}{\tau} = \dfrac{1}{RC} \end{cases}$$

(Eq 12.3.1)

where ω_0 is the "cut-off frequency" of the filter, defined as the frequency at which the output signal is attenuated by 3dB respect to its level in the pass-band.





(Fig 12.3.1: first order low-pass filter)

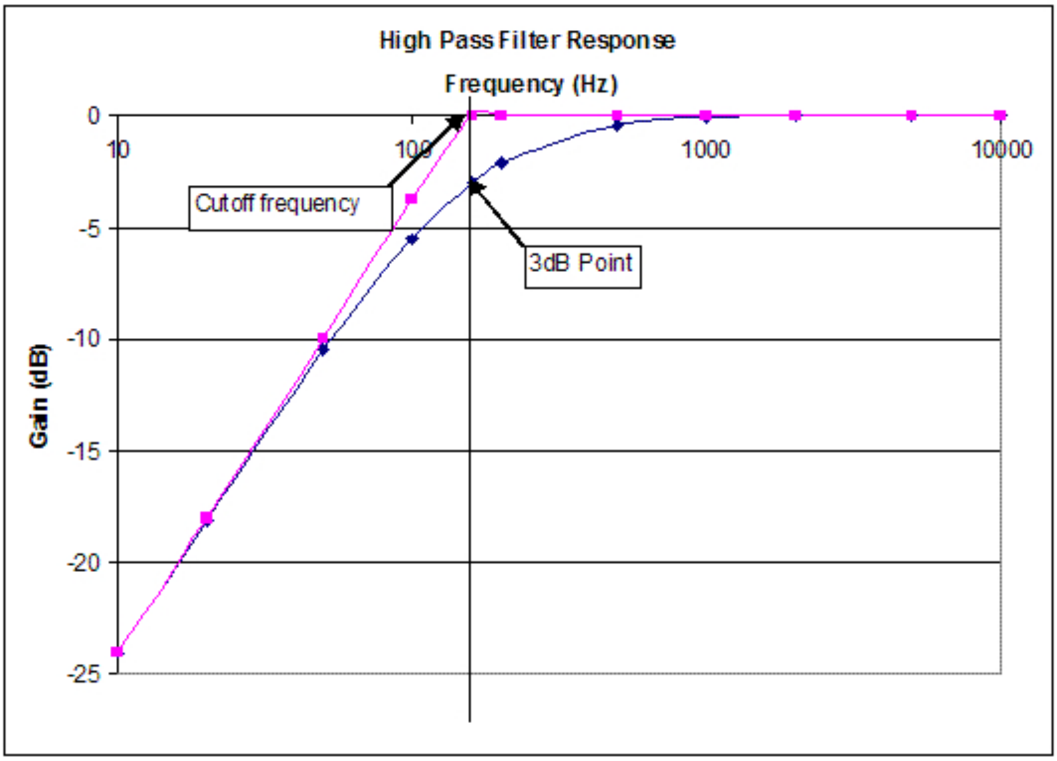
12.4 Filters - High pass filter

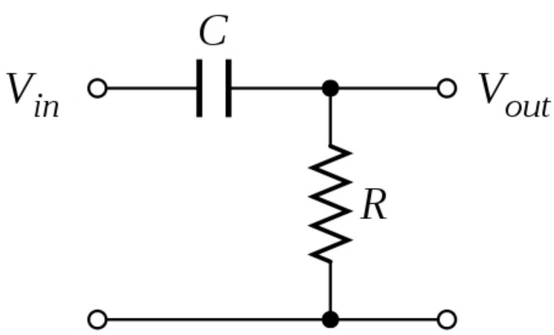
The high pass filter rejects frequencies below its cut-off frequency. It can be implemented as a polynomial filter:

$$\begin{cases} A(\omega) = \dfrac{s\tau}{1+s\tau} \\ \omega_0 = \dfrac{1}{\tau} = \dfrac{1}{RC} \end{cases}$$

(Eq 12.4.1)

where ω_0 is the "cut-off frequency" of the filter, defined as the frequency at which the output signal is attenuated by 3dB respect to its level in the pass-band.

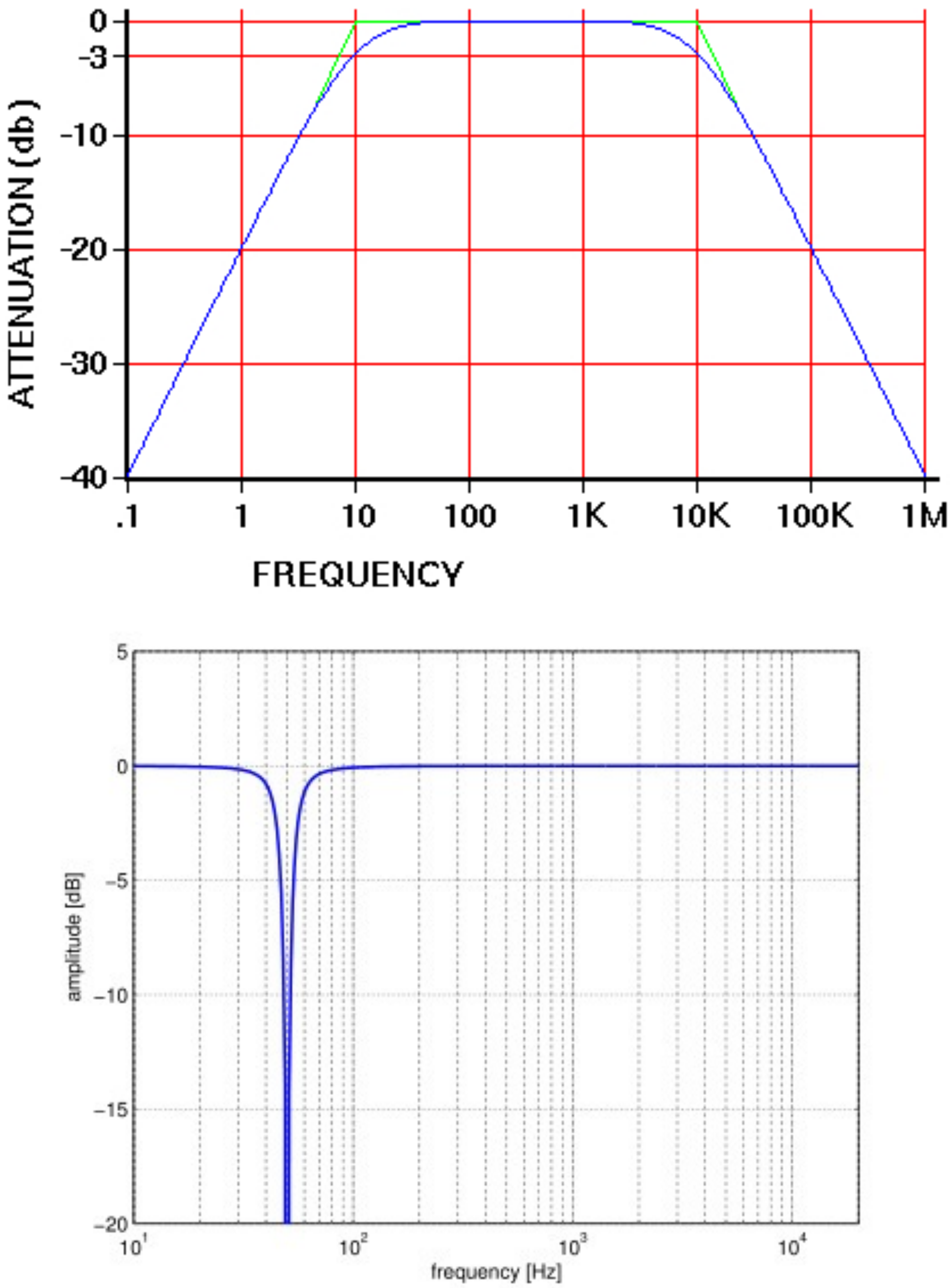




(Fig 12.4.1: first order high-pass filter)

12.5 Filters - Band pass and notch filters

By cascading low pass and high pass filters having different cut-off frequencies it is possible to implement either a "band pass" or a "notch filter":

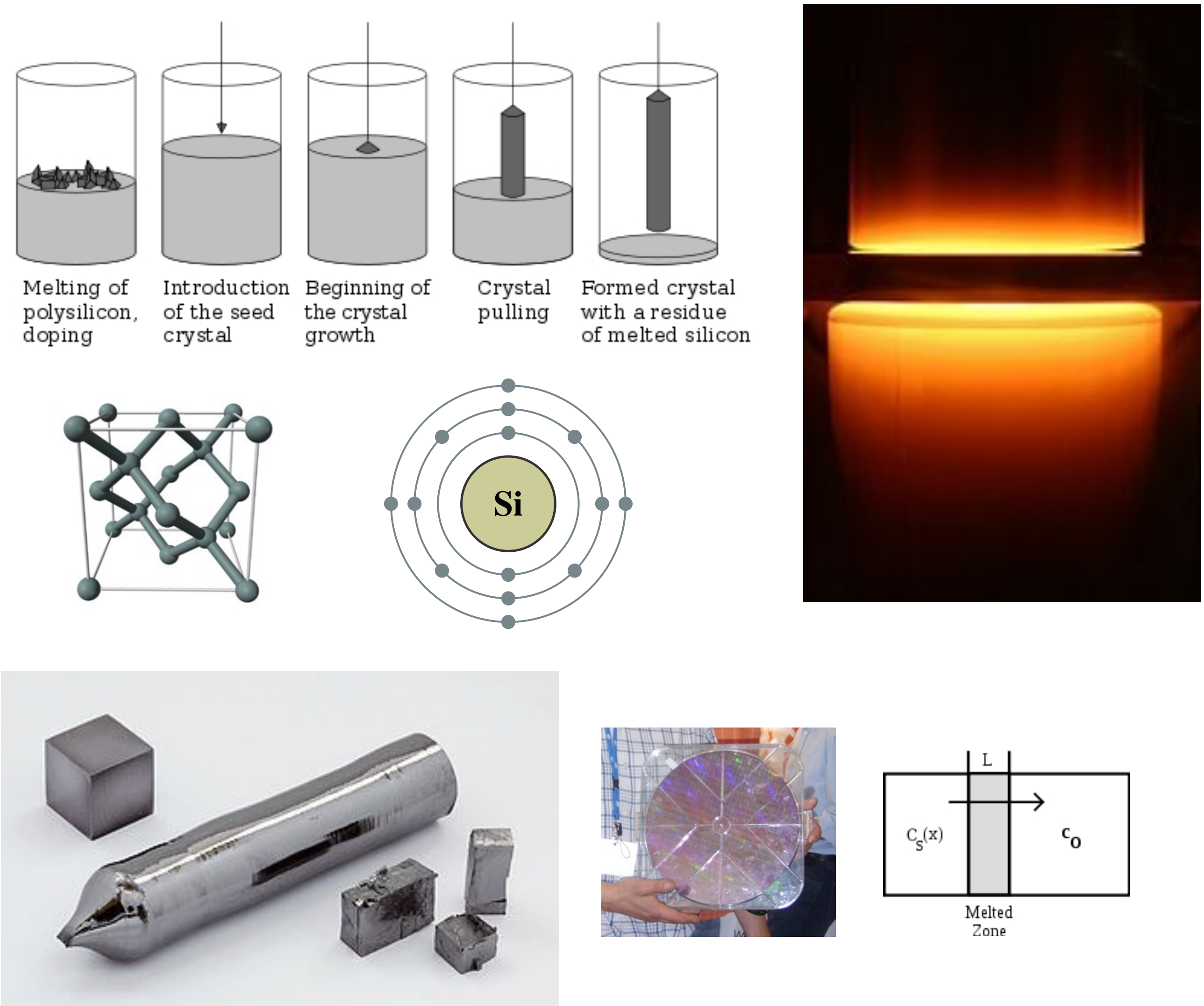


(Fig 12.5.1: band-pass -upper panel- and notch -lower panel- filters)

Notice that an "all pass filter" exists too, which does not change the amplitude of any frequency component but can change their phase. If specifically designed it can work as a "delay line", which is a circuit able to delay all frequency components of a signal by a defined amount of time equal for each frequency component.

13.1 Semiconductors - Silicon wafers

In order to build semiconductor devices a very high quality silicon is necessary. Two common methods of production are the "Czochralski" (CZ) and the "floating zone" (FZ) processes:



(Fig 13.1.1: CZ and FZ crystal growing methods)

In the CZ method a "crystal seed" is pulled up from a melting of "poly-silicon", creating a pure "mono crystal" silicon rod.
In the FZ method an impure silicon rod is scanned up and down by a ring in which high frequency current flows: the high frequency induce a strong magnetic field in the silicon rod, which melts it creating a floating zone having a gradient of temperature from the inside to the outside of the axis of the rod. This gradient of temperature corresponds to a gradient of concentration of impurities, which are this way pumped off axis by osmosis. The impurities then falls down by gravity. Scanning up and down the rod makes the bottom of it collecting all the impurities. At the end of the process the bottom is cut off and the remaining part is a pure mono crystal.

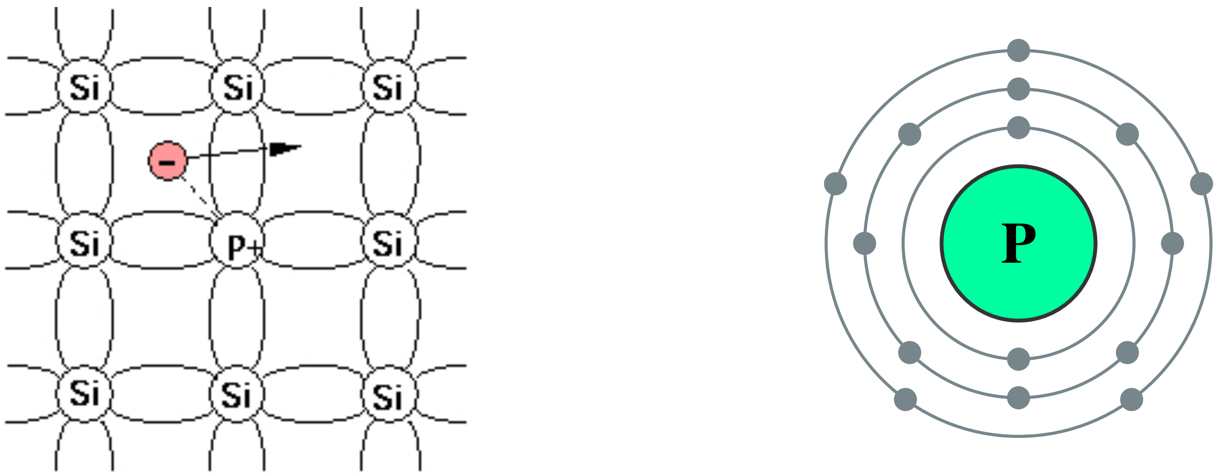
The mono crystals are later cut in circular slices, called "wafers". The semiconductor devices are then built on the wafer by means of photolithographic process localised in many small square-like areas which are later cut into small chips, the so called "microchips".

13.2 Semiconductors - Doping

Pure crystal silicon is called "intrinsic". In intrinsic silicon at room temperature there is some probability for an electron to cross the energy gap and become a conduction electron. This probability can be changed dramatically by inserting some "dopant" atoms in the crystal lattice.

N-TYPE SEMICONDUCTORS

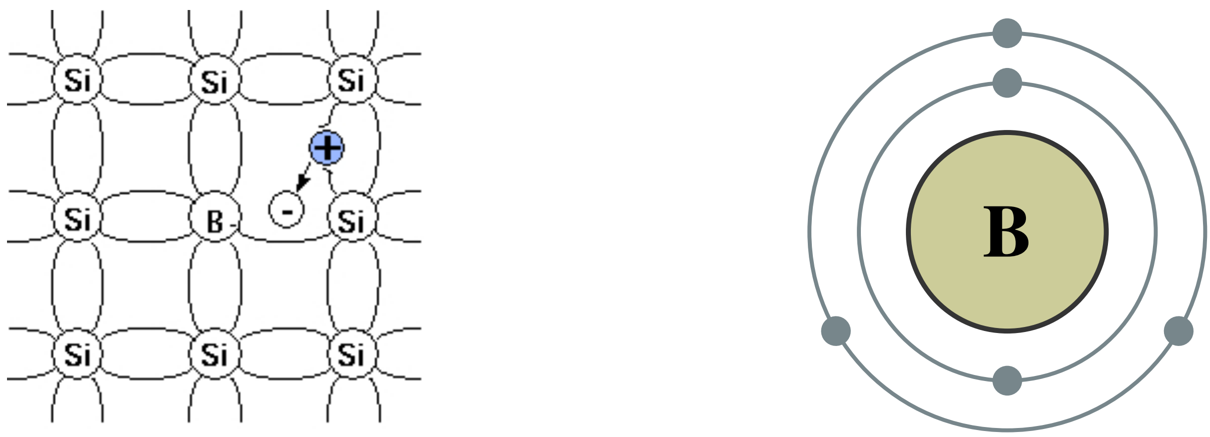
By substituting a silicon atom with a phosphorus atom the latter one becomes positively ionised and liberates one electron in the conduction band, adding it to the lattice as a free electron:



(Fig 13.2.1: N-type doped silicon)

P-TYPE SEMICONDUCTORS

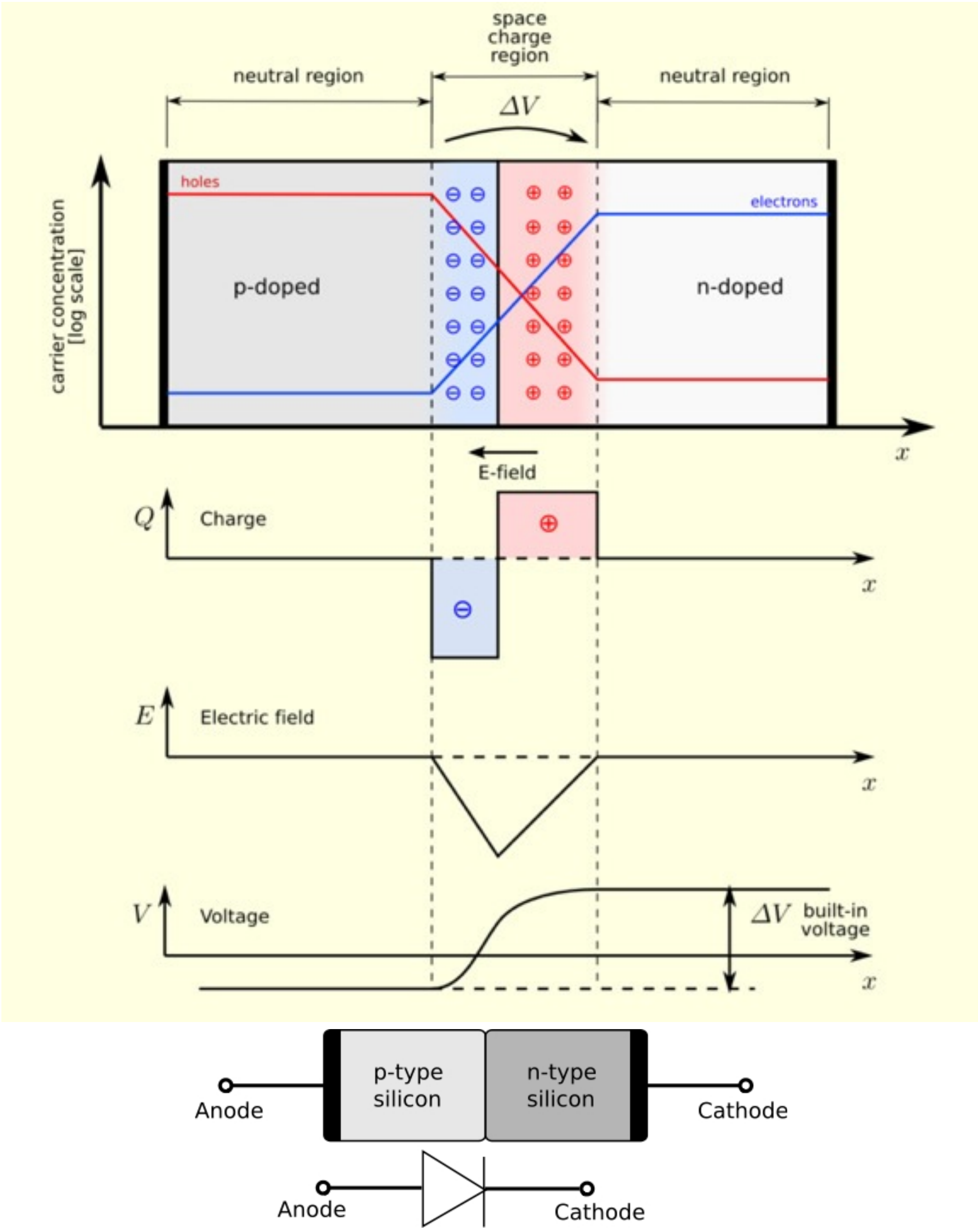
Similarly, by substituting a silicon atom with a boron atom the latter attracts a neighbour electron and becomes negatively ionised. By doing so it leaves a "hole" (also called "electron vacancy") which will be subsequently filled by another electron, and so on creating a moving positive charge in the crystal:



(Fig 13.2.2: P-type doped silicon)

13.3 Semiconductors - PN-junctions

By welding together a P-type region and an N-type region a PN-junction is obtained:



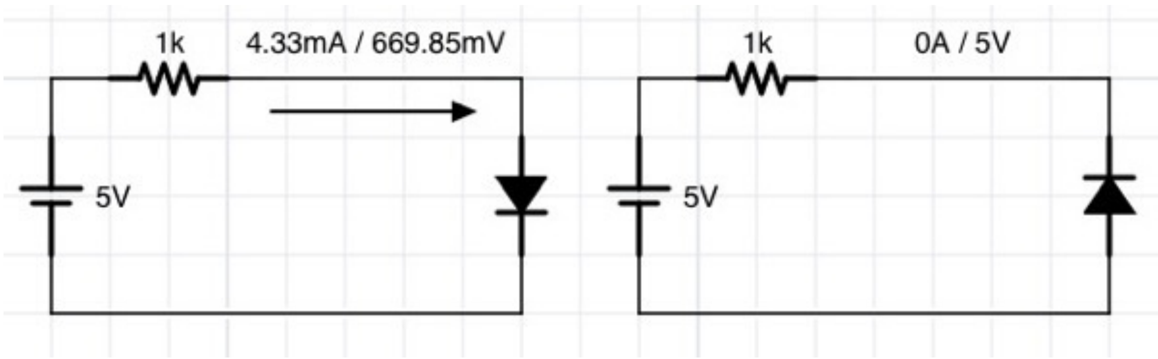
(Fig 13.3.1: the semiconductor silicon diode)

The free electrons in the N-region tend to recombine with the free holes in the P-region, changing the charge distribution across the junction and hence creating an electric field. The existence of such electric field is detected by the presence of the built-in voltage, which arises from it and which value depends on the dopants. Usually for silicon it is about 0.7V at room temperature.

Notice the junction is not obtained by simply putting a P-type silicon in contact with a N-type silicon: the two parts needs to have the crystal lattice to be in perfect continuation in order to make a proper PN-junction. When correctly done, whatever the used technique, this corresponds to have the two parts physically "welded" together.

13.4 Semiconductors - Diodes

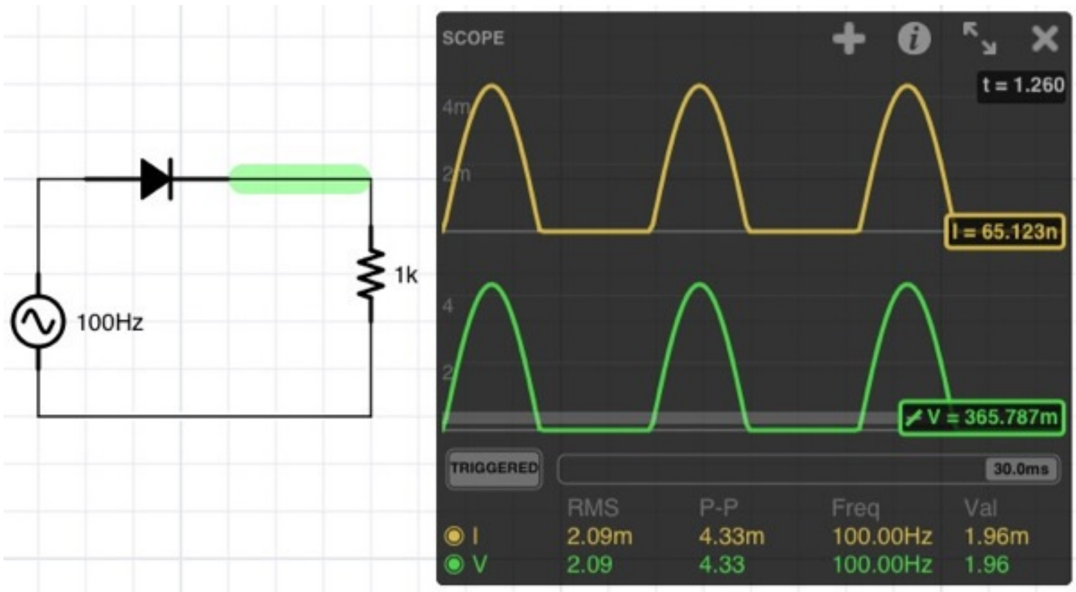
A diode is a device for which the conduction is made asymmetric by the conditions of polarisation of a PN-junction. Because of this property, diodes mainly conduct electricity only in one direction:



(Fig 13.4.1: forward bias)

In the forward bias, the electric field generated by the battery pushes the holes in the P-region and the electrons in the N-regions towards each other, shortening the charge unbalance region across the junction. If the voltage of the battery is greater than the built-in voltage of the PN-junction then the energy gap across the junction is zeroed and current can flow.

In reverse bias, the electric field generated by the battery makes the charge unbalance region across the junction becoming bigger and no current can flow:

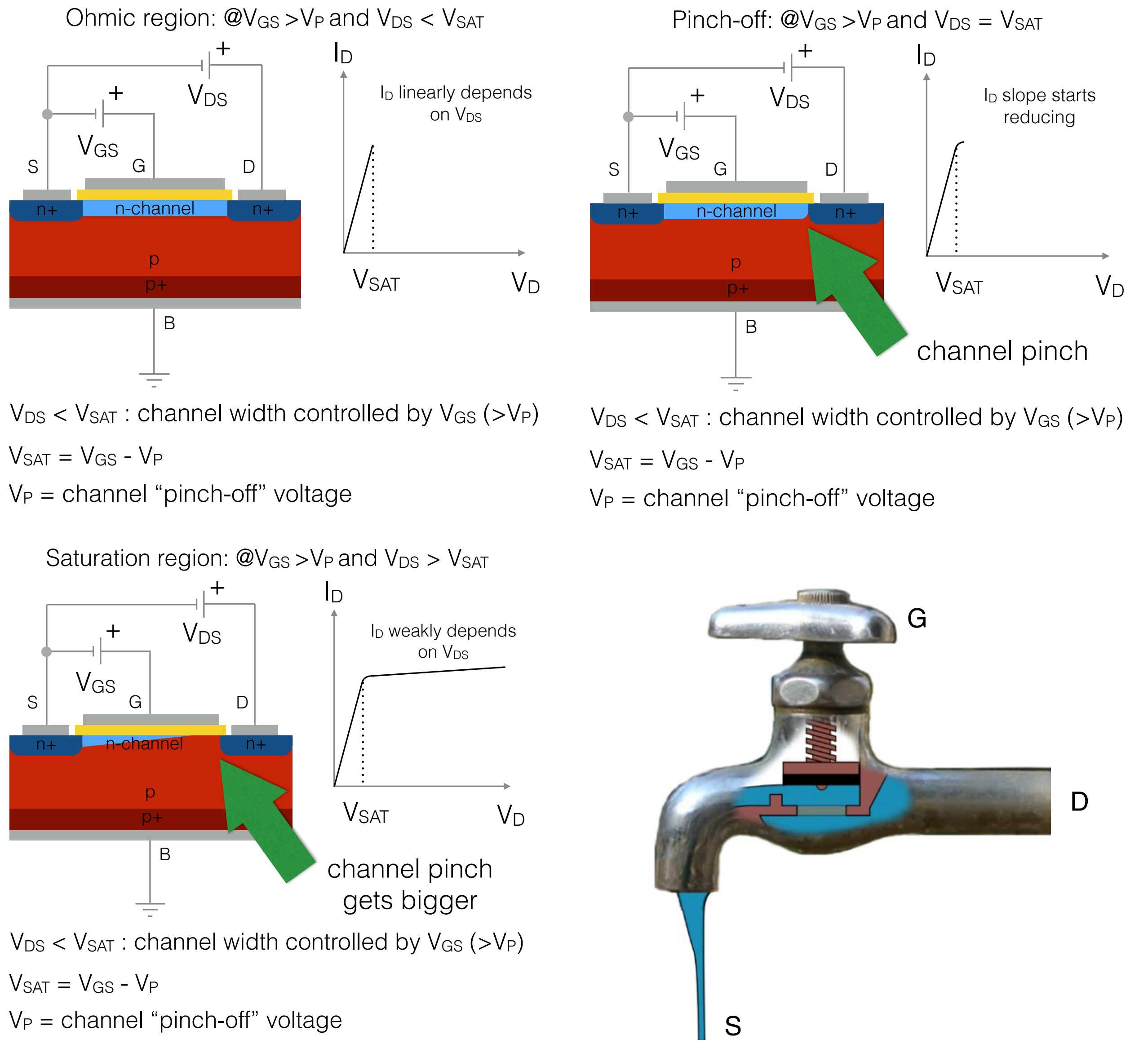


(Fig 13.4.2: inverse bias)

Notice that in both P and N regions there is always a certain small amount of so called "minority charge carriers". In the P-region the "majority charge carriers" are the holes, while in the N-region they are the electrons. The minority charge carriers are responsible of a small current which is always opposite to the main stream due to the majority charge carriers. This makes the net forward current to be a little bit smaller than the one purely due to the majority charge carries; conversely in reverse bias the current is not exactly zero but there is always a certain small small amount of "reverse current" due to the minority charge carriers.

13.5 Semiconductors - Transistors

By using photolythographic processes it is possible to build the following structure on a silicon wafer:



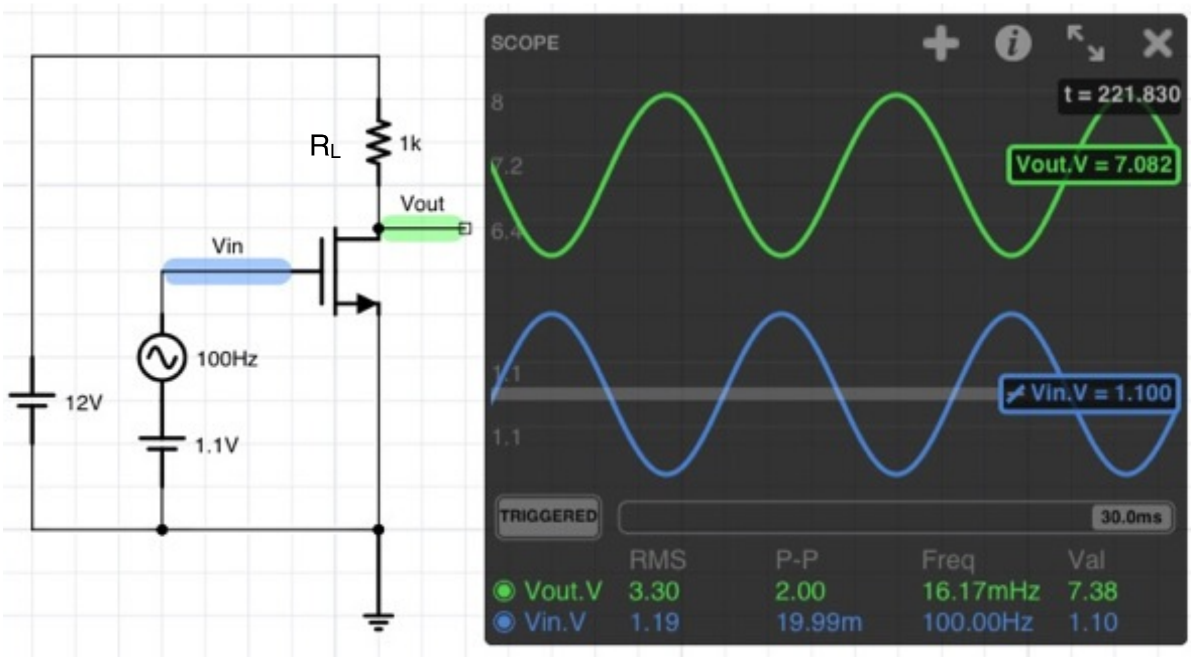
(Fig 13.5.1: MOSFET operative regions and water tap analogy)

This is a MOSFET - Metal Oxide Semiconductor Field Effect Transistor (Transfer resistor). This device can work in two different regions: "ohmic", where in acts as a voltage-controlled linear resistor and "saturation", where it works as linear amplifier (see later).

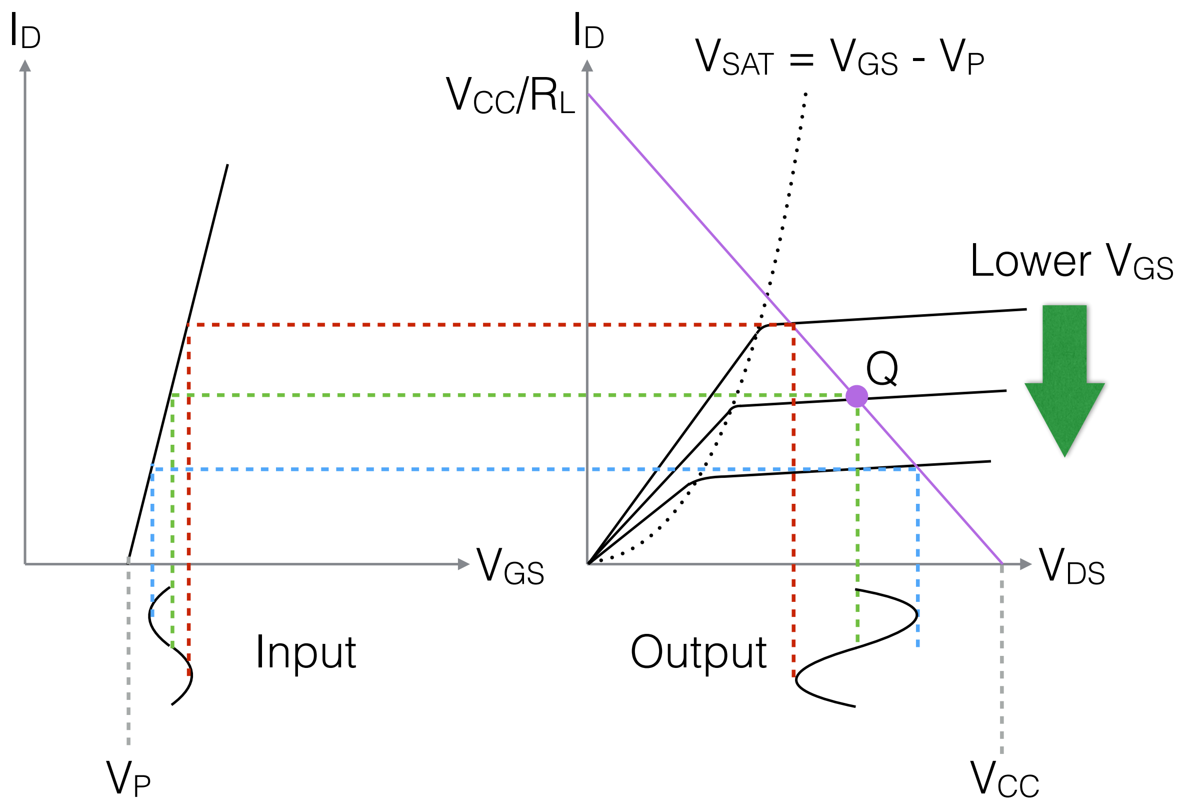
In a P-type MOSFET, a N-channel appears as a consequence of a positive gate repelling holes in the bulk material. The gate voltage controls the width of the channel, hence its conductivity. A voltage applied between the source and the drain makes a second current flowing between them, whose magnitude is modulated by acting on the channel according to the ohmic or saturation operative modes. Notice that the "source" and "drain" terms refer to the charge carriers in the channel: hence for a P-type MOSFET the source terminal is a source of electrons; similarly for the drain terminal.

14.1 Amplifiers - Single transistor amplifiers

The fundamental amplifier is made of a single transistor:



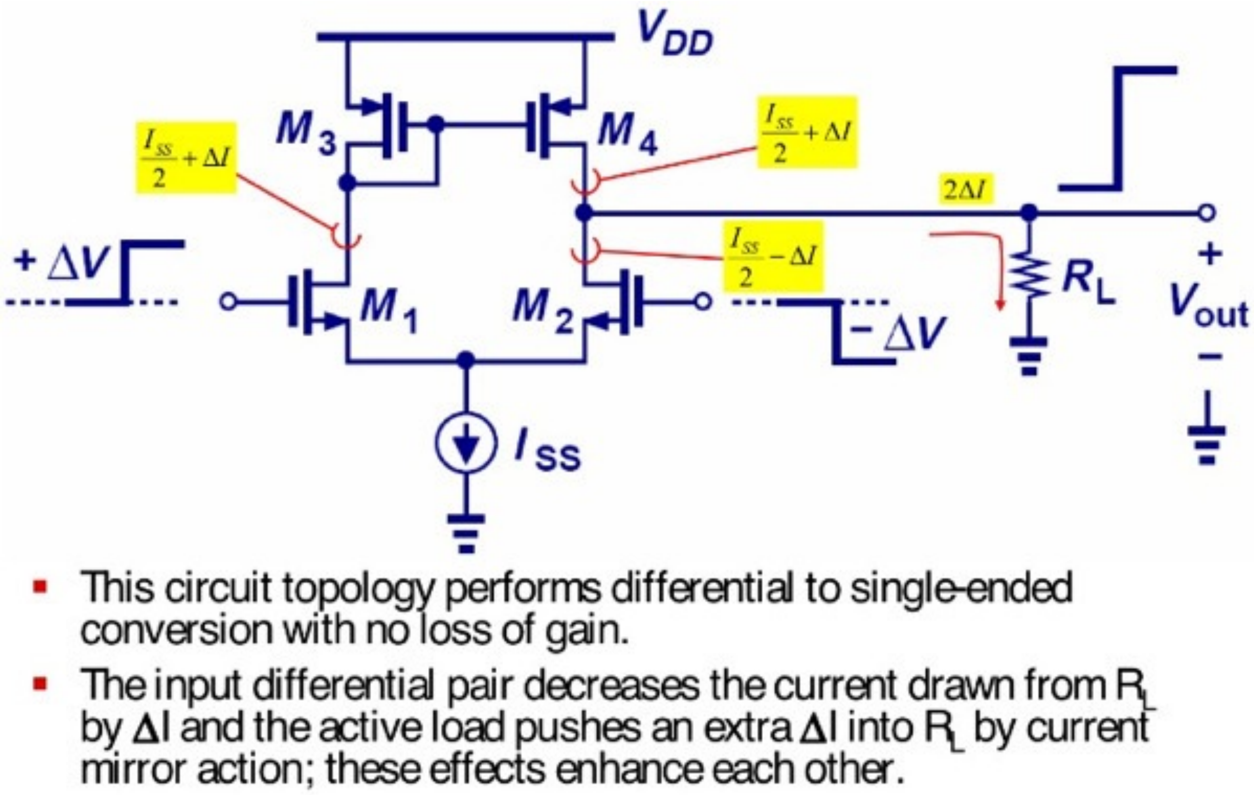
(Fig 14.1.1: single amplifier circuit)



(Fig 14.1.2: single amplifier transfer characteristic curves)

14.2 Amplifiers - Differential amplifiers

A differential amplifier can be obtained by combining two single-ended amplifiers:



(Fig 14.2.3: differential amplifier)

The mosfets M1 and M2 at the input stage make a very high input impedance, while the mosfet M4 makes a very low output impedance.

The circuit made of M3 and M4 is called "current mirror" and its function is to copy the drain current of M3 and setting an equal drain current in M4.

The M3 can be seen as a single transistor amplifier having M1 as variable source resistor. The operating point of M3, due to the negative feedback occurring because of its gate-to-source loop connection, will self adjust on the load line, whose angle depends on the equivalent resistance of M1, so that its drain current would be the only possible value for which its gate voltage matches its source voltage, since they are the same.
The M4 will then copy that current on the right arm of the circuit, hence the current in the R_L is controlled from both the M1, via the copied current, and the M2 according to the 1st

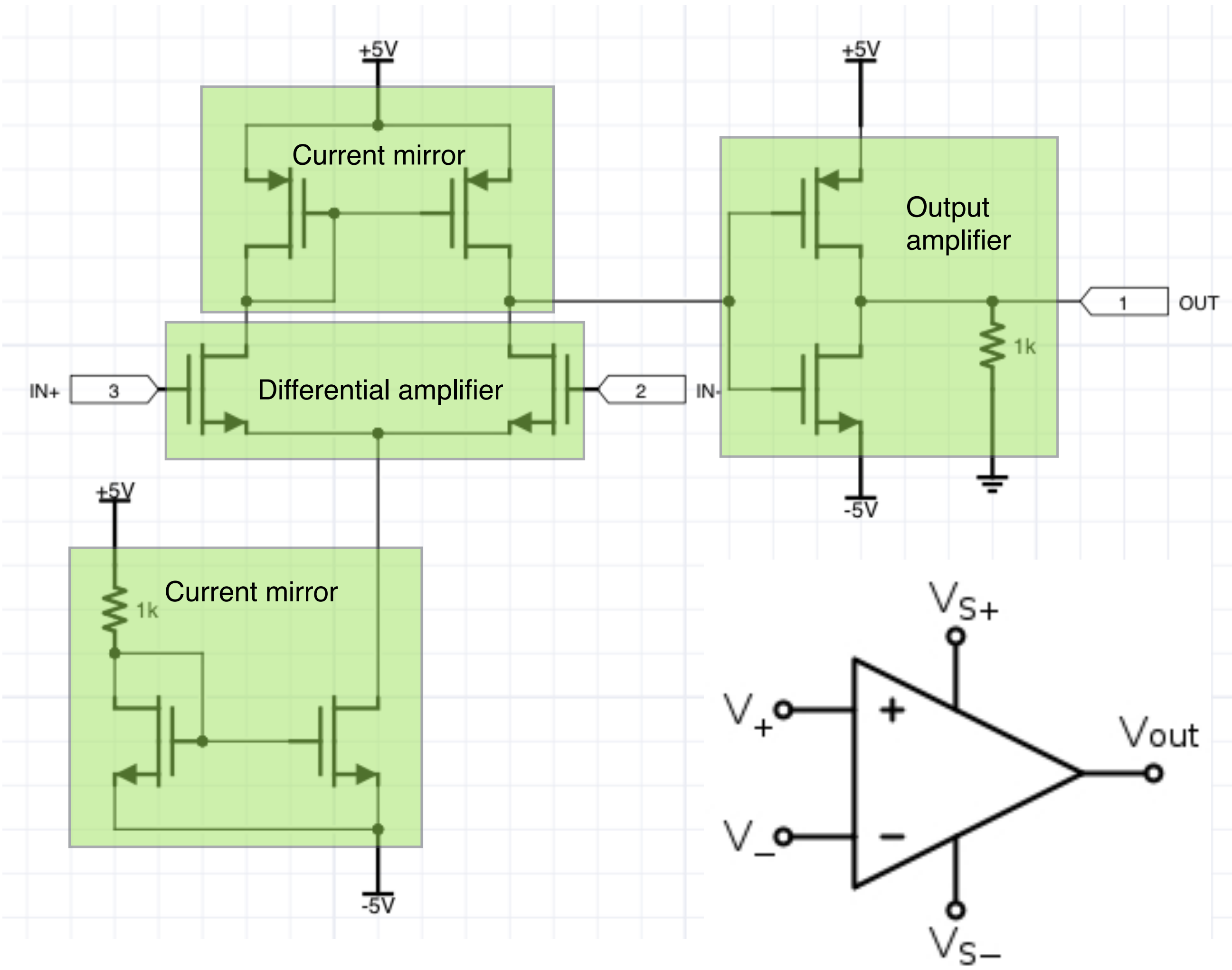
Kirchoff's law applied to the output node.

14.3 Amplifiers - Operational amplifiers

The operational amplifier is a differential amplifier with single-ended output and has got a very high gain (e.g. 10^8). It is also characterised by having a very high input impedance at both inputs and a quite low output impedance:

$$\begin{cases} A \longrightarrow +\infty \\ Z_{IN} \longrightarrow +\infty \\ Z_{OUT} \longrightarrow 0 \\ V_{OUT} = A\left(V_+ - V_-\right) \end{cases}$$

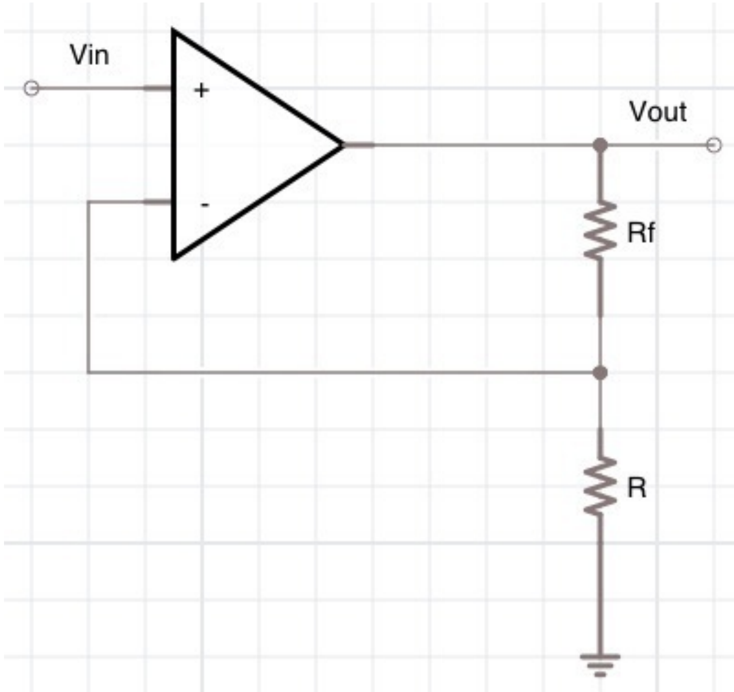
(Eq 14.3.1)



(Fig 14.3.1: operational amplifier internal diagram)

14.4 Amplifiers - Non-inverting op-amp configuration

The typical use of an operational amplifier is the following one:



(Fig 14.4.1: non-inverting op-amp configuration)

$$\begin{cases} V_{OUT} = A\left(V_+ - V_-\right) \\ V_+ = V_{IN} \\ V_- = \dfrac{R}{R+R_f} V_{OUT} \end{cases}$$

(Eq 14.4.1)

$$G = \frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\dfrac{R}{R+R_f}}$$

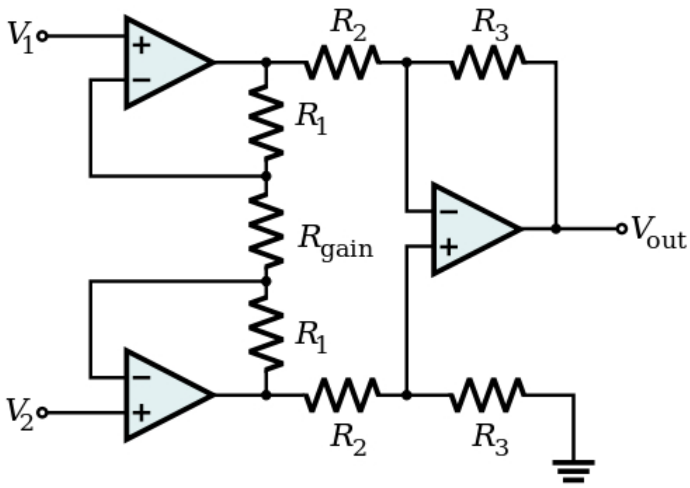
$$\lim_{A \to +\infty} G = \frac{R+R_f}{R} = 1 + \frac{R_f}{R}$$

(Eq 14.4.2)

Notice that the "closed loop gain" G does not depend on the "open loop gain" A, if the latter one is big enough.

14.5 Amplifiers - Instrumentation amplifiers

The ultimate topology of amplifier which is very important in order to amplify very weak signals in the "instrumentation amplifier":



(Fig 14.5.1: instrumentation op-amp)

It has the advantage of having a very high input impedance due the two input op-amps, it is differential, it is very linear due to the feedback and the output has a good low impedance that means it can drive a load without having the output voltage affected. The first two op-amps compute respectively GV_1 and GV_2 while the third one computes the difference of the two previous results, hence:

$$V_{OUT} = GV_1 - GV_2$$

(Eq 14.5.1)

The V_1 input is the called signal input, while V_2 is the reference input.

In neural recordings V_1 would be connected to the signal electrode, while V_2 would be connected to the reference electrode. Notice the ground is different from the reference: both signal and reference are measured *versus* ground, which should be at a stable potential.

This is particularly interesting to study the effect of the external noise cancellation:

$$V_{OUT} = G\left(V_{neuron} + V_{noise}\right) - G\left(\overline{V_{all\ neurons}} + V_{noise}\right) = G\left(V_{neuron} - V_{local}\right)$$

(Eq 14.5.2)

which is valid when both the signal and the reference electrodes are **close enough in order to pick up the same noise** and at the same time the reference electrode is collecting the time average of all the neural activities of the local population of neurons, which corresponds to the "local field potential".

15.1 Basic digital circuits - Boolean algebra

The Boolean algebra is a set of rules and theorems describing mathematical operations in the binary numeric system. In this system numbers are made by sequences of only two different symbols: "0" and "1".

Here is an example of binary-to-decimal numeric conversion:

$$11001 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 16 + 8 + 0 + 0 + 1 = 25$$

(Eq 15.1.1)

and here is an example of decimal-to-binary numeric conversion by successive integer divisions:

$$\frac{25}{2} = 12\Big|_{R=1}, \frac{12}{2} = 6\Big|_{R=0}, \frac{6}{2} = 3\Big|_{R=0}, \frac{3}{2} = 1\Big|_{R=1}, \frac{1}{2} = 0\Big|_{R=1}$$

(Eq 15.1.2)

The backward sequence of successive remainders gives the binary representation of the decimal number.

The fundamental operations are "AND", "OR" and "NOT":

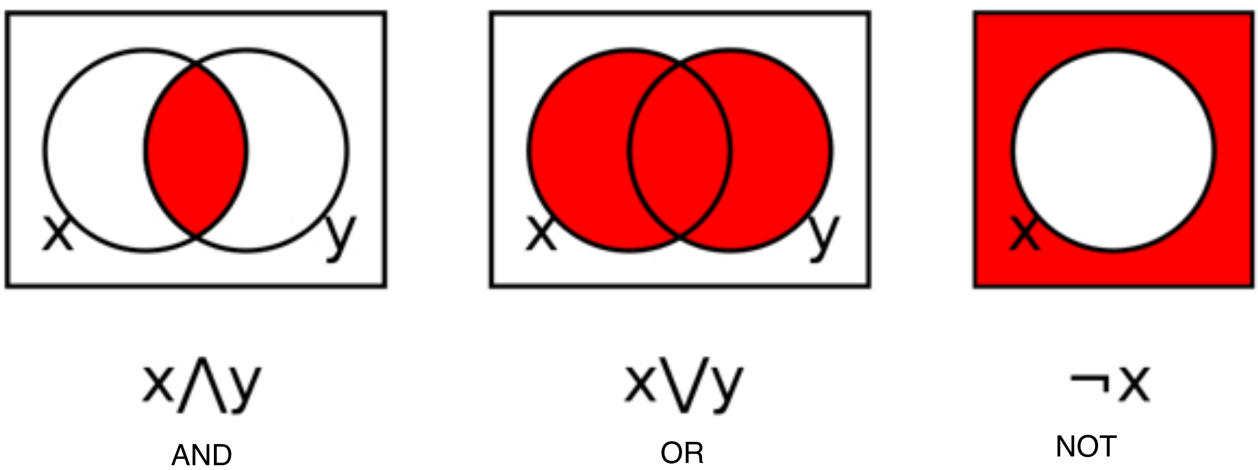
(Fig 15.1.1)

and the fundamental relations between those operations are the "De Morgan's laws":

NOT(X AND Y) = (NOT X) OR (NOT Y)

NOT(X OR Y) = (NOT X) AND (NOT Y)

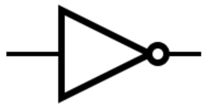
(Eq 15.1.2)



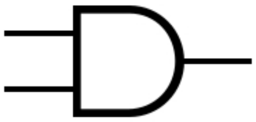
(Fig 15.1.1: fundamental logic operations)

15.2 Basic digital circuits - Logic gates

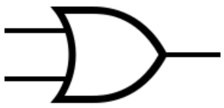
The fundamental digital operations can be implemented by electronic circuits known as "logic gates", whose behaviour is described by "truth tables":



INPUT	OUTPUT
X	NOT(X)
0	1
1	0



INPUTS	OUTPUT
X,Y	AND(X,Y)
00	0
01	0
10	0
11	1

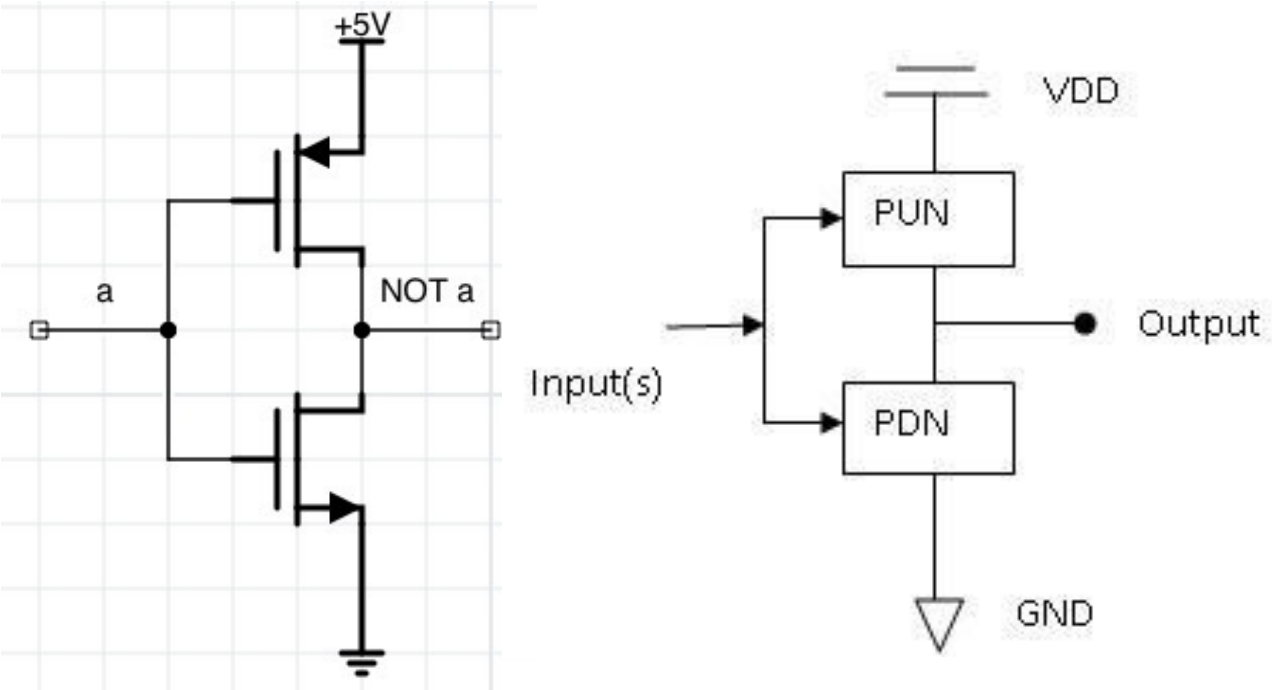


INPUTS	OUTPUT
X,Y	AND(X,Y)
00	0
01	1
10	1
11	1

(Fig 15.2.1: basic logic gates truth tables)

15.3 Basic digital circuits - NOT gate

Logic gates are implemented using a complementary circuit topology.
A "pull-up" network (PUN) is connected to a "pull-down" network (PDN), making a so called "CMOS": "Complementary MOS (Metal-Oxide Semiconductor)" circuit topology:

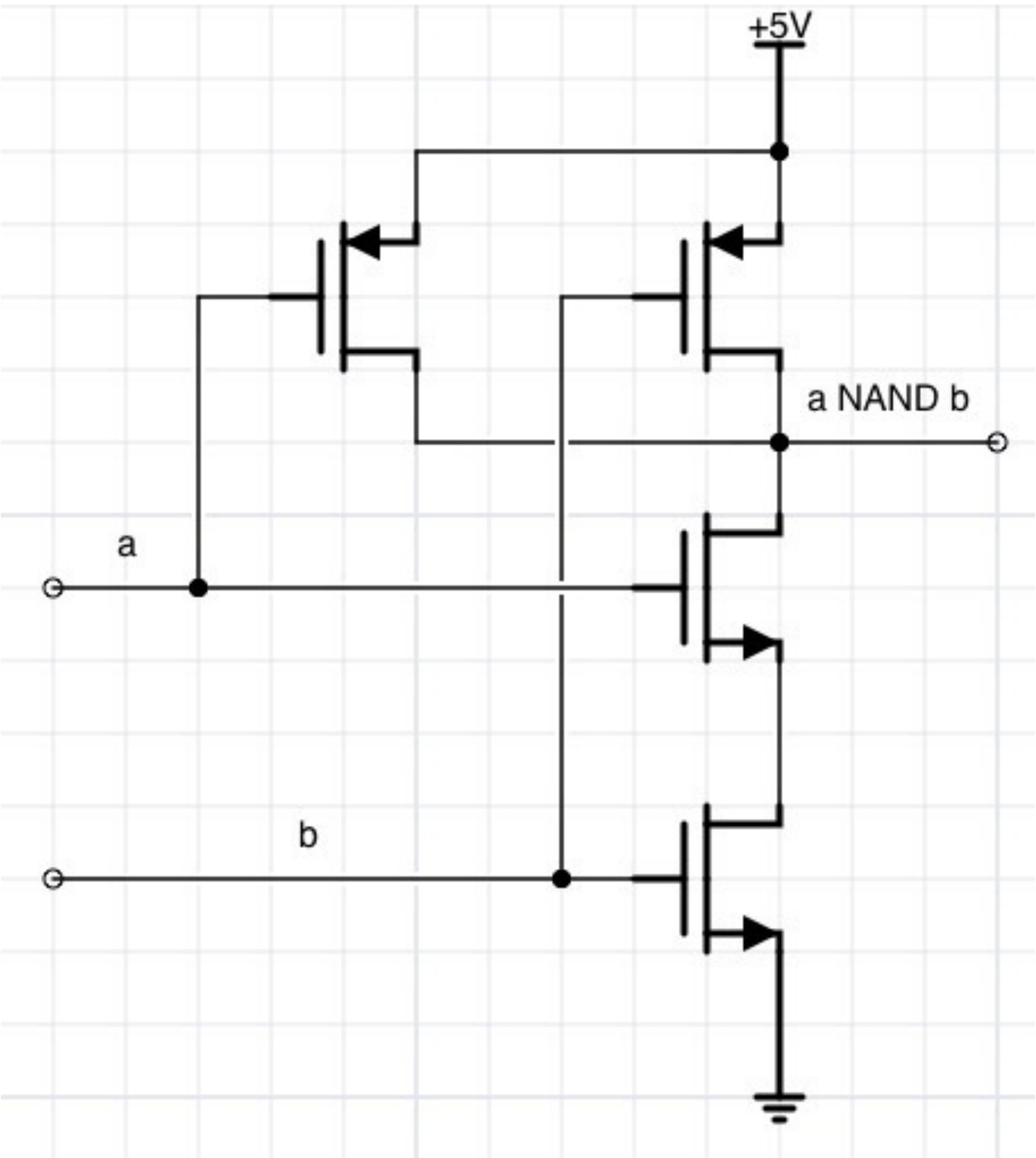


(Fig 15.3.1: the NOT gate)

The NOT gate is also called inverter.

15.4 Basic digital circuits - NAND gate

Within the CMOS technology it is not possible to directly implement an AND gate by using the PUN and PDN circuit topology. Only "negative logic" can be implemented this way:

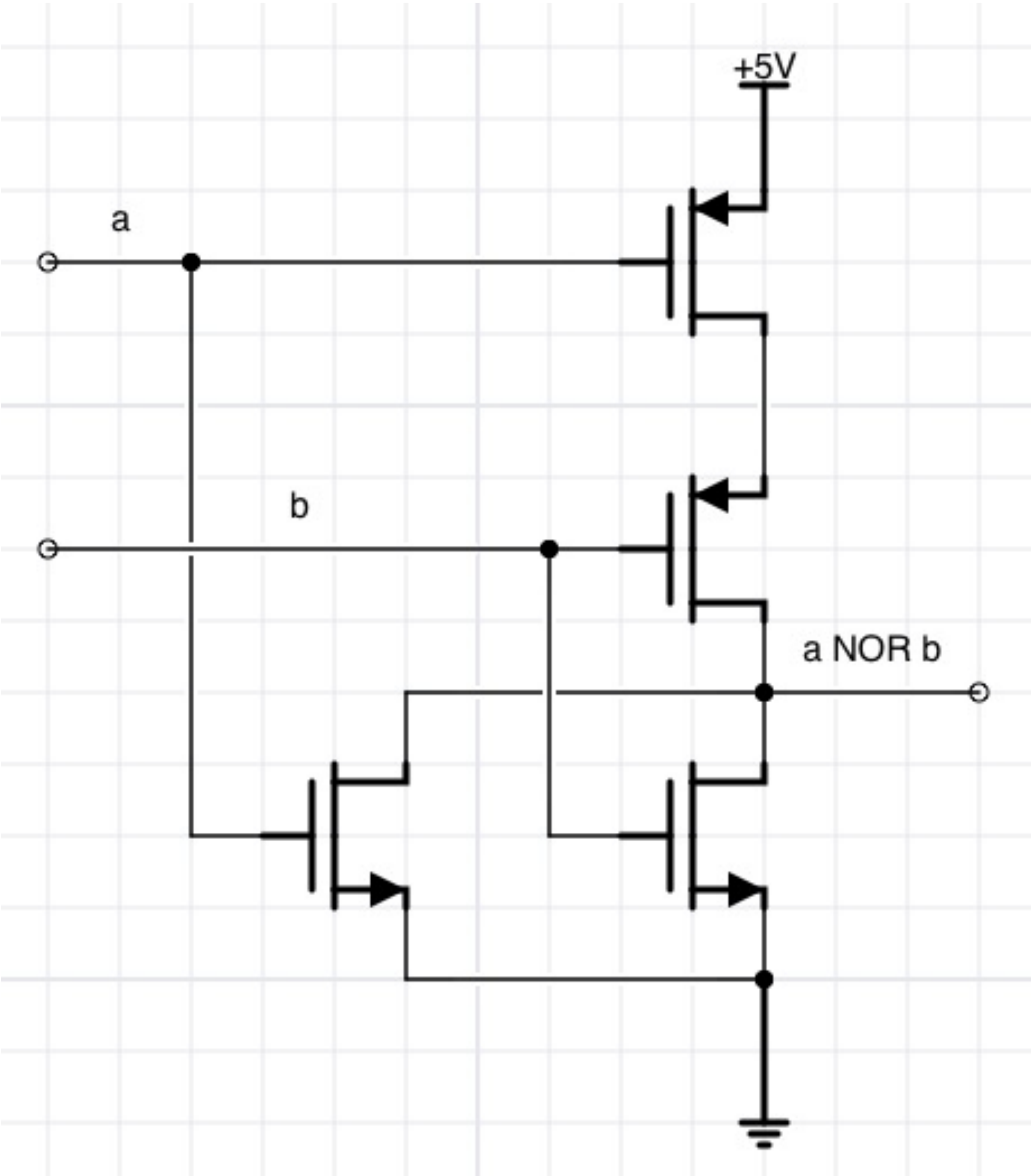


(Fig 15.4.1: the NAND gate)

Anyway, an AND gate can be implemented by a cascade of a NAND gate and NOT gate.

15.5 Basic digital circuits - NOR gate

Similarly to the NAND gate, in CMOS technology only a NOR gate can be implemented:



(Fig 15.5.1)

Similarly, an OR gate can be implemented by a cascade of a NOR and a NOT gate.

16.1 Advanced digital circuits - Combinational logic

Combinational logic circuits are digital circuits where the status of the outputs depends only on the status of the inputs and not on the time. They can be synthesised by writing the truth table:

sensor inputs

A	B	C	Output	
0	0	0	0	
0	0	1	0	
0	1	0	0	
0	1	1	1	$\overline{A}BC = 1$
1	0	0	0	
1	0	1	1	$A\overline{B}C = 1$
1	1	0	1	$AB\overline{C} = 1$
1	1	1	1	$ABC = 1$

$\text{Output} = \overline{A}BC + A\overline{B}C + AB\overline{C} + ABC$

(Fig 16.1.1: truth table of a combinational circuit)

The logic function for the output can be obtained by summing the "minterms", which are built by reading the rows of the truth table corresponding to a "1" output, negating the variables which are "0", and multiplying them together:

$$Y = \overline{A}BC + A\overline{B}C + AB\overline{C} + ABC$$

(Eq 16.1.1)

Alternatively, the same function can be obtained by multiplying the "maxterms", which are built by reading the rows of the truth table corresponding to a "0" output, negating the variables which are "1" , and adding them together:

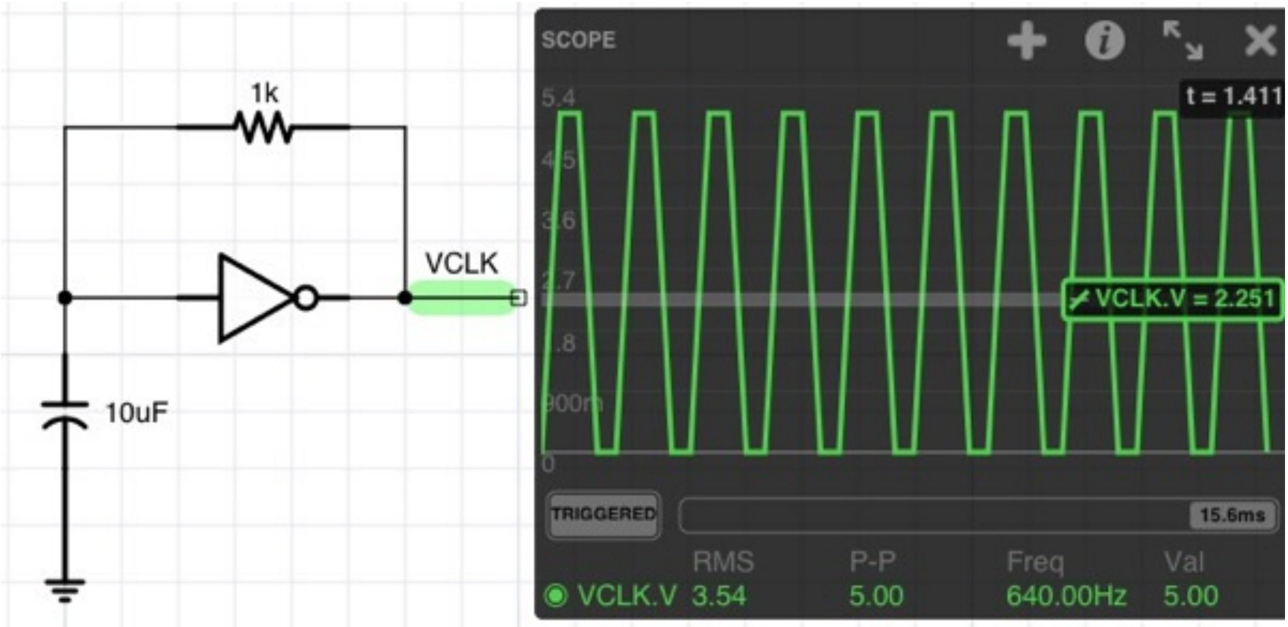
$$Y = \left(A+B+C\right)\cdot\left(A+B+\overline{C}\right)\cdot\left(A+\overline{B}+C\right)\cdot\left(\overline{A}+B+C\right)$$

(Eq 16.1.2)

16.2 Advanced digital circuits - Sequential logic

Sequential logic circuits are circuits whose status of the outputs depends on the status of the inputs and also the time. They are characterised by feedback connections:

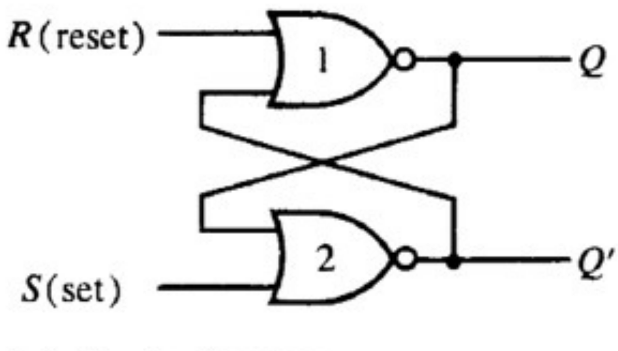
e.g. THE CLOCK OSCILLATOR



(Fig 16.2.1: the oscillator)

The capacitor is charged by the output of the NOT gate. When the voltage across the capacitor is enough this makes the output changing state, making the capacitor discharging. When the voltage across the capacitor is then low enough the output changes state and a new cycle begins.

e.g. THE FLIP-FLOP



(a) Logic diagram

S	R	Q	Q'	
1	0	1	0	
0	0	1	0	(after $S = 1, R = 0$)
0	1	0	1	
0	0	0	1	(after $S = 0, R = 1$)
1	1	0	0	

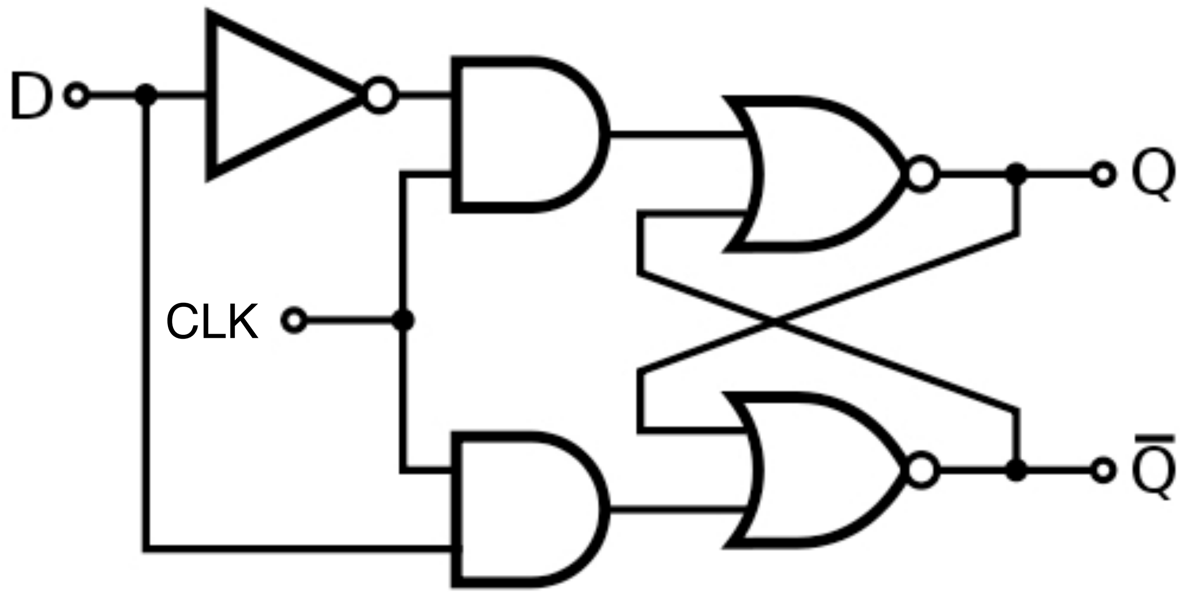
(b) Truth table

(Fig 16.2.2: the SR flip-flop)

The flip-flop is one of the simplest RAM - Random Access Memory.

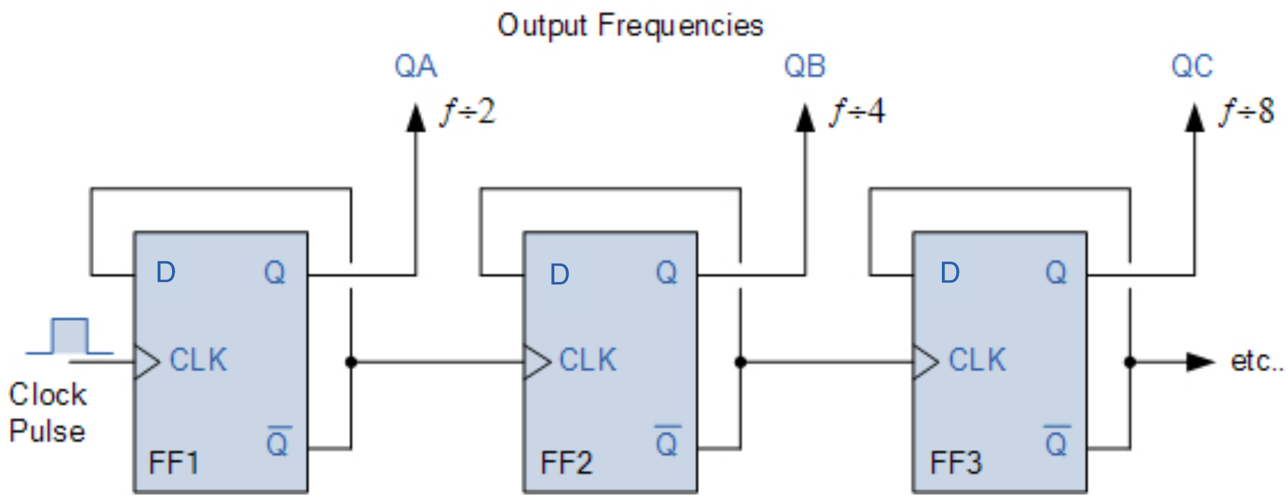
16.3 Advanced digital circuits - Counters

The D-type flip-flop latches the value of the input data (D) into the output (Q) every time there is a clock (CLK) cycle:



(Fig 16.3.1: the D-type flip-flop)

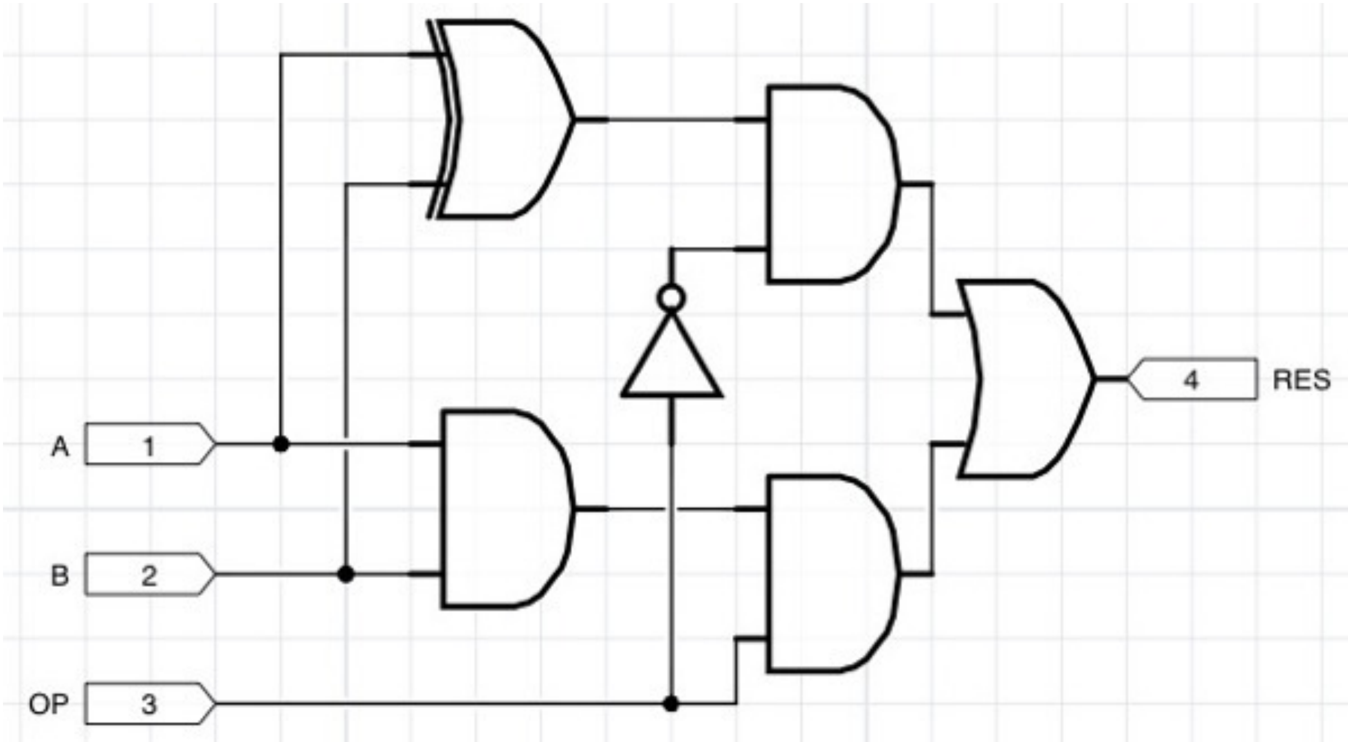
By cascading flip-flops in a closed loop a binary counter is obtained:



(Fig 16.3.2: the binary counter)

16.4 Advanced digital circuits - Arithmetic Logic Unit

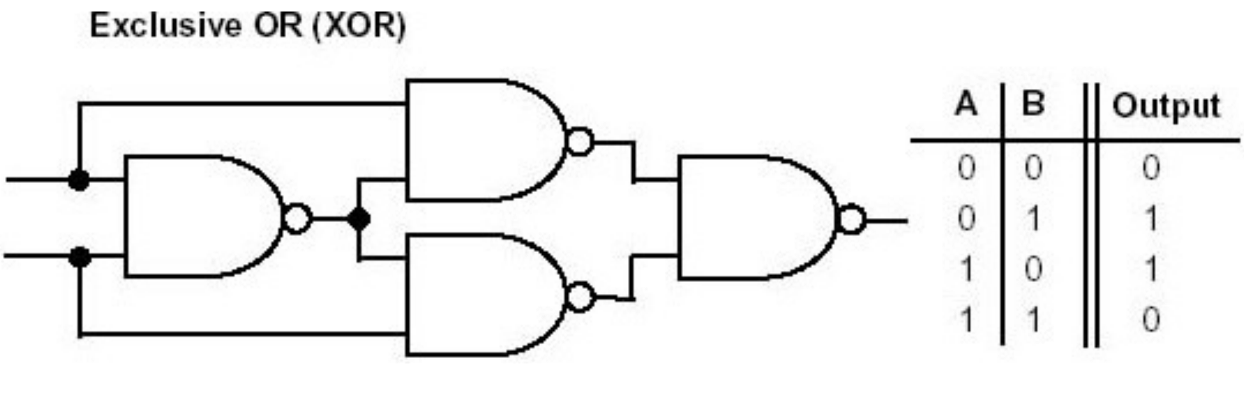
An "Arithmetic Logic Unit - ALU" is a circuit which takes "operands" and "operations" as inputs and produces an output corresponding to the evaluation of the operands according to the selected operation:



(Fig 16.4.1: a basic arithmetic logic unit)

The inputs A and B feed an AND gate, which performs a two-bit binary multiplication, and also a XOR gate, which performs a two-bit binary addition (without carry). The "OP" input selects either one of the two other AND gates in gating configuration, making either the multiplication or the addition result passing at the "RES" output.

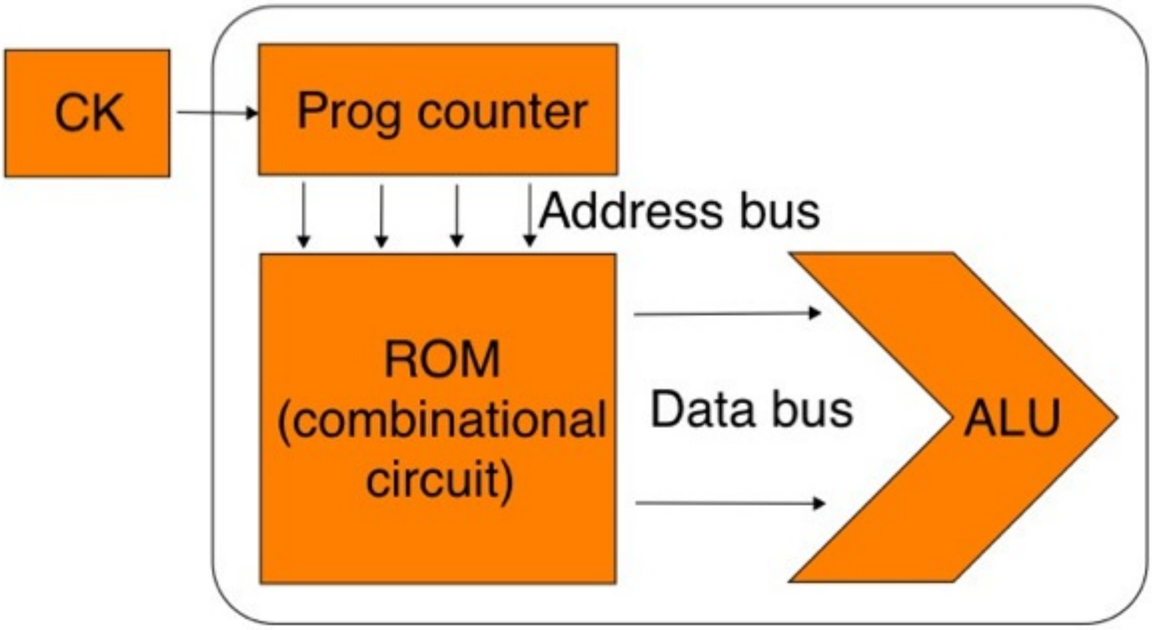
This is just a very simple example which explains how ALU work. Notice also the XOR gate:



(Fig 16.4.2)

16.5 Advanced digital circuits - Central Processing Unit

The "Central Processing Unit - CPU" is a device able to perform computation by reading instructions made of operands and operators from a memory in form of an indexed list called "program".



(Fig 16.5.1: CPU model)

A clock oscillator makes a program counter incrementing an address. A combinational circuit takes the address lines as inputs and produces an output data flow according to boolean functions which depends on the information coded by the program. The arithmetic logic unit hence decodes the operands and the operators from the data flow and selects the appropriate combinational, or sequential, circuits which implements the operation. The final output can be in principle read from a display device or stored back to a memory, according to the instruction written in the program.

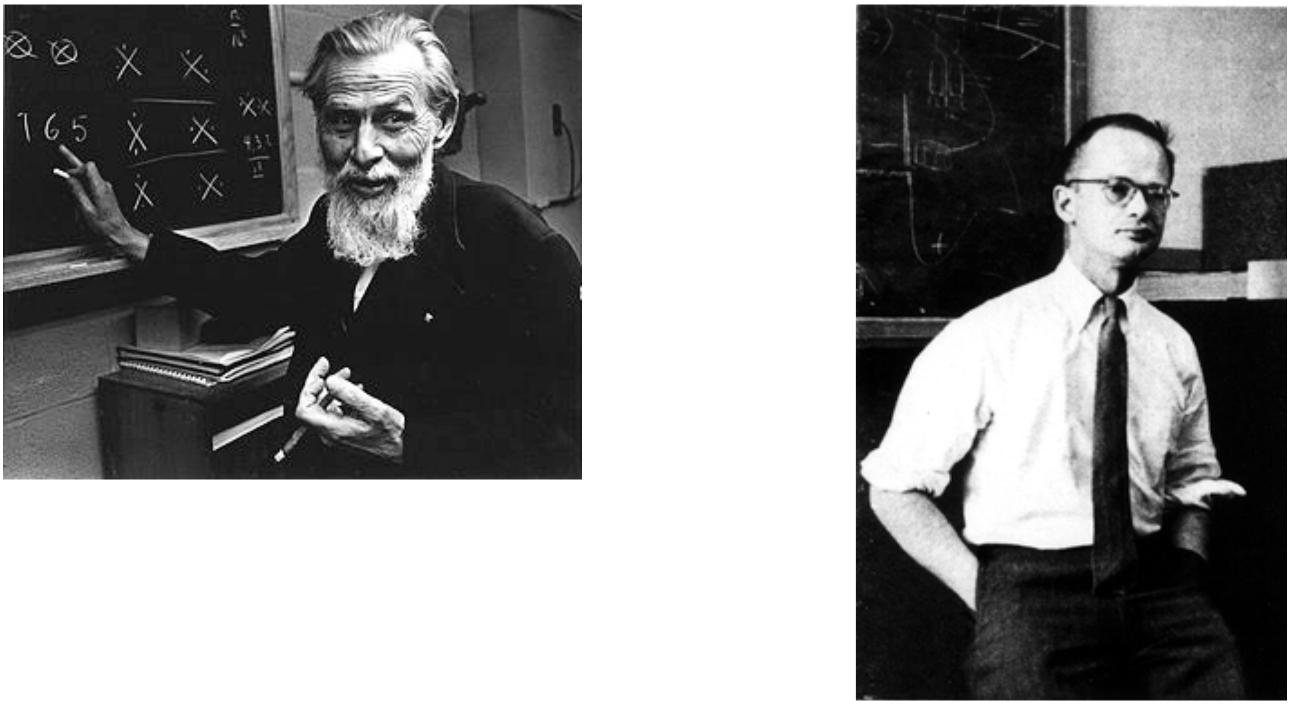
In the example above a very simple CPU architecture has been described, which is hereby only able to read a program from a ROM - Read Only Memory and calculate simple operations while outputting the results in form of serial stream of binary data.

More generally, the circuit resulting from of the application of both combinational and sequential circuits is a so called "state machine".

The status of a state machine is fully described by the instantaneous value of all its boolean variables.

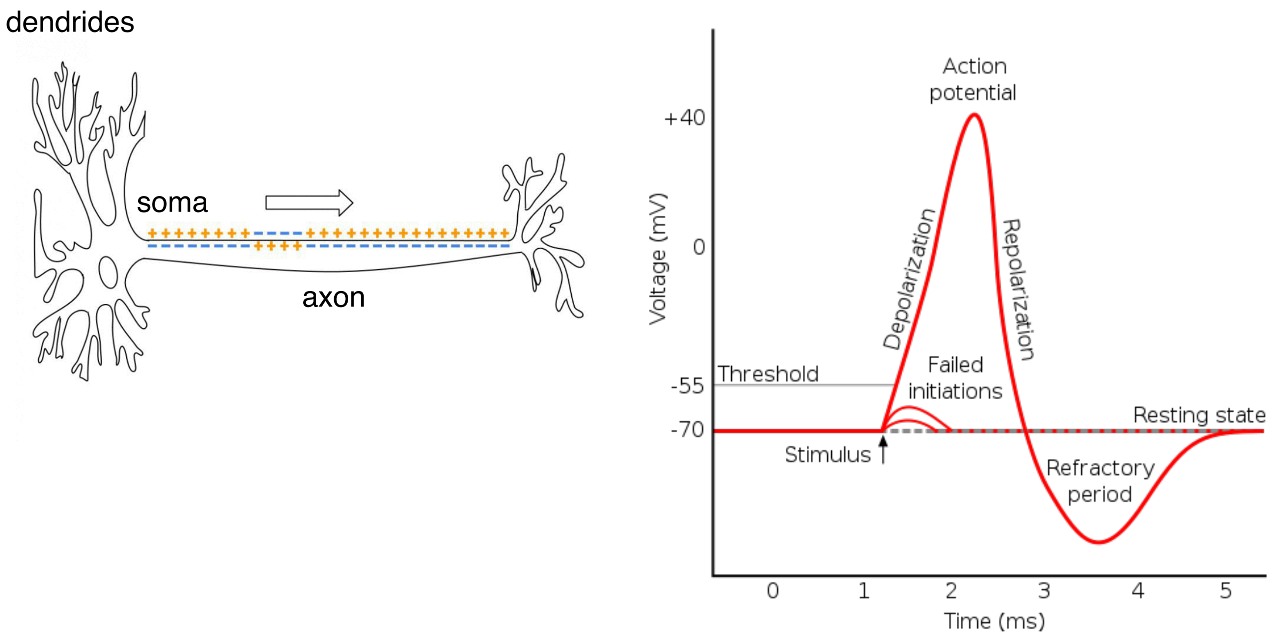
17.1 Neural networks - The action potential

The modern era of the neural networks began with the pioneer work of McCulloch and Pitts (1943).



(Fig 17.1.1: McCullock -left- and Pitts -right-)

They described a neuron as "device" having inputs and outputs and they recognised the "action potential" as the elementary piece of information building the neural signals.



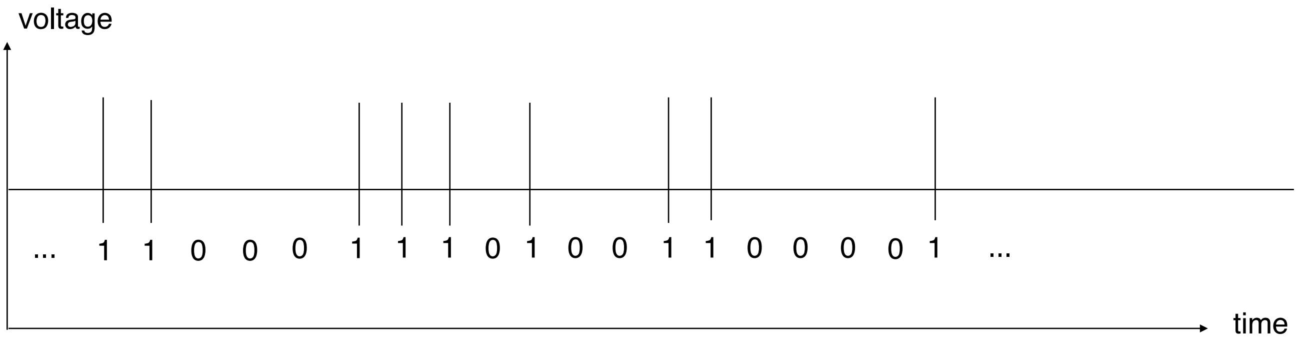
(Fig 17.1.2: a neuron and an action potential)

17.2 Neural networks - Modelling an artificial neuron

The McCullock-Pitts model (MCP) consists in both the description of an action potentials in terms of time-varying variables and the neural networks as circuits of neurons.

The action potential has been described in the MCP as an event: something which can be present or not present at a given time instant. A boolean variable keeps tracks of its presence as time goes by: if the action potential is not present then value of the variable is "0", if it is present the value is "1".

This way a boolean variable can effectively describe a neural spike train:



S(t) = ...1100011101001100001...

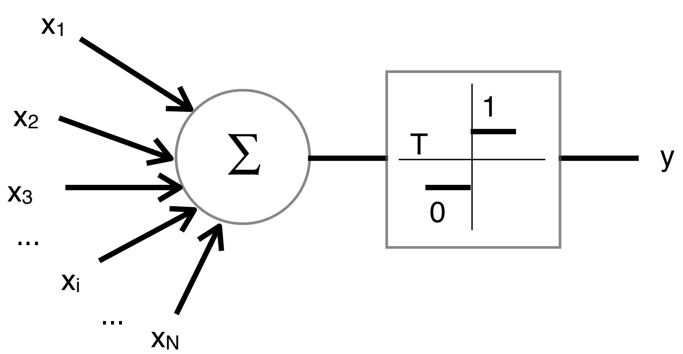
(Fig 17.2.1: neural spike train)

Notice the time can be "discrete", because the neurons do not fire again before their refractory period has passed: hence the spikes cannot overlap each other over time.

The neuron's "dendrides" has been recognised as the inputs of the system: they take the value of the variable the see and they multiply it for a "weight" coefficient.

The neuron's "soma" has been recognised as a functional centre which summates all the weighted inputs.

The neuron's "axon" has been recognised as the output of the system: it generates an action potential according to the comparison of the result of the summation operated by the soma and an internal "threshold", as follows:

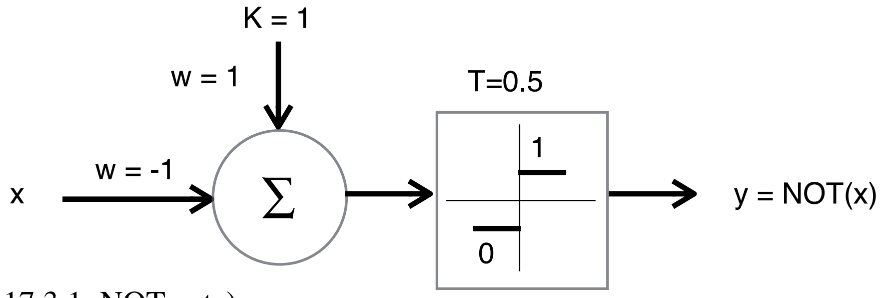


$$\begin{cases} S = \sum_i w_i x_i \\ y = \begin{cases} 0, \text{ if } S < T \\ 1, \text{ if } S \geq T \end{cases} \end{cases}$$

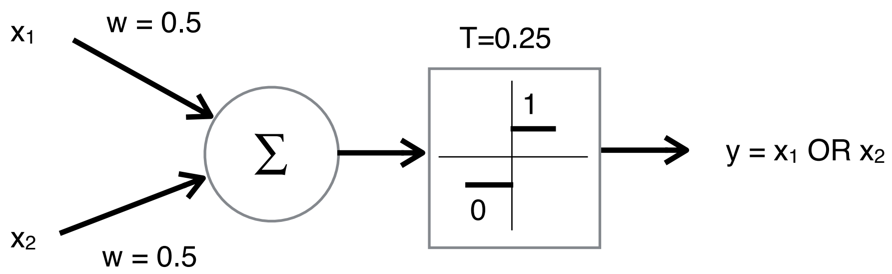
(Fig 17.2.2: MCP neuron) (Eq 17.2.1)

17.3 Neural networks - Neural boolean logic

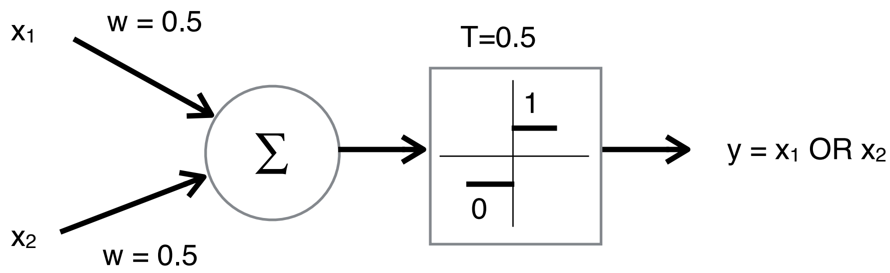
By choosing the appropriate weights and thresholds, neutrons can implement logic gates:



(Fig 17.3.1: NOT gate)



(Fig 17.3.2: OR gate)



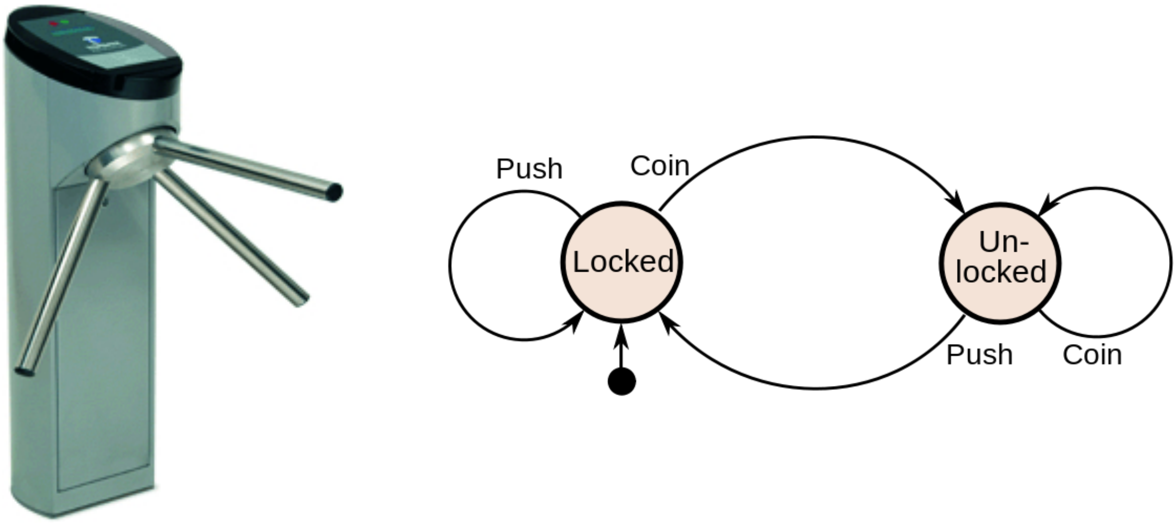
(Fig. 17.3.3: AND gate)

Notice that in the NOT gate, the weight attached to the "x" input is negative: this corresponds to an "inhibitory" input. The "K" input is an input which is constant in time and equal to "1": this means a continuously spiking train has been attached to that input.

17.4 Neural networks - State machines

As a result of the neural boolean logic, neural networks are in fact an implementation of state machines. More precisely it has ben proved (Kleene - 1956, Minsky - 1967 and Kremer - 1995) that "every state-machine is equivalent to and can be implemented by some neural network".

An example of state machine is, for instance, a turnstile:



(Fig 17.4.1: a turnstile and its corresponding state diagram)

In this example the behaviour of the turnstile has been described by a "state diagram". Starting from the black dot the machine goes in the "locked" state. Pushing the turnstile will not make the machine exiting from that state, unless a coin is inserted. Then the machine goes to the "unlocked" state. Putting other coins will not make the machine exiting from the latter state but pushing the turnstile will allow the person to pass and will make the machine returning to its original "locked" state.

This can also be described in terms of its "state transition table":

Current state	Input	Next state	Ouput
Locked	Coin	Unlocked	Unlock turnstile so customer can push through
	Push	Locked	None
Unlocked	Coin	Unlocked	None
	Push	Locked	When customer has pushed through lock turnstile

(Fig 17.4.1: state transition table of a turnstile)

17.5 Neural networks - Neural network implementation

According to the turnstile example, the same corresponding state machine can be implemented in different ways. At least two of them are interesting:

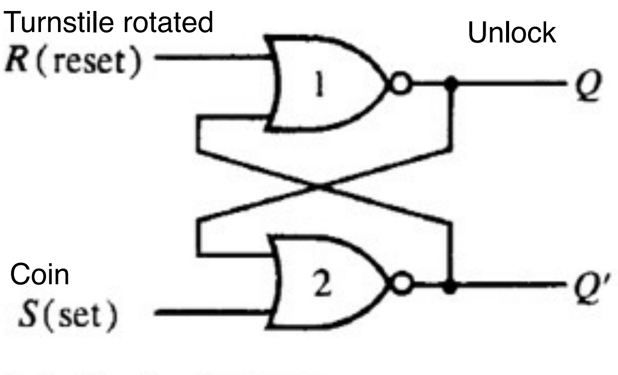
- ALGORITHMIC APPROACH:

The following program (here showed in "pseudo-code" language) might be interpreted by a CPU system:

```
1:	state = locked;
2:	turnstile = NOT(pushed);
3:	rotate = FALSE;
4:	IF (coin inserted) THEN
5:		(state =  NOT(locked));
6:	ELSE
7:		(state = locked);
8:	IF (turnstile = pushed AND state = NOT(locked)) THEN
9:		rotate = TRUE;
10:		state = locked;
11:	ELSE
12:		rotate = FALSE;
13:	GOTO(4);
```

- ARCHITECTURAL APPROACH:

The same problem can be solved by the following circuit:



(a) Logic diagram

S	R	Q	Q'	
1	0	1	0	
0	0	1	0	(after $S = 1, R = 0$)
0	1	0	1	
0	0	0	1	(after $S = 0, R = 1$)
1	1	0	0	

(b) Truth table

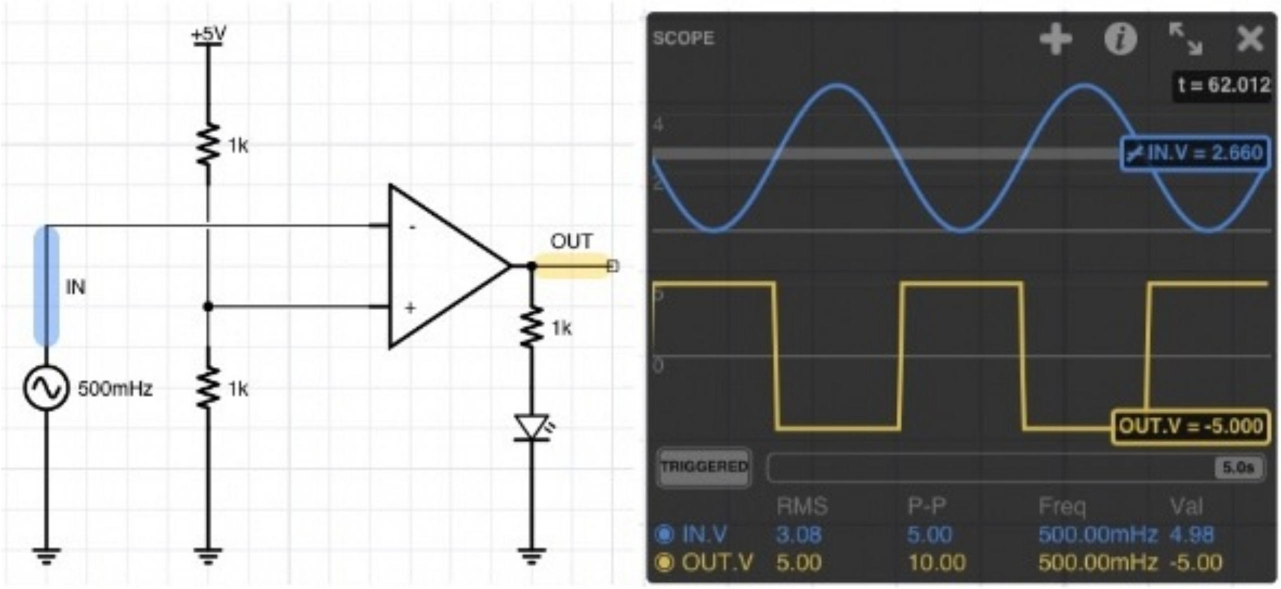
(Fig 17.5.1: a S_R flip-flop solving the turnstile state machine)

It is the S-R flip-flop. Compared to the algorithmic solution the architectural one is much more efficient in terms of components used for the implementation: this is not necessarily true and it introduces the topic of "optimisation" of computation.

18.1 Digital signal processing - Comparators

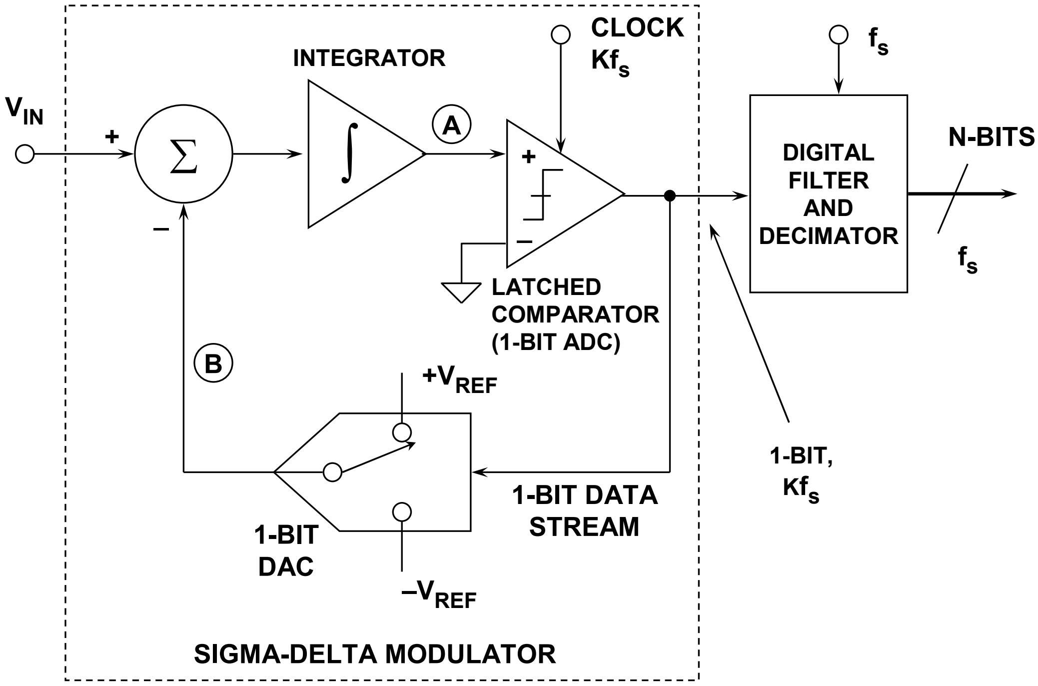
The comparator circuit compares an analog input signal with a programmable threshold (by properly settings the value of the resistors in the voltage divider), and gives at the output a digital signal which status indicates whether the analog signal is above or below the threshold:

(Fig 18.1.1)



18.2 Digital signal processing - Analog to digital converters

The analog-to-digital conversion is the procedure of transforming an analog time varying signal into a digital time stream of binary numbers corresponding to discrete voltage levels of the analog signal. This operation is possible through the use of special circuits called ADC - Analog to Digital Converters. There are many different types of ADC circuits, but the most important nowadays is the "Sigma-Delta".



(Fig 18.2.1: sigma-delta ADC)

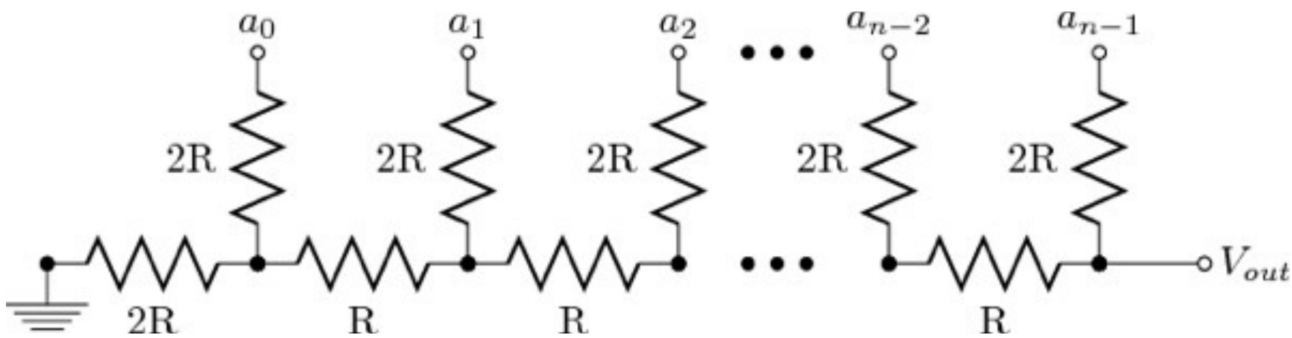
It works as follows:
1. Assume there is a certain almost constant voltage input V_{IN} during a period of CLOCK.
2. The output "A" of the integrator is then ramping up or down, according to the polarity of V_{IN}. The ramping speed depends on the magnitude of V_{IN}.
3. The output of the 1-bit comparator is fed back and controls a switch (a 1-bit DAC).
4. The output voltage "B" of the switch jumps up or down according to the comparator's output: this makes the average value of "B" to be equal to the constant input V_{IN}.
5. The summing node subtracts "B" from V_{IN}, the latter voltage is then averaged by the integrator
6. Because the averaging (integration) is a linear operation, the average of the subtraction is equivalent to the subtraction of the average: hence the result in "A" is the "error voltage" $V_{ERR} = V_{IN}$ - AVG(V_{IN}).
7. V_{IN} is almost constant in CLOCK period, hence V_{ERR} is small and stays around zero, which is the voltage versus which the comparators compares.
8. The 1-bit serial data stream at the output of the comparator is therefore a digital representation, in terms of density of "1s" and "0s", of the AVG(V_{IN}) during each CLOCK period: there are more "1s" when V_{IN} is close to V_{REF} and more "0s" when V_{IN} is close to -V_{REF}.
9. The binary data stream is then numerically processed by a CPU system.

So, even if you have only 1 bit available but you are fast enough then you can still do a lot of things: never underestimate the power of a bit!

18.3 Digital signal processing - Digital to analog converters

The digital-to-analog conversion is the procedure of transforming a digital time stream of binary numbers into an analog time varying signal made of discrete voltage levels proportional to the numeric values in the digital time stream. This operation is possible by using DAC - Digital to Analog Converters circuits.

The simplest DAC circuit is the R-2R resistor ladder network:



(Fig 18.3.1: R-2R ladder based DAC)

Assuming the binary number is the N-bit number:

$$number = \left| a_0, a_1, \ldots, a_{n-1} \right\rangle$$

(Eq 18.3.1)

then the voltage output value is:

$$V_{out} = V_{ref} \frac{number}{2^n}$$

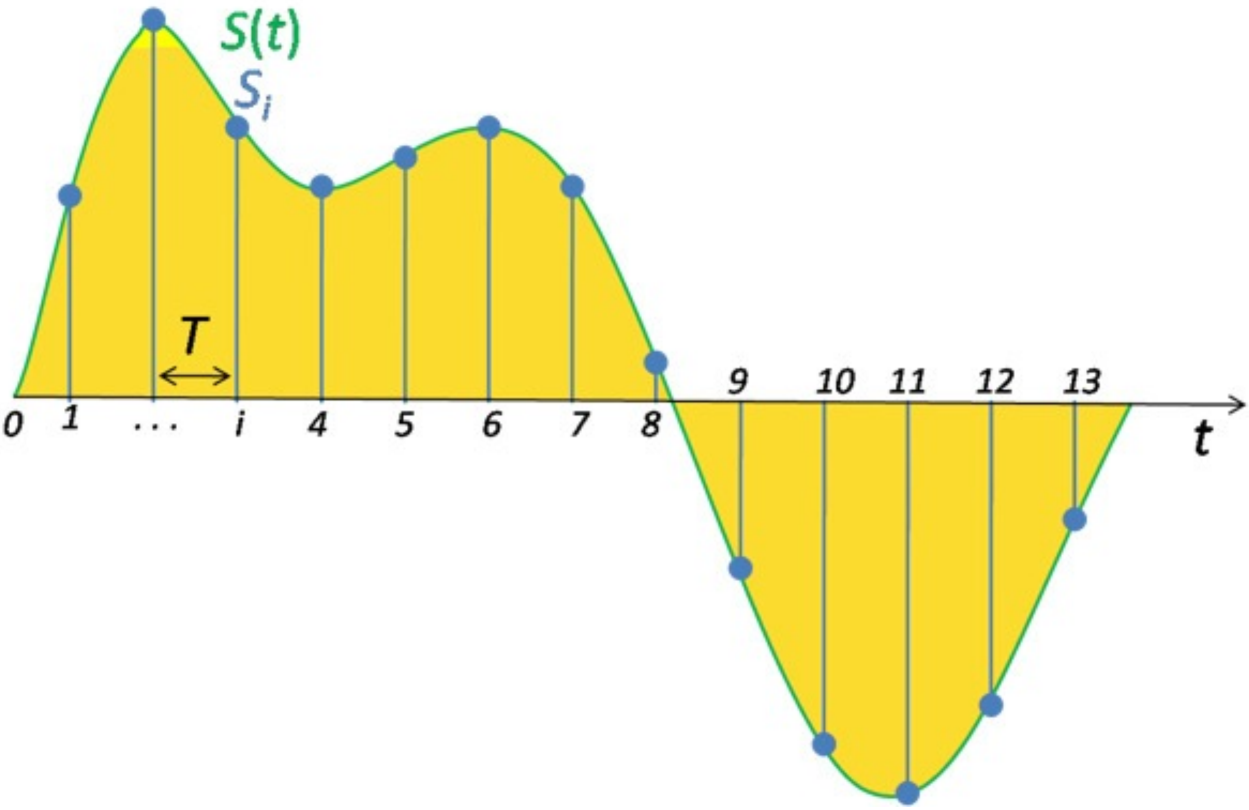
(Eq 18.3.2)

where V_{REF} is a fixed value corresponding to the voltage level of each bit.

18.4 Digital signal processing - Sampling

Both ADCs and DACs circuits implement a signal conversion in the time domain.

However, such kind of conversions cannot be done continuously in time, because this would require an infinitive amount of information stored either in analog or digital form. Since this it not physically possible, signals have to be approximated by a time-sampling:



(Fig 18.4.1: sampling)

Notice the time sampling (horizontal axis) is not related to the quantisation of the sampled value (vertical axis).

The vertical axis quantisation corresponds, in the analog world, the smallest amount of value the instrument (e.g. signal amplifier) can resolve. In the digital representation this corresponds to the encoding of the sampled value represented as a number having a certain number of bits.

The horizontal axis quantisation, namely the time-sampling, sets the "sampling rate".

The "data rate" is the amount of data per unit of time which corresponds to information the system has to store, either in analog or digital form.

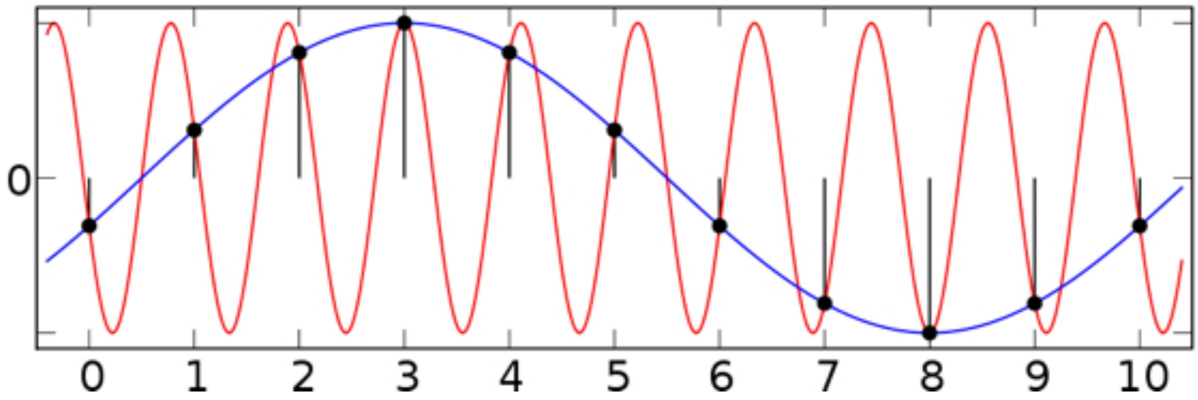
e.g.
input range = -10V...+10V divided in 4096 intervals --> resolution = (10 - (-10))/4096 = 4.88mV
number of bits per sample = log2(4096) = 12bits
sampling time = 50us --> sampling frequency = 1/50us = 20kHz = 20000 samples/s
data rate = 12bits * 20000 samples/s = 240000 bits/s --> 240000/8 = 30kbytes/s

18.5 Digital signal processing - Aliasing

In order to avoid artefacts in their approximate representation, signals must be sampled according to their bandwidth and the sampling theorem:

the sampling frequency must be at least twice the bandwidth of the signal

If not, a phenomena called aliasing will occur:



(Fig 17.5.1: aliasing)

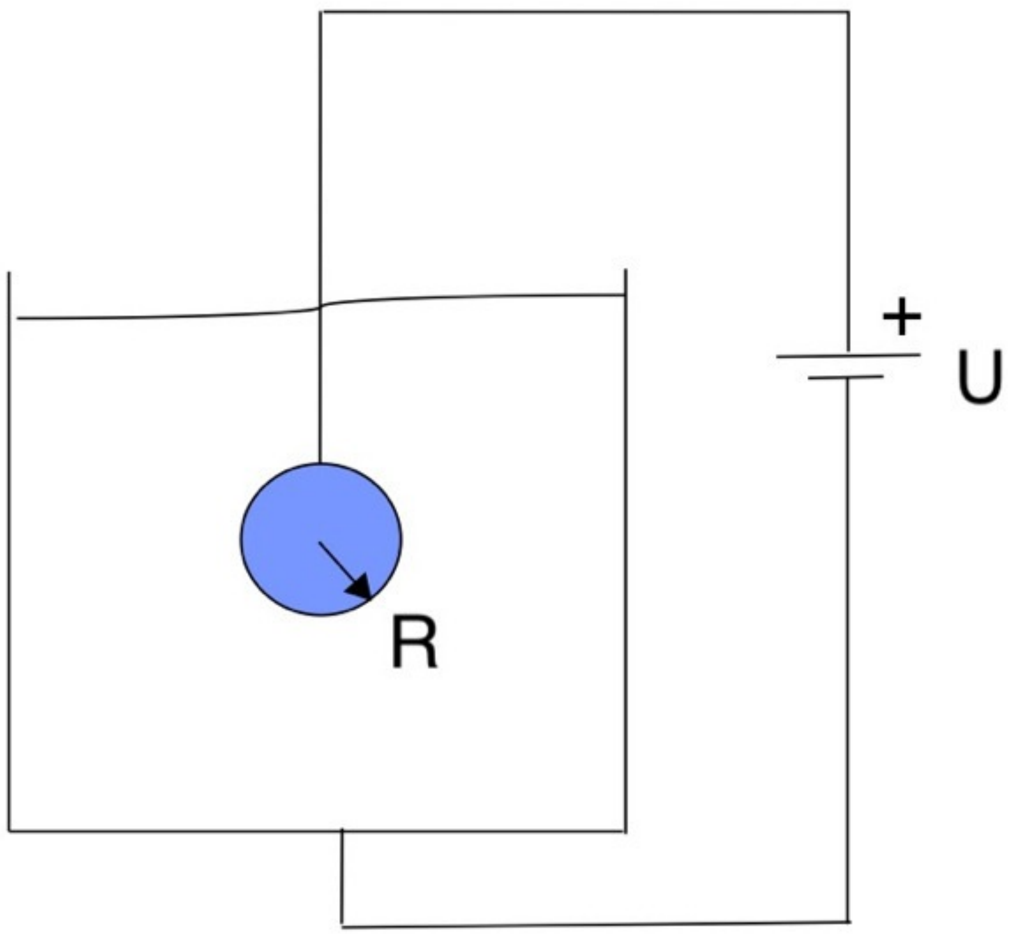
In this case a high frequency (red) signal has been sampled with a too low sampling frequency. The effect is that a reconstruction of the sampled signal, via interpolation of the collected samples, will result in a wrong low frequency (blue) signal.

Notice that, in order to be correctly sampled, the signal of interest must have a limited frequency bandwidth. Only if this is true then the correct sampling frequency can be chosen.
If not, as usually when working with physical signals, the bandwidth of the signal must be limited before sampling it by using an appropriate analog low-pass filter, called anti-aliasing filter.

The same for the analog reconstruction of a digital time sampled stream of data: the analog output signal must be filtered with a similar analog low-pass filter, according to the sampling.

19.1 Electrodes - Modelling an electrode

Metal electrodes immersed in an electrolyte fluid are characterised by having a certain impedance. In order to evaluate it let us consider a simple model, where the cathode of battery is connected to a perfect conductor electrode through a perfect conductor
wire and the electrode is a sphere having radius R. The anode of the battery is connected to a metal container, much larger than the dimension of the electrode and whose walls are far from it filled with an electrolyte having dielectric constant ε and conductivity g:



(Fig 19.1.1: simple model of an electrode)

The wire connecting the battery to the sphere is supposed to be insulated from the fluid.

19.2 Electrodes - Electric field

Due to the assumptions made in the simplified model, the electric field in the proximity of the electrode can be obtained as a function of the radial distance r by applying the Faraday's law to the conductive sphere:

$$E(r)=\frac{Q}{4\pi\varepsilon r^{2}}$$

(Eq 19.2.1)

Since the electric field is, by definition of potential, opposite to the gradient of its potential it is possible to calculate a relation between them:

$$E(r)=-\frac{dV}{dr}\Leftrightarrow V(r)=-\int_{r}^{+\infty}E\,dr=\frac{Q}{4\pi\varepsilon r}$$

(Eq 19.2.2)

Hence the relation is:

$$E(r)=\frac{V(r)}{r}$$

(Eq 19.2.3)

19.3 Electrodes - Current density

Because the electrolyte has a certain conductivity g there is a current flow from the electrode to the (conductive) walls of the container. The current density can be calculate by means of the local Ohm's law:

$$j = gE$$

(Eq 19.3.1)

Hence the current I is obtained by integration over a spherical surface surrounding the electrode:

$$I = \int_S j\,dS = \int_S gE\,dS$$

(Eq 19.3.2)

Since this is valid for all concentric spheres centred in the centre of the electrode, it is possible to choose the one having radius R where we know the potential V(r) to be equal by the voltage U set by the battery:

$$I = 4\pi RgU$$

(Eq 19.3.3)

19.4 Electrodes - Impedance of an electrode

Finally the impedance can be calculated as the ratio between the voltage and the current:

$$Z_{electrode} = \frac{U}{I} = \frac{1}{4\pi R g}$$

(Eq 19.4.1)

Notice the so calculated impedance is a real number, therefore it is a pure resistance:

$$R_{electrode} = \frac{\rho}{4\pi R}$$

(Eq 19.4.2)

where ϱ = 1/g is the "resistivity" ***of the fluid***.

19.5 Electrodes - Comments

There are some interesting things to notice:

1. The impedance of an electrode is purely resistive, so let us call it resistance

2. The resistance of an electrode is inversely proportional to the curvature of its tip

3. ***The resistance of an electrode is not the resistance of the wire it is made of***

4. The resistance of the electrode depends on the resistivity of the fluid in which it is immersed, but not on its dielectric constant

5. The resistance of the electrode is, in fact, the equivalent resistance of a system made of an electrode, a fluid and a conductive container which is supposed to be very large and very far from the electrode itself. This latter assumption would correspond to the fact the container would act as an ideal ground.

6. Considering the latter note, also notice the resistance of an electrode does not explicitly depend on the geometry of the ground electrode: ***if the ground electrode is made of a perfect, large and far piece of metal then the value of the equivalent resistance of the electrode does not depend on the geometry and the location of the ground electrode.***

20.1 Neural recordings - Sensitivity of an electrode

Considering the brownian motion of the constitutive particles (ions and electrons) of electrical resistor, there is always random white-noise gaussian-distributed current in it due to its temperature which leads to the following average voltage drop measurable across any open resistor:

$$V_{JN}(T) = \sqrt{4k_B TR\Delta f}$$

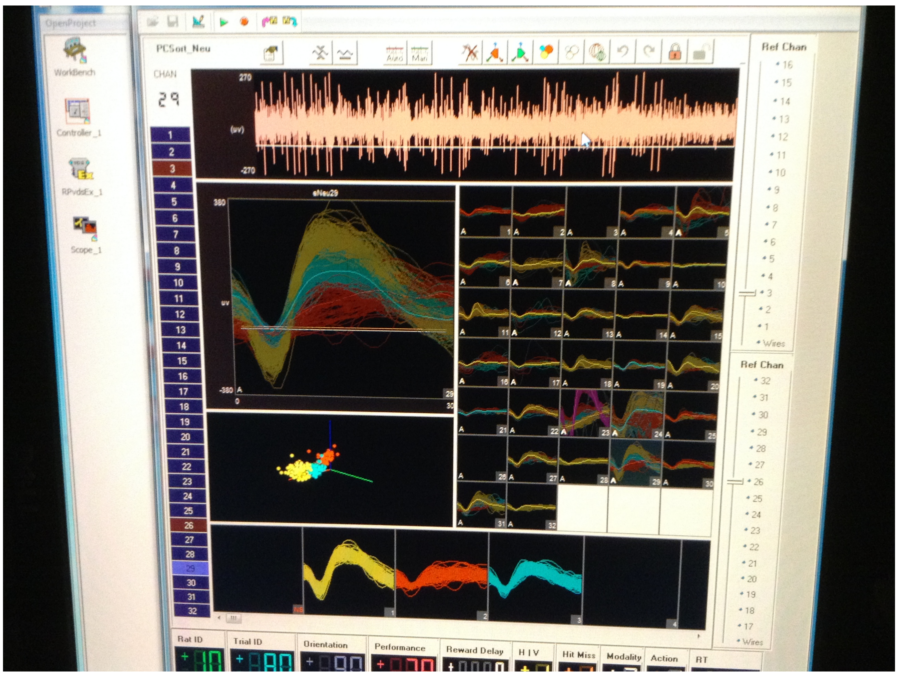
(Eq 20.1.1)

which is called the "Johnson-Nyquist" noise, where $k_B = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann's constant, T is the absolute temperature (in Kelvin), R is the resistance and Δf is the considered bandwidth of the white noise. Typical conditions for a neural signal electrode are: a resistance of 1Mohm, room temperature T = 300K Δf = 20kHz, Under those conditions the thermal noise is:

$$V_{JN} = \sqrt{4 \cdot 1.38 \cdot 10^{-23} \cdot 300 \cdot 10^{6} \cdot 2 \cdot 10^{4}} \simeq 18\mu V$$

(Eq 20.1.2)

It means whatever signal lower than that value remains buried inside that noise for that electrode. This is very useful in electrophysiology, because this is what makes neural signal isolation possible: far and weak neurons will not be captured as a signal, near and strong neurons will pass that threshold and will become visible, as shown in the following picture:



(Fig 20.1.3: neural spikes)

Notice here the polarity of the action potentials are reversed: it is because this is the case of an extracellular recording, so the electrode is outside the cell membrane.

20.2 Neural recordings - Reference and ground electrodes

The purpose of the reference electrode is to make the noise cancellation (see before).
It works on two assumptions:
A) the reference is close enough to the signal electrode, in order to pick-up the same noise.
B) the reference "listens" to all neurons in the area of interest, collecting the local field potential

--> hence its impedance must be ***lower*** than the signal electrode, otherwise it would not be possible for it to capture many signals (even from relatively far and weak neurons) converging to it.

Using a high impedance electrode as reference will not do the job: it would just record the signal of another neuron, and then the differential amplifier will just measure the difference between two (or a few) neurons, not the difference between the neuron of interest and its local field potential pedestal. The same apply for the ground electrode.

Possible sources of artefacts in neural recordings are:

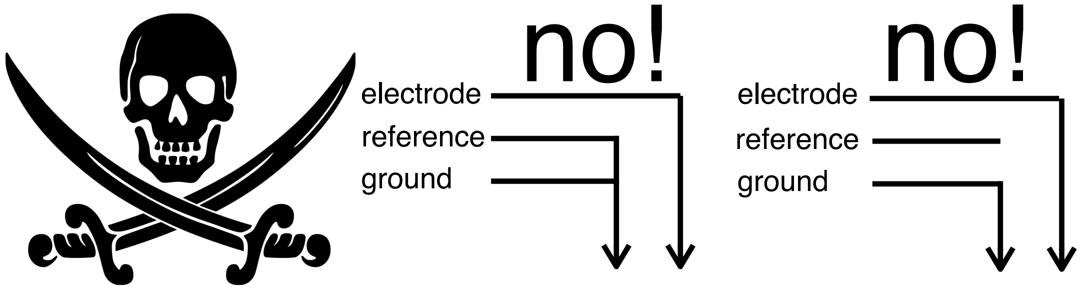
1. wrong impedance of the signal electrode
2. wrong impedance of the reference electrode
3. wrong impedance of the ground electrode
4. wrong location of the signal electrode
5. wrong location of the reference electrode
6. wrong ground of the ground electrode

Short circuiting the reference with the ground electrode or, even worse, removing the reference electrode leaving its corresponding input floating are common misconceptions.

It is therefore very important to:

1. think about the correct impedance and location of all the electrodes
2. make use of a good ground
3. always properly connect and use the reference electrode

there is really nothing else to know in order to have good neural recordings: all the rest are just pirate workarounds.



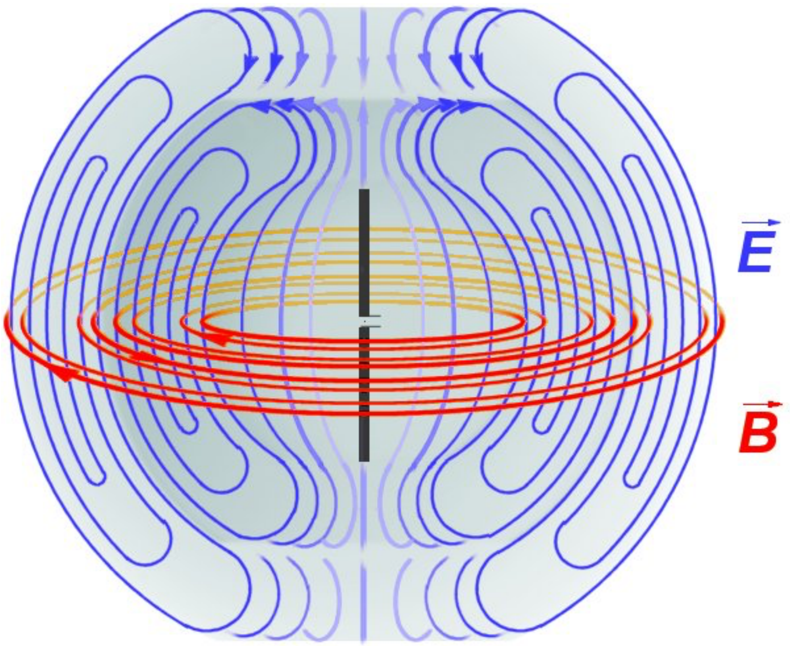
(Fig. 20.2.1: common pirate workarounds misconceptions in electrophysiology)

20.3 Neural recordings - Electromagnetic radiation

Electrical signals are not restricted to propagate into electric cables or electronic components: they can also propagate through many different media, including insulator materials, air and vacuum.

They do this without net transportation of charge, but modulating the electric and the magnetic fields which permeate the surrounding space. The transportation of information, which is related to a transportation of energy, corresponds to the phenomena of "electro-magnetic radiation".

In the figure below, a "radiating dipole" is shown:



Maxwell's equations:

$$\begin{cases} \nabla \cdot E = \dfrac{\rho}{\varepsilon_0} \\ \nabla \cdot B = 0 \\ \nabla \times E = -\dfrac{\partial B}{\partial t} \\ \nabla \times B = \mu_0 \left(J + \varepsilon_0 \dfrac{\partial E}{\partial t} \right) \end{cases}$$

(Fig 20.3.1: a radiating dipole) (Eq 20.3.1)

When the polarity of the dipole is changed over time by means of an oscillator, the electric field lines around it propagate apart from it in a kind of expanding "bubbles" detaching from the dipole and travelling at the speed of light through the surrounding media. This is energy, in terms of photons of a certain wavelength, which the radiator looses and which has to be constantly provided by the power supply powering up the oscillator.

The core mathematics and physics of this phenomena is described by the "Maxwell's equations".

20.4 Neural recordings - Electromagnetic current induction

The electro-magnetic radiation of an oscillating dipole can make another distant dipole oscillating, due to the electric and magnetic forces associated to the field: this is the basic concept of "radio transmission". Besides the radiated energy, there is always a part of energy which does not go away from the transmitter but stay close to it. The wave length which is associated to the frequency of oscillation sets the boundary for the different arrangements of the energy:

$$\lambda = \frac{1}{n} \cdot \frac{c}{f}$$

(Eq 20.4.1)

where c = 299792485 m/s is the speed of light and "n" is the "index of refraction" of the media. For the vacuum n = 1, for air n = 1.000293. For other materials it can be bigger. As a first approximation, objects staying inside a range equal to λ are in the so called "near field" or "induction field" region, while the ones staying outside that range are in the so called "far field" or "radiation" region. Considering a closed circular ring made of an electrical conductor in presence of an electro-magnetic field and having radius "R", cross section "A" and conductivity "g" it is possible to show there is an electric current "induced" at distance on it:

$$\left\{ \begin{array}{l} \int_s \nabla \times E \, dS = \oint_\gamma E \cdot d\gamma = 2\pi RE \\ \nabla \times E = -\frac{\partial B}{\partial t} \end{array} \right\} \Leftrightarrow 2\pi RE = \int_S -\frac{\partial B}{\partial t} dS = -\pi R^2 \frac{\partial B}{\partial t}$$

(Eq 20.4.2)

Hence, recalling the definition of current and the local Ohm's law:

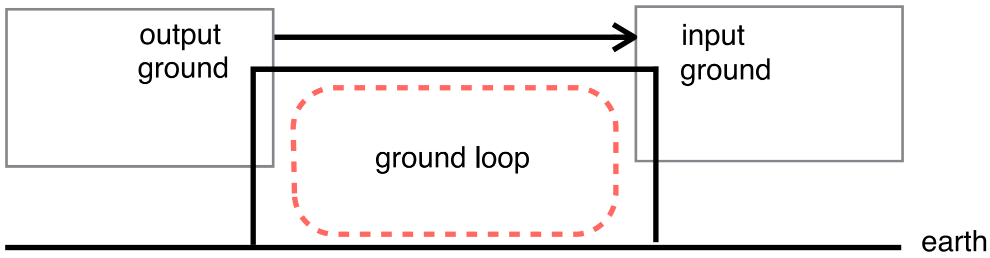
$$\left\{ \begin{array}{l} j = gE \\ I = Aj \\ 2\pi RE = -\pi R^2 \frac{\partial B}{\partial t} \end{array} \right\} \Leftrightarrow I(t) = -\frac{1}{2} AgR \frac{\partial B}{\partial t}$$

(Eq 20.4.3)

Which means a time variable magnetic field, and its associated electric field, "induces" a time variable current I in a closed loop circuit coupled to them.
Notice the bigger the radius, the bigger the induced current. Also notice the faster the variation of B, the bigger the induced current.

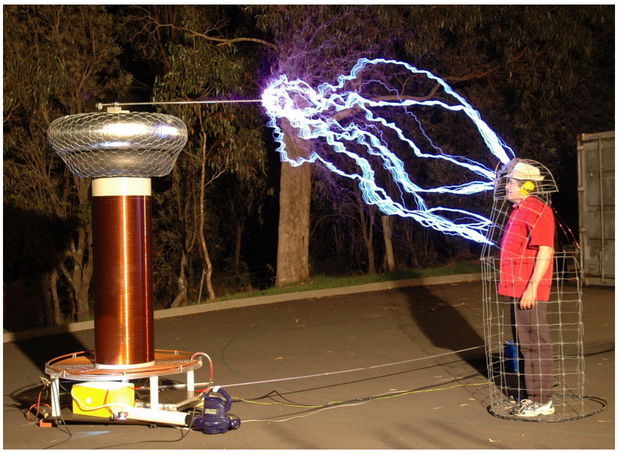
20.5 Neural recordings - Electromagnetic shielding

The result of the computation of the electric current induced in a loop of wire by a time-varying electro-magnetic field shows that the induction effect is proportional to the radius of the loop, hence by shrinking all possible loops the effects of the induction are minimised. This is particularly important in the case of "ground loops" which may occur between two distant pieces of equipment connected together:



(Fig 20.5.1: ground loop)

Another possibility, is to put whatever device is sensitive to inductions inside a "Faraday cage". Faraday cages are basically 3-dimensional topologically closed empty balls made of a good electrical conductor. Since they are 3-dimensional closed manifolds, they do not exhibit any open surface through which a net flux of either electric or magnetic external field can induce currents or polarise charges. They work fine in the electrostatic case or in the radiative case when in the far field region. In the near field region the magnetic field scan be strong in a way that the electric conductivity of the metal of a Faraday cage might be not enough high in order to reflect the incoming electromagnetic waves. In this case it can be useful to surround the Faraday cage with another one made with a good ferromagnetic material (e.g. iron). The latter cage will work against the strong magnetic field by trapping it and making circulating it inside the thickness of its ferromagnetic walls, due to the high magnetic permeability of the material. An "electric" Faraday cage is usually enough, but the "magnetic" one might become important too in case of strong low frequency noise, for which the corresponding wavelength could be so long that everything is within the near field radius. For the same reason, sometimes electric Faraday cages are made in metal mesh and not full solid metal: if the pitch of the mesh is much finer than the wavelength of the radiated noise signal then it will not pass through it: it basically means that in those circumstances the average conductivity of the cage, even considering the holes, is enough.



(Fig 20.5.2: Faraday cage shielding an engineer in front of his Teslacoil. Hint: do not try it at home)