COURSE NOTES BY MATTEO TOIA

Origin of bio signals

Most processes in the body are regulated by electrical signals sent through the nervous system from the brain

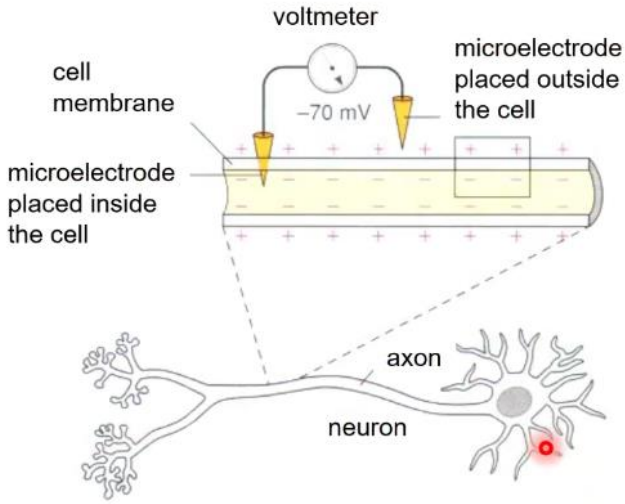
Neurons

Biosignals originate and are transported by neurons.
Neurons can be divided into 2 parts

- A peripherical connection which connects to the other neurons with **synapses**
- A ling segment called **axon**

The mechanism by which the signal travels in the neuron has 3 main components

- Sodium-potassium pump
- Diffusion of ions
- Consequent development of electrical gradient



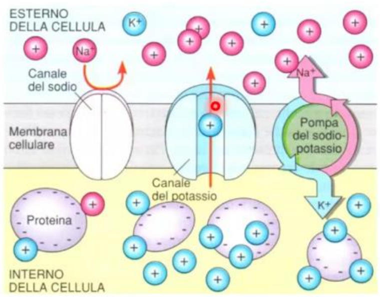
Resting potential

We can measure a voltage drop between the external and the internal part of the neuron, the potential inside of the cell will be $70mV$ lower than the outside, so we have $-70mV$ from outside to inside.

This static voltage is called **resting potential** and it is sustained by the neuron utilizing ion's pumps to maintain a concentration gradient of different ions between inside and outside the neuron.

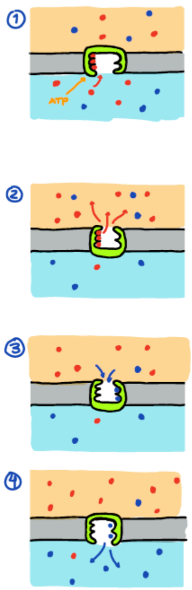
The main ions that influence this process are sodium and potassium ions (calcium plays an important role in the heart tissues).

Sodium potassium pump



Channels and permeability

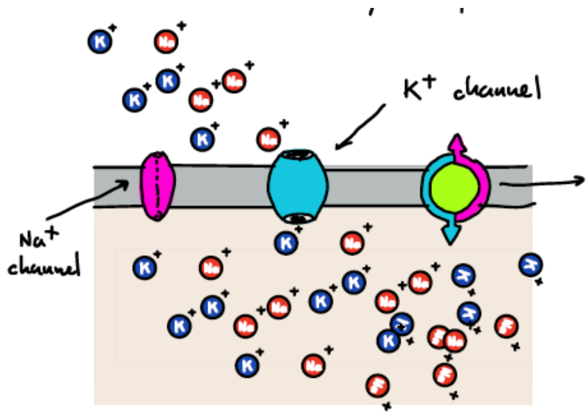
In the cellular membrane there are several channels specific for a type of ions and that can either be open or closed. We call **permeability** the **capacity of the membrane to allow passage to a specific ion**, **this property depends on how many channels are open.**



Pumps

Ions pumps actively pump in or out a type of ions.
We are interested in **sodium potassium pumps** which are pumps that when active consume ATP to pump potassium ions K^+ into the cell and sodium ions Na^+ out.
The operation is the following

1) The pump takes 3 Na^+ ions and sends them to the outside of the cell consuming ATP.
2) As the Na^+ ions are released 2 K^+ ions are attached to the pump and sent inside of the cell



This process is then repeated, we end up accumulating more positive charge outside the cell

Resting condition

In a resting condition we have that

- Low permeability of the membrane to Na^+
- High permeability of the membrane to K^+

This means that while we end up having

- A low diffusion of Na^+ ions from outside to inside
- A high diffusion of K^+ ions from inside to the outside

When we consider the potassium K^+, we have that the anionic component (negative) remains trapped inside the cell as it is not able to diffuse through the K^+ channel, this means that eventually a potential is born from the accumulation of negative charge inside the cell, this potential eventually becomes high enough to match the diffusion so we reach an electrochemical equilibrium

Total contribution of all the ions

$$E_m = \frac{RT}{F} \ln \frac{P_k[K_o^+] + P_{na}[Na_o^+] + P_{cl}[Cl_i^-]}{P_k[K_i^+] + P_{na}[Na_i^+] + P_{cl}[Cl_o^-]}$$

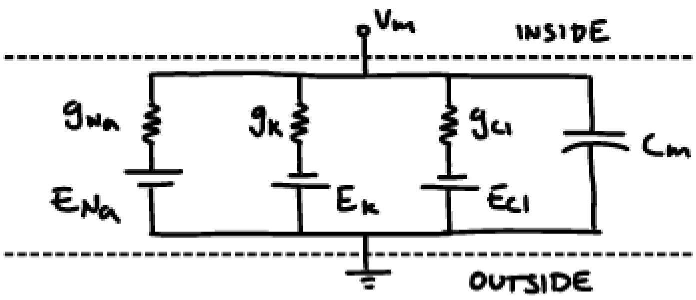
equation of Goldman-Hodgkin-Katz (GHK)

P_X: membrane permeability to ion X
(es. $P_{na+} = 2\times10^{-8}$ cm/s, $P_{K+} = 2\times10^{-6}$ cm/s)

Ion	Extracellular concentration (mM)	Intracellular concentration (mM)	Equilibrium potential (mV)
Na+	145	12	67
K+	4	155	–99
Cl-	120	4	–92

We can use the equation above to determine the contribution to the resting potential of each kind of ion.

Electrical model of the cell membrane

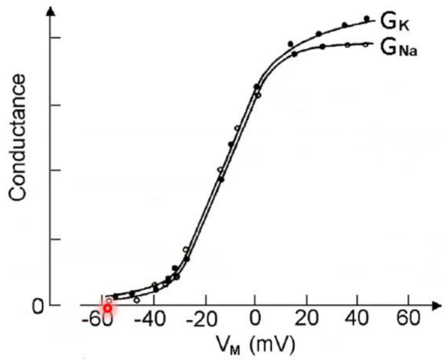


We indicate the interplay between the membrane permeability and the rest potential.

The capacitance C_m is charged at to the value of the rest potential at any given time, any change in g of the ion concentration will change the value of V_m

Action potential

Let's now consider a negative pulse applied to the exterior of the neuron so that the potential between interior and exterior increases becoming positive.

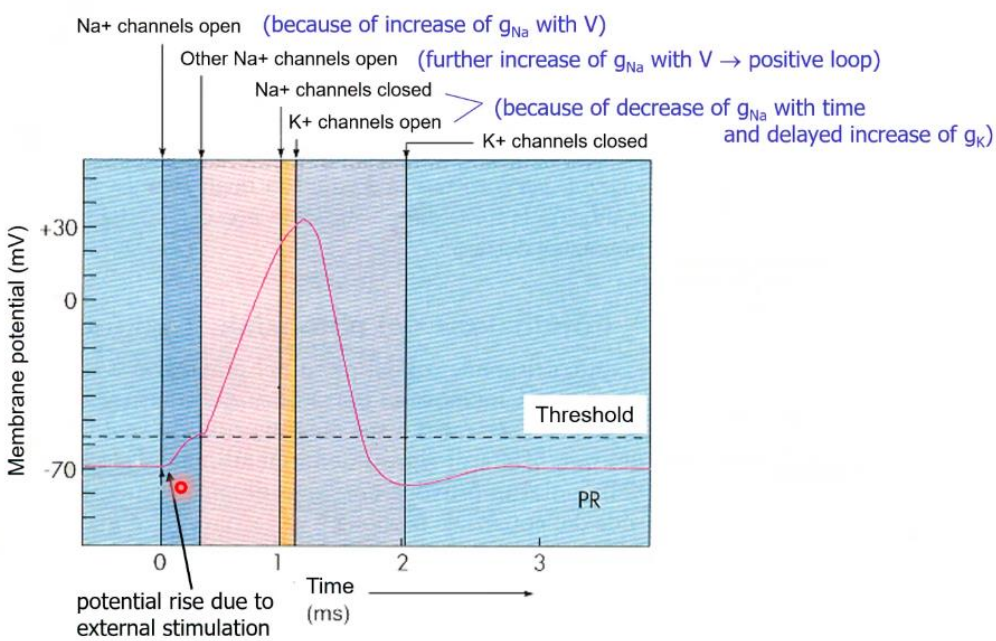


The behavior of the neuron is determined by behavior of the cell membrane conductance for potassium and sodium.

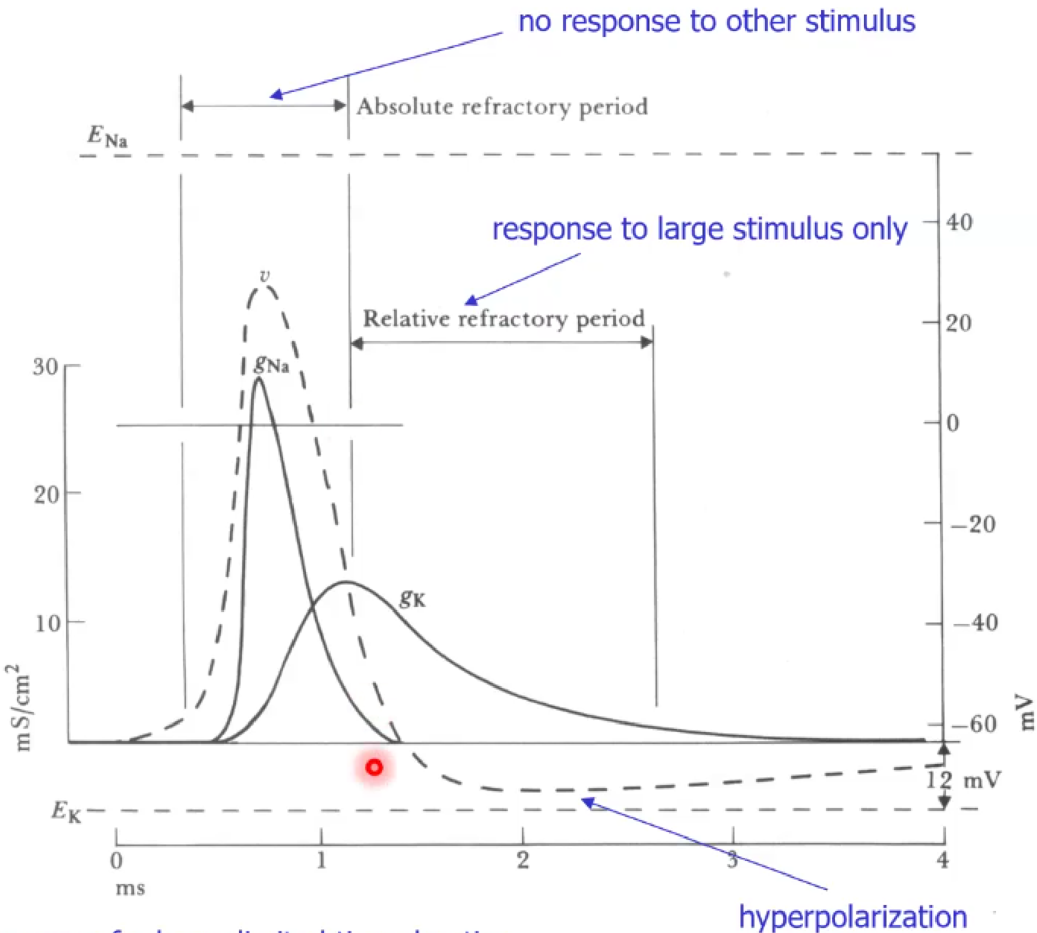
As we can see as the membrane potential increases both the conductance also increase with the one of sodium starting first while the one of potassium follows later.

Operation

- As the potential increases first the conductance for sodium ions increases.
    - As the sodium channels open the ions are pushed inside the cell by the electric potential and the concentration gradient.
- Since the ions are positive the potential inside the neuron increases creating positive feedback for the conductance that increases again allowing more sodium ions to pass.
- Eventually the N_a^+ sodium charge that can flow inside the membrane is finished and instead the potassium ions can flow outside counteracting the increase in potential which shortly after starts to decrease.
- We first reach a smaller voltage than the resting potential at which point the K^+ channels close and then we can go back to the resting condition.



Below we can see the conductance of the membrane during the different moments of the stimulation.

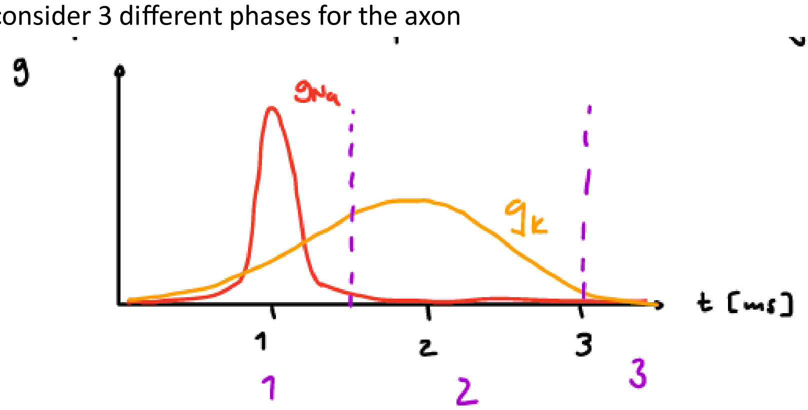


note: the increase of g has a limited time duration

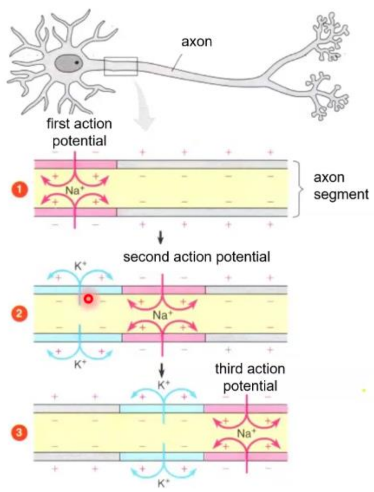
Potential action

After the signal is transmitter we need to consider 3 different phases for the axon

Absolute refractory period



During this interval the membrane is unable to respond to another stimulus because the Na concentration g_{Na} is not in the rest value.



Relative refractory period

When g_{Na} returns to its balance value the membrane is again able to receive only a large stimulus as g_k is still not fully recovered

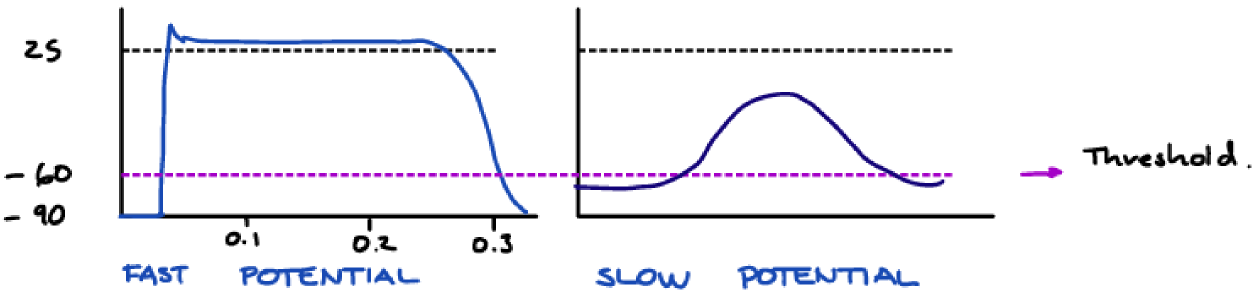
Third action potential

During this third period we have the relaxation of the membrane and the restoration of the rest potential

The presence of these refractory periods prevent the signal to travel backwards in the neuron

Fast Action potential

There is no single action potential for example cardiac fibers will be moved by fast action potential while others may be moved by slow action potential.



The different shape between the 2 graphs is cause by the effect of additional ions like Ca^{++} which intervene when the sodium ions are over.

Synapsis

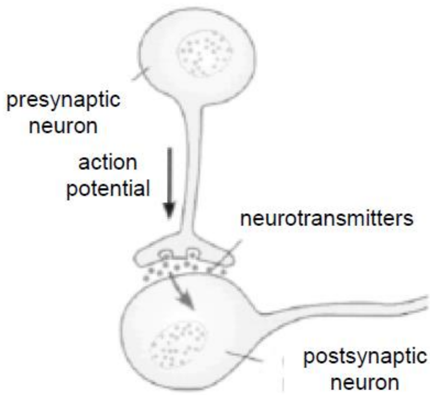
Electrical synapsis

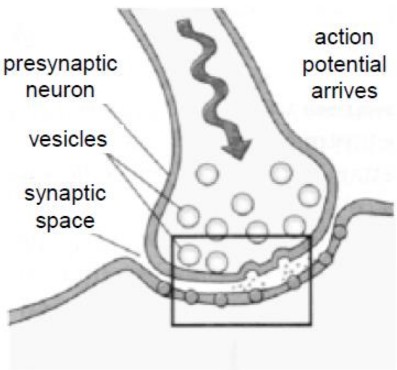
When neurons are very close $d < 3nm$ the action potential can be transmitted between the neurons by means of the Na^+ ions.

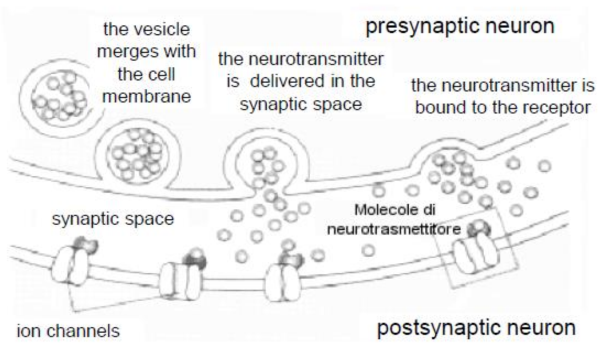
Chemical synapsis

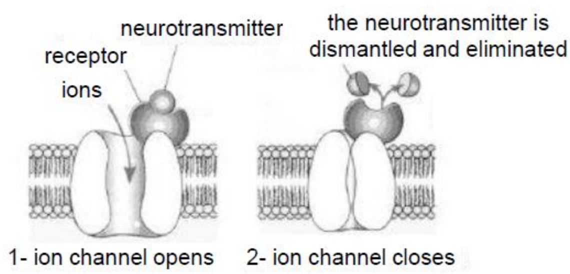
When the distance is greater the signal is converted into a chemical signal though the use of neurotransmitters.

Inside the synapses there are neurotransmitters contained in vesicles which are released and the break inside the other neuron.









Speed of signal transfer in the neurons

All signal will travel at the same speed so to measure the transfer speed we apply 2 pulses at 2 different points of the neuron placed at a distance D from each other and the we measure the arrival time at the same point.
The times are L_1 and L_2 if we use their difference to divide the distance D we obtain the speed of the signal

$$u = \frac{D}{L_1 - L_2}$$

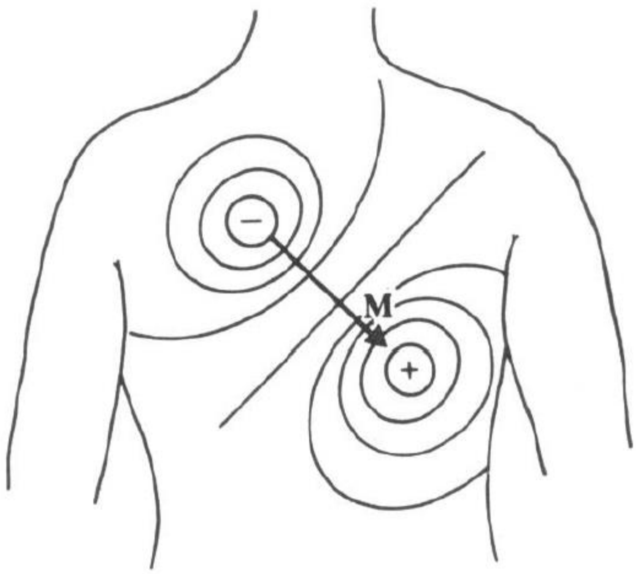
Conduction of signals in the volume

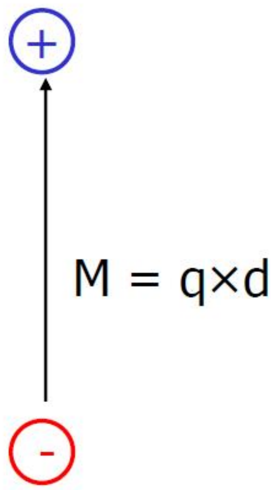
To truly understand how signal travel in the body is important to know how they are propagated in the body.
So let’s consider the heart as an example

Electrical model of the heart

We want to define an electrical model for the heart, to do so we need to consider that the heart electrical activity is based on a strong flow of calcium ions Ca^{++}.

The heart can be considered an electric dipole since only half of it will contract at a time, the intensity and the direction of the dipole will change with time.

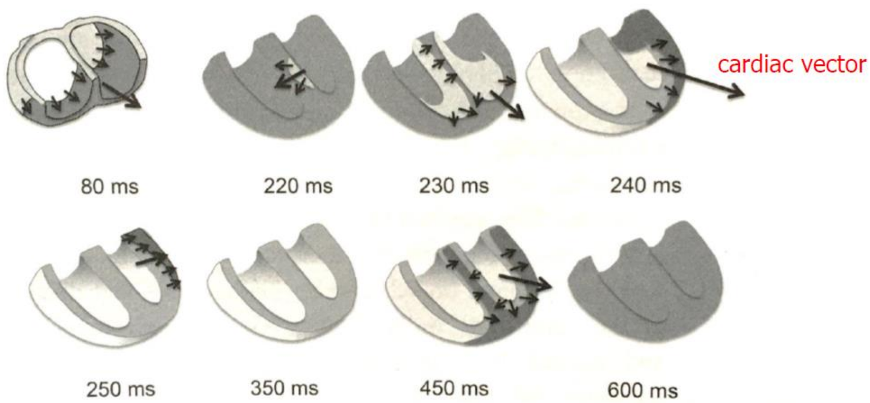




Steps in depolarization

We have that the light areas are depolarized while the dark areas are polarized.

We can see that the polarization follows vectors along the walls of the heart, the total cardiac vector is the sum of each individual component.



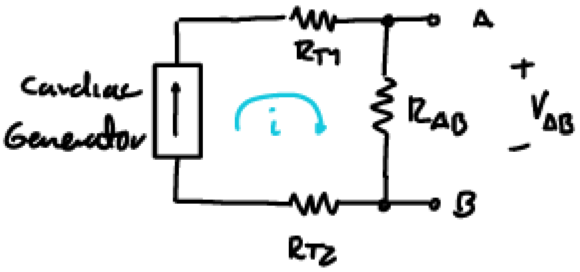
And it can be determined by measuring the voltage potentials at different points of the body.
The total cardiac vector is given by

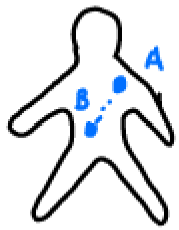
$$\vec{M} = q \cdot \vec{d}$$

Oriented from – to +-

Measurement of the cardiac vector

To measure the cardiac vector we measure the surface potentials of the body. This is challenging because the potentials are highly attenuated by the resistance between generation point and sourface of the body as well as by the resistance between measurement points.





The equivalent circuit is the following

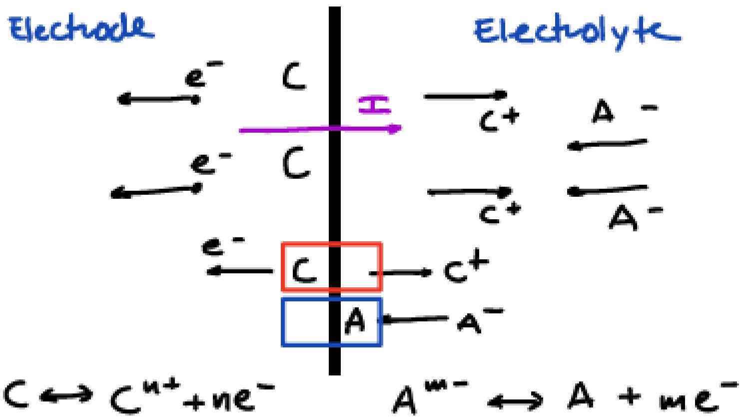
Electrodes for biopotentials

The electrodes act as the interface between the tissue and the instrument used.
During measurement current flows from the tissue to the instrument, it is important to note that

- In the tissue the current is carried by ions
- I the instrument the current is carried by electrons

The electrode needs to be able to convert the ion current into an electron current

Electrode/electrolyte interface



The electrolyte contains ions (C^+) of the metal forming the electrode with a matching number of anions A^- since we need to have charge neutrality.

We have that at equilibrium the rate of oxidation is equal to the rate of reduction.

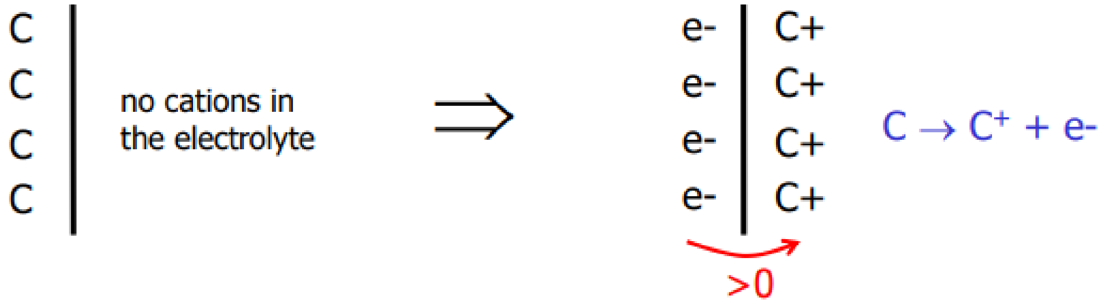
$$C \Leftrightarrow C^{n+} + ne^-$$

The current unbalances the chemical reaction to one side or the other according to the direction, the rearrangement of the ions towards equilibrium creates a fixed charge at the interface which will in turn produce a potential difference called **half cell potential**.
The value of this potential depends on may factors like

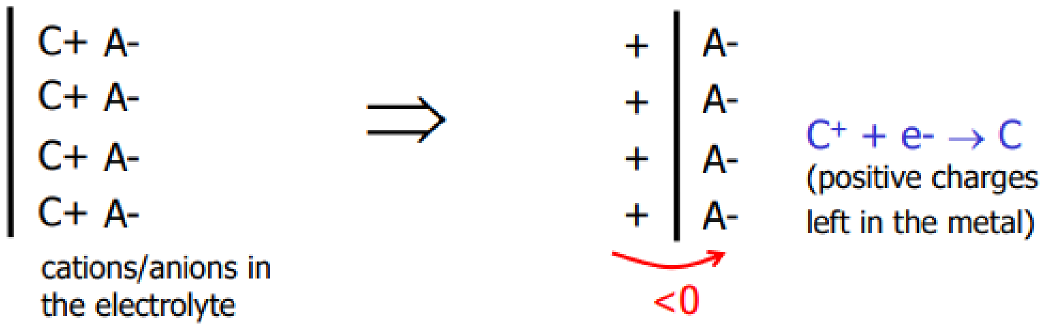
- Type of metal
- Concentration of ions in the electrolyte
- Temperature

Example

Example 1



Example 2



Electrical modelling of an electrode

The potential observed at the electrode contact can be determined by the **Nernst equation**

$$E = E_0 + \frac{RT}{nF} ln\left(\frac{a^{C_n^+}}{a^C}\right)$$

Where

- a stands for the activity of the element which in some cases can be approximated with its concentration
- E_0 is the standard potential

Semi element potential

In order to measure the potential of an electrode we need another electrode so it is impossible to measure in principle, to get around this limitation we **define potentials relative to a standard electrode** so that we can write

$$V_{pot} = V_{elec} - V_{std}$$

In electrochemistry the standard is **gaseous H_2 over a platinum electrode**.

Polarization of electrodes

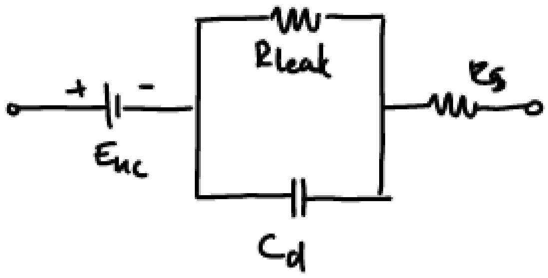
The flow of current can modify the potential at the interface of an electrode, we divide electrodes in 2 categories

1) **Perfectly polarizable**: no charge flows through the junction when current flow is present (behavior of a capacitor)
2) **Perfectly non polarizable**: the current flows freely at the interface without modifying the potential difference

Equivalent circuit model of an electrode

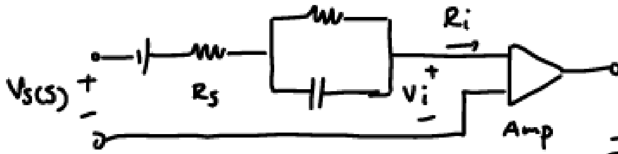
In this model

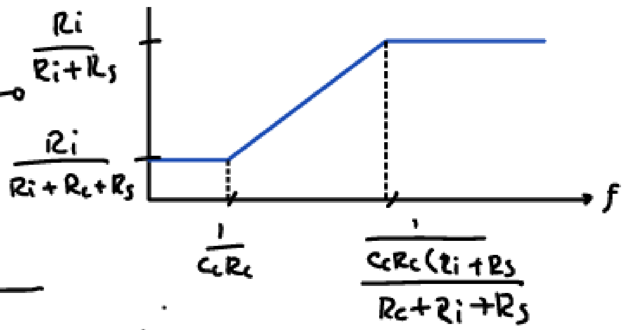
- C_d: surface change capacitance
- R_s: resistance seen by the electrolyte
- E_{hc}: half cell potential



Frequency response

The model we introduced exhibits a pole zero transfer function and thus is frequency response will be the following





$$\frac{V_i}{V_s} = \frac{R_i}{R_i+R_s} \cdot \frac{s + \frac{1}{R_c C_c}}{s + \frac{1}{C_c\left(R_c \parallel (R_i+R_s)\right)}}$$

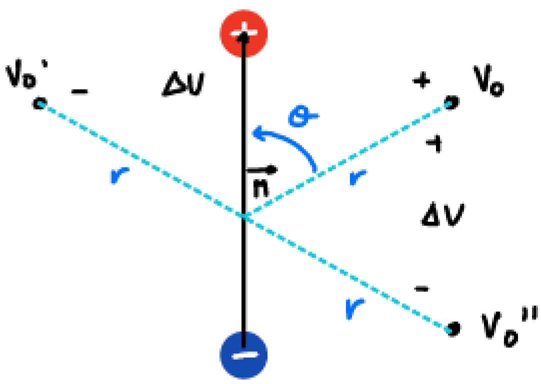
We are going to have a zero at low frequency and a pole at high frequency

Electrocardiograph fundamentals

As we have said the heart can be modeled as an electric dipole which changes direction and orientation over time

$$\vec{M} = q \cdot \vec{d}$$

Components of the cardiac vector



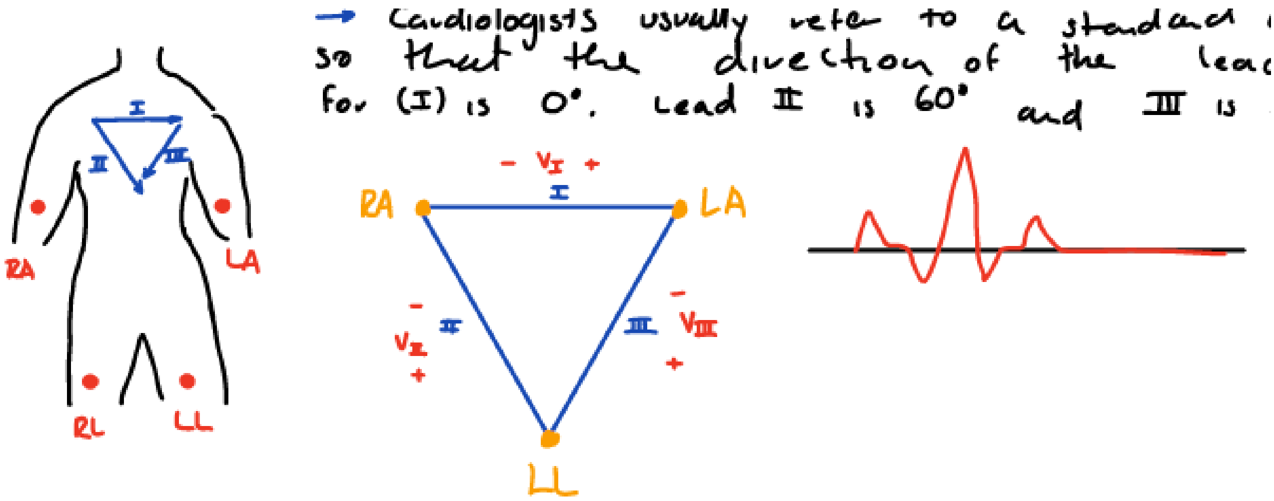
The components of the cardiac vector can be determined by measuring the voltage difference between different points

$$V_O = \frac{M \cdot r}{4\pi\varepsilon r^3} = \frac{H\cos(\vartheta)}{4\pi r^2} \quad for\ large\ r$$

The Eindhoven triangle

Cardiologist usually refer to a standard direction so that the direction of the lead vector is

- For lead I is 0°
- For lead II is 60°
- For lead III is 120°



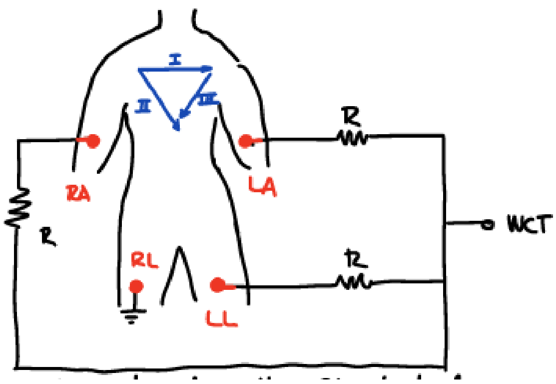
These measurements are differential potentials between 2 points, after taking these we extract the unipolar potentials

Unipolar potentials

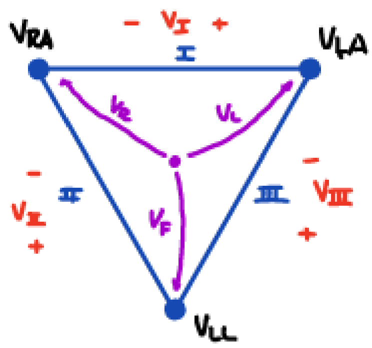
Unipolar potentials measurements are taken with respect to the **Wilson's central terminal WCT**. Whose potential is obtained as

$$V_{WCT} = \frac{1}{3}(V_{LA} + V_{RA} + V_{RL})$$

Typically this measurement is done connecting the RL (right leg) electrode to ground, as shown in the picture below



The resulting vectors are rotated by 30 degrees with respect to the original ones



Example

Let's consider the unipolar potential for the left arm

$$V_L = V_{LA} - WCT$$

This is equal to

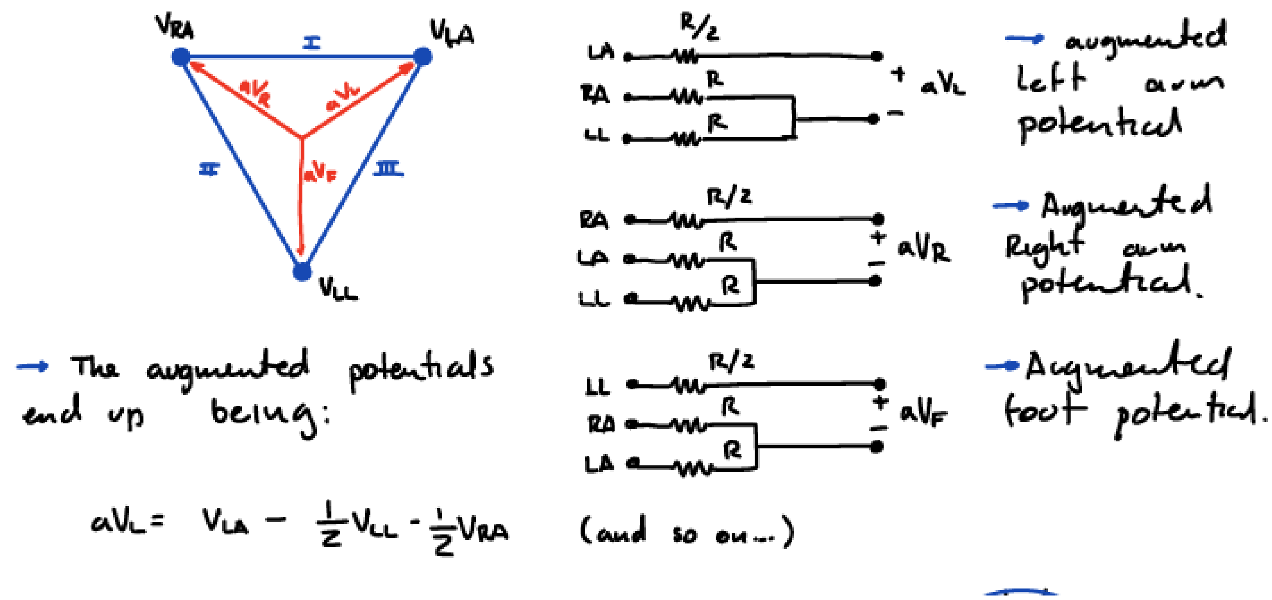
$$V_L = V_{LA} - \frac{1}{3}(V_{LA} + V_{RA} + V_{RL})$$

So we get

$$V_L = \frac{2}{3}V_{LA} - \frac{1}{3}(V_{Ra} - V_{LL})$$

Augmented potentials

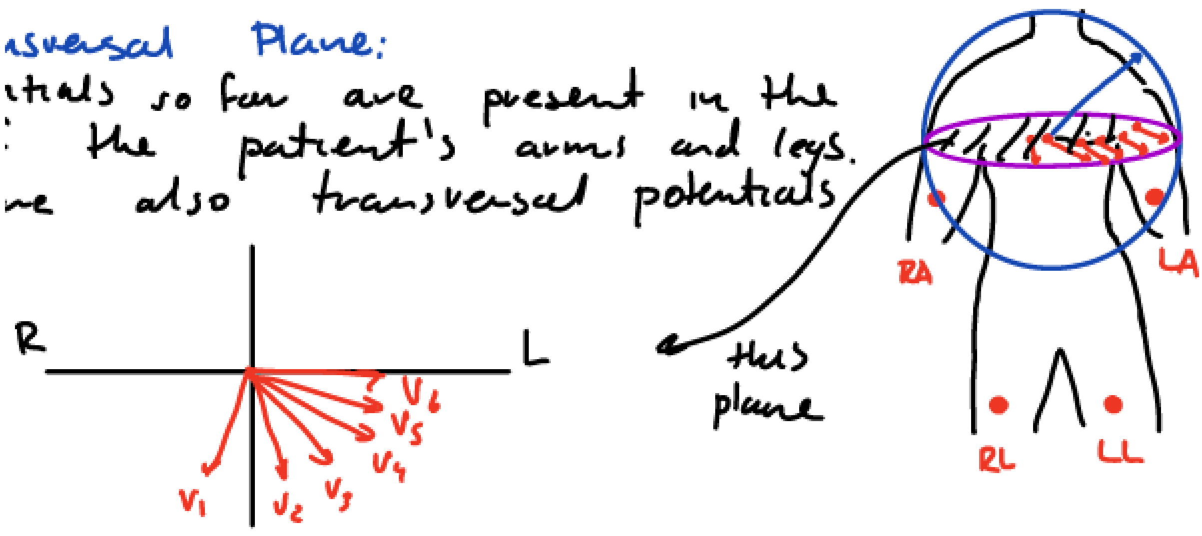
The augmented potentials utilize the differential measurements to greatly improve the signal and are defined as follows



We have that rather then measuring from the Wilson's node we measure from one point to the average of the other 2, this allows us to obtain a voltage which is increased by 50% with respect to the normal measurement as now the electrode on the body from which we are measuring sees the high impedance of the voltage tester rather than the resistance R connecting it to the Wilson node.

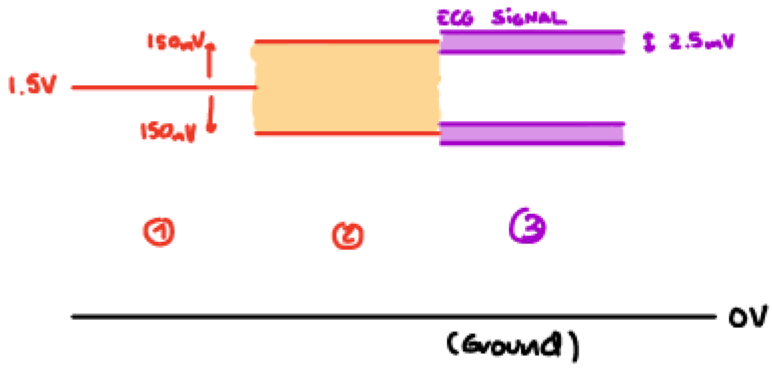
The transversal plane

The potentials measured so fare are present in the plane of the patient arms and legs but we also need to consider transversal potentials.



Electrocardiograph signal characteristics

We have discussed the placement of the electrodes now let's focus on the signal we are going to measure and its composition.



Common mode input $(50/60Hz\ 1,5V)$ (1)

The common mode signal arises mostly due to the capacitive coupling of the patient and the power lines

Electrode offset (DC signal $\pm 300mV$) (2)

The electrodes are polarized by the half cell potential_{l difference (contact between electrode and tissue). Electrodes also become partially polarized by the bias current of the amplifier.

ECG signal $(0,05 \ldots 150\text{Hz}, 2,5\text{mV})$ (3)

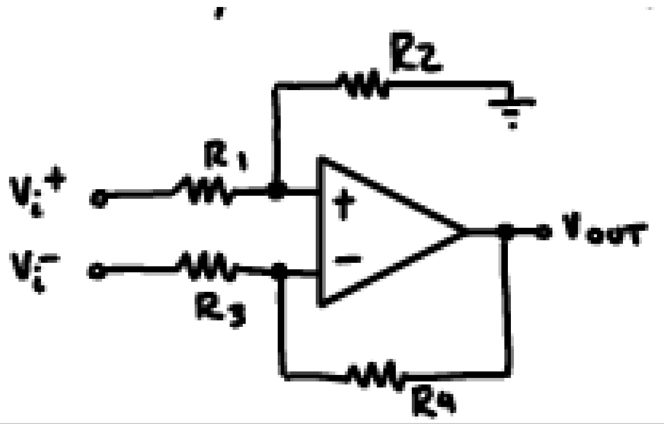
The signal from the cardiac vector we are interested in measuring. The value of this voltage is quite small when compared to the other contributions, this means that we are going to need a very precise filter to avoid losing it.

INA

The instrumentation amplifier or Ina is the best solution for the ECG as it presents an high input impedance and CMRR.

To highlight the importance of using an INA let's first start by utilizing a normal OPAMP in differential configuration

Example with OPAMP



Let's consider the 2 inputs independently and then superimpose the outputs

V^+

Here we have a non-inverting configuration and a voltage divider between R_2 and R_1 so the total transfer function will be

$$V_{out} = V_i^+ \cdot \frac{R_2}{R_1 + R_2} \cdot \left(1 + \frac{R_4}{R_3}\right)$$

V^-

In this case we have a simple inverting configuration so we get

$$V_{out} = V_I^- \left(-\frac{R_4}{R_3}\right)$$

We can just sum the 2 results and obtain the output

$$V_{Out} = V_i^+ \cdot \frac{R_2}{R_1 + R_2} \cdot \left(1 + \frac{R_4}{R_3}\right) - V_I^- \frac{R_4}{R_3}$$

And if we set $\frac{R_4}{R_3} = \frac{R_2}{R_1}$ we simplify and obtain

$$V_{Out} = (V_i^+ - V_I^-)\frac{R_4}{R_3}$$

CMRR

In theory if all the component were ideal the CMRR would be infinite, in reality this is not the case and as we will show even a small mismatch between the real value and the nominal values will cause an unacceptable common mode gain.

Let's set $V_i^+ = V_i^-$ and consider $R_1 = R_3 = R_4 = R$ while $R_2 = 0{,}999R$ so a mismatch of 0,1%.

If we repeat the calculations for the gain using the new resistance value we obtain

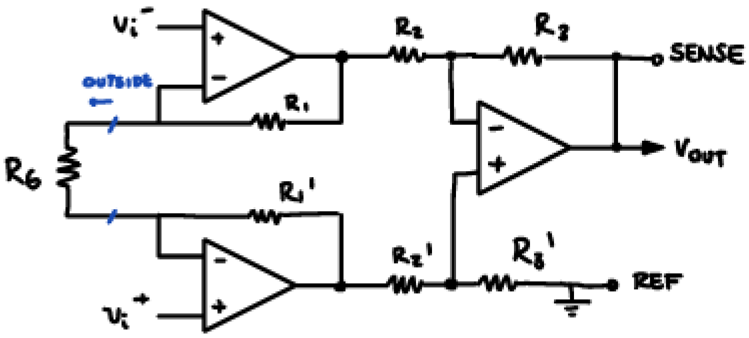
$$\frac{V_{CM}}{V_i} = \frac{0{,}999R}{1{,}999R}\left(\frac{2R}{R}\right) - \left(\frac{R}{R}\right) = 0{,}0005$$

In this condition the CMRR is

$$CMRR = \frac{1}{0{,}0005} = 66dB$$

This result is not good enough for our application

Structure of the INA



The instrumentation amplifier is composed by 3 OPAMPs divided into 2 stages, the 2 input OPAMP increase the CMRR of the final circuit while the third provides the gain.

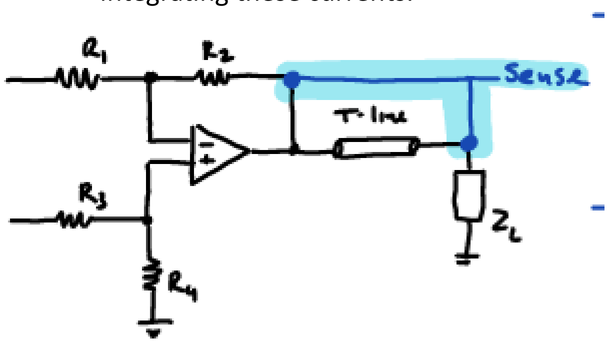
If we utilize BJT to implement the input amplifier we obtain a very large input impedance, this however will cause an input bias current, so we are going to need a path to ground to avoid integrating these currents.

Sense electrode

We can see from the graph that the INA presents 3 outputs, a reference pin a sense pin and an output pin.

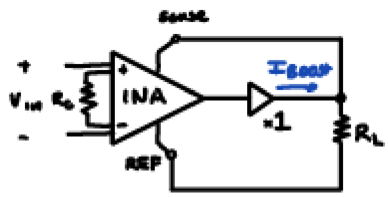
The sense pin is very important as it allows us to render negligible all non idealities and parasitism that we may encounter between the output of the INA and the LOAD.

By connecting the sense directly at the load we extend the loop to include the parasitism whose effects will now be attenuated by the negative feedback.



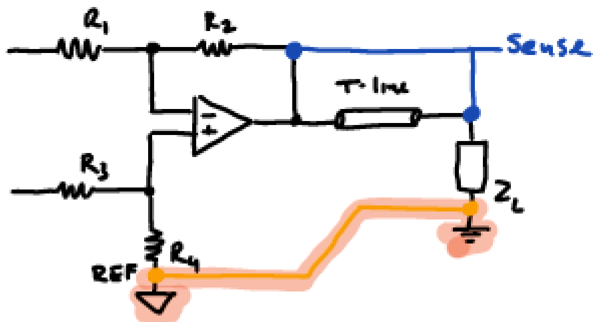
Current booster

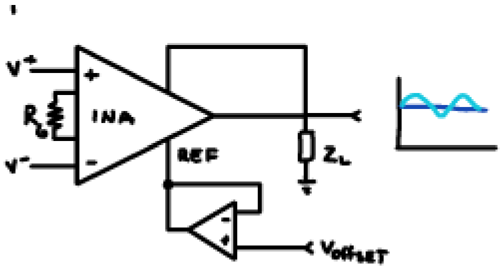
The extended loop allows us also to include a current booster at the output, we have that now we can include a buffer at the output in case we need to provide more current than the INA is capable of and since the booster will also be included in the loop the circuit will automatically regulate itself so that the output remains the same.



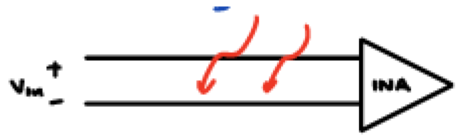
The reference node

The circuits on some applications may have a separate ground in the input and output nodes, in order to keep the same ground we connect the reference node to the ground of the load.





Additionally the reference can also be used to provide an offset to the output.

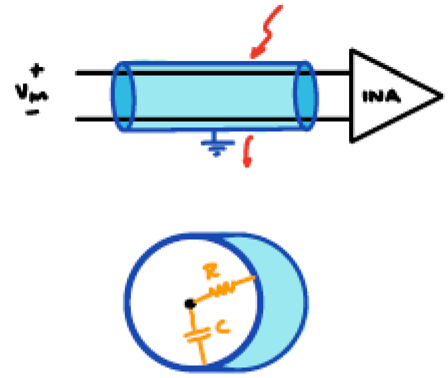


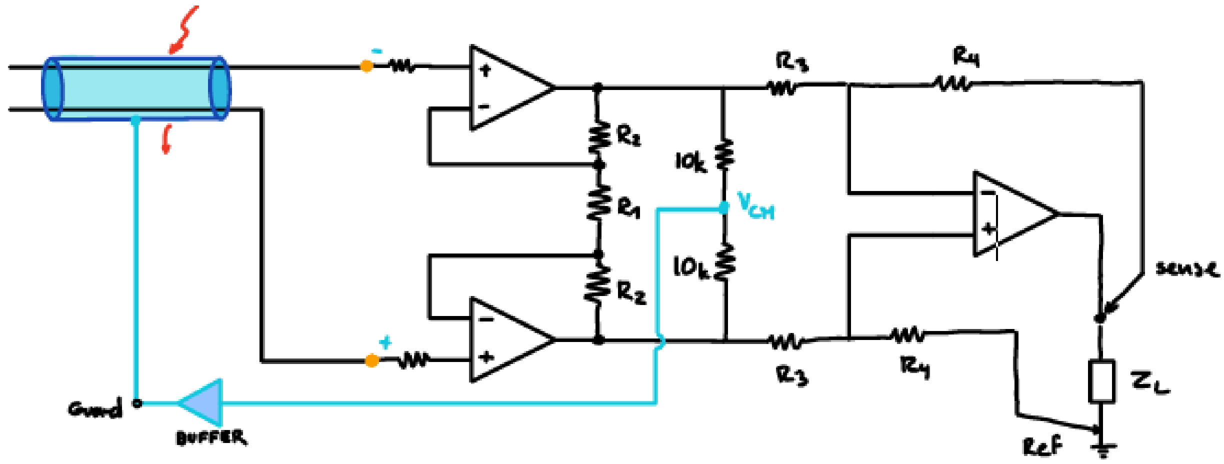
Shielding input cables

The cables connecting the electrodes to the INA can acts as antennas and become coupled with the external electromagnetic interferences.

To avoid this we shield the cable and connect it to ground so that any charge generated by the external waves will be discharged before reaching the cables.

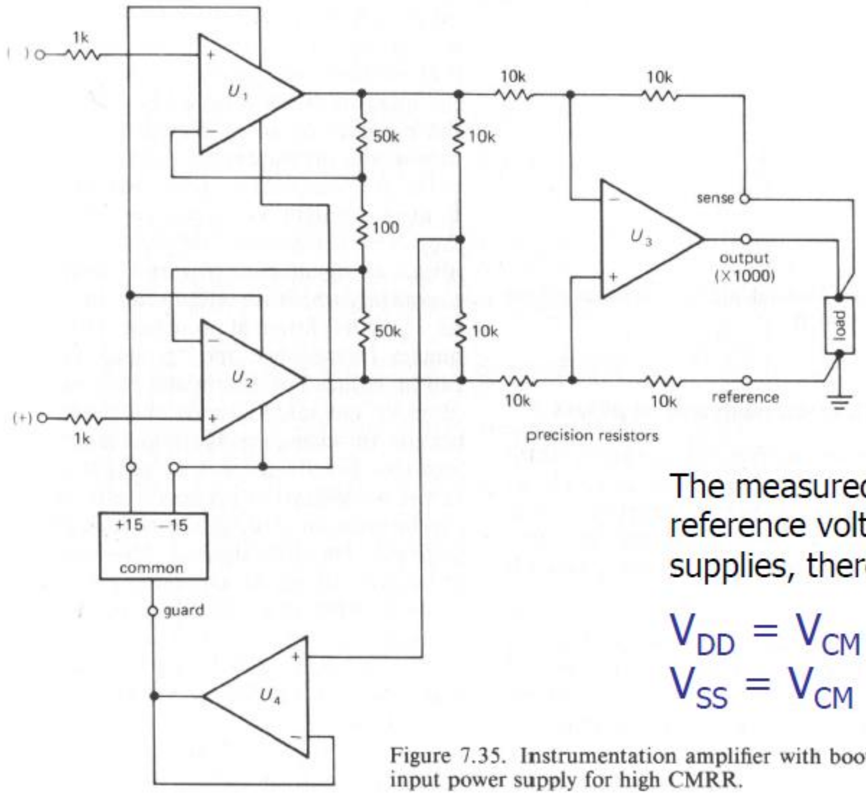
This however create a capacitive coupling between the cable and its shielding, to avoid this we connect the shield not to the ground but the guard pin of the INA whose voltage is equal tot eh common mode voltage of the circuit.





Bootstrap power supply

We can connect the power supply to the guard pin to further reducing the effect of the V_{CM} variations.



Bootstrap of power supplies (to reduce the effect of V_{CM})

The measured V_{CM} is provided to the reference voltage of the power supplies, therefore:

$V_{DD} = V_{CM} + 15V$
$V_{SS} = V_{CM} - 15V$

Figure 7.35. Instrumentation amplifier with bootstrapped input power supply for high CMRR.

(U_3,U_4 biased as usual as CMRR has lower impact than for U_1,U_2)

Common issues in ECG instrumentation

Saturation

A large voltage transient produced by some current (like in the case of defibrillation) can cause the INA to saturate, in this case measurements have no physical significance and an interval is required before they go back to normal

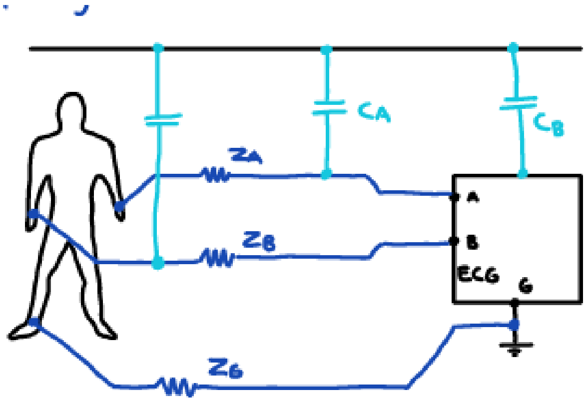
Common source interference

Other instrumentations and the $50{,}60Hz$ power line can interfere with eh ECG operation

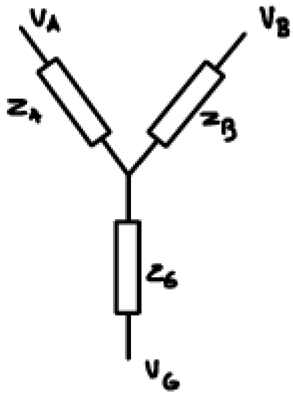
Coupling of the power line on the ECH

We could encounter a capacitive coupling between the power line and the electrodes.
This is a safety issue as the patient becomes a discharge path for the current since one of the electrodes is connected to ground through the patient.
We have both differential and common mode signals generated by the power line coupling.

Differential signal due to power line coupling



Because of the coupling effect a current flow in the electrodes A and B passing through the skin and discharging through the electrode G connected to ground.
We are going to have a different voltage drop depending on the value of the impedances of the 2 electrodes



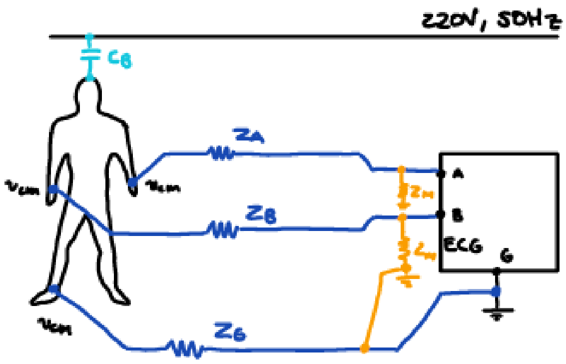
$$V_A - V_B = i_{d1}Z_1 - i_{d2}Z_2$$

So if $i_{d1} = i_{d2}$ we get

$$V_A - V_B = i_{d1}(Z_1 - Z_2)$$

This can be reduced by limiting the electrode impedances and shielding the cables.

Common mode signal due to power line coupling



The power line coupled with the body induces a common mode voltage.

$$V_{CM} = i_{dB} \cdot Z_G$$

While this original signal is in common mode some of it becomes differential because of the mismatch in the electrode impedances, we get

$$V_A - V_B = V_{cm}\left(\frac{Z_{in,INA}}{Z_{in,INA} + Z_1} - \frac{Z_{in,INA}}{Z_{in,INA} + Z_2}\right)$$

We have that if $Z_1, Z_2 \ll Z_{in}$ we get

$$V_A - V_B = V_{cm}\frac{Z_2 - Z_1}{Z_{in,INA}}$$

This means that the output voltage of the INA will be equal to

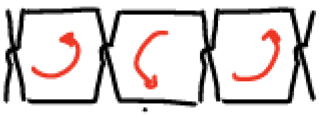
$$V_{OUT} = G_d \cdot V_{bio} + \frac{G_D \cdot V_{CM}}{CMRR} + G_D V_{CM}\left(1 - \frac{Z_{1M}}{Z_{1M}+Z_1-Z_2}\right)$$

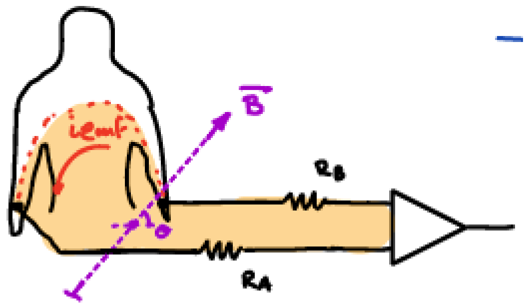
limited CMRR of the INA.

diff. signal due to mismatched impedance.

Magnetic coupling

Magnetic fields may induce MF in our wires, to prevent tis the wires are twisted together so that the total magnetic flux is zero



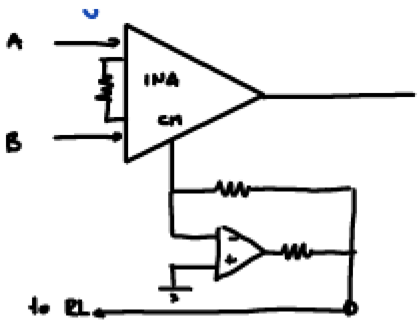


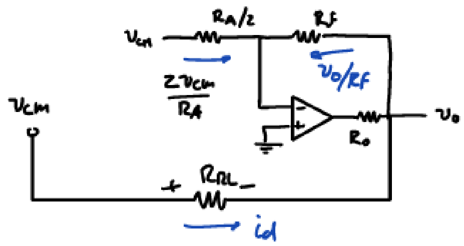
Components for overvoltage protection

Another important issue is protecting the circuit from high voltage transient this can be done by using diodes to block the voltage when it increases too much.

Reducing the common mode voltage

We can further reduce the common mode voltage by utilizing the common mode node of the INA to take the voltage with which bias the load.





$$\frac{2V_{CM}}{R_a} + \frac{V_o}{R_f} = 0$$

$$V_o = -\frac{2R_f}{R_a} V_{cm}$$

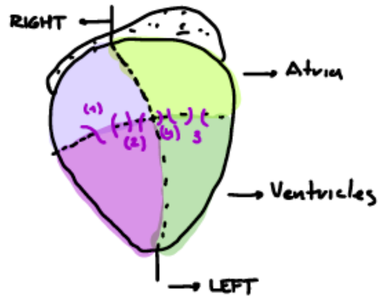
$$V_{cm} = R_{RL} \cdot i_d + V_o$$

$$V_{cm} = \frac{R_{RL}\, i_d}{1 + \frac{2R_f}{R_a}}$$

Improvement factor

The heart

Structure of the heart



the heart can be divided into compartments each with a different role within the cardiac cycle.

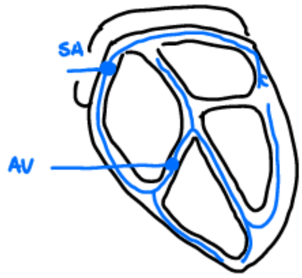
- The upper compartment are called atria while the ones on the lower part are called ventricles
- The compartments are connected by valves

Operation of the heart

The cardiac activity is regulated by electrical pulses, the distribution of these pulses is achieved through a series of nodes.

The 2 most significant nodes are

- The **sinoatrial node SA**
- The **atrioventricular node AV**



Intrinsic properties of the cardiac cycle

A healthy heartbeat is characterized by 2 intrinsic properties

1) **Automaticity**: capability to start a beat autonomously
2) **Ritmicity**: regularity in the beats

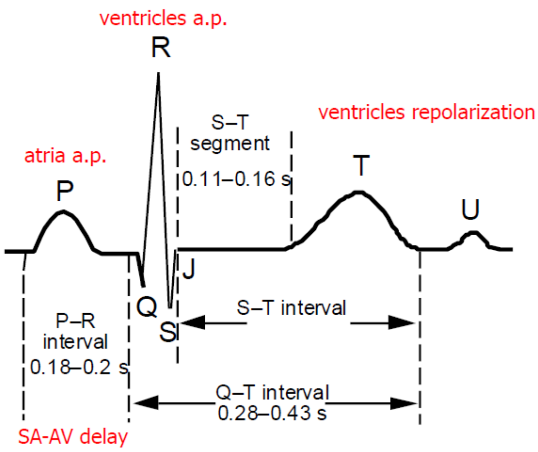
Spontaneous depolarization

The heart nodes have each different natural rhythm

- **S.A. node** around 70bpm
- **A.V. node** around 50…60bpm
- **Purkinje fibers** around 25…40bpm

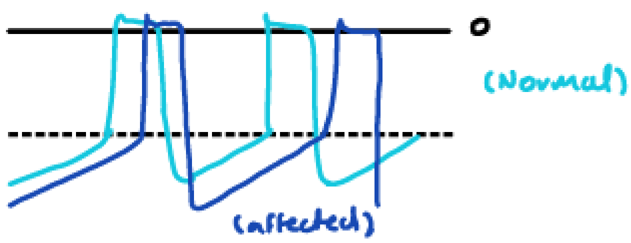
The rhythm of the while cardiac cycle is determined by the self Ritmicity of the S.A node as the cardiac stimulus starts at the SA node and propagates through the ventricles.

Cardiac cycle ECG waveform

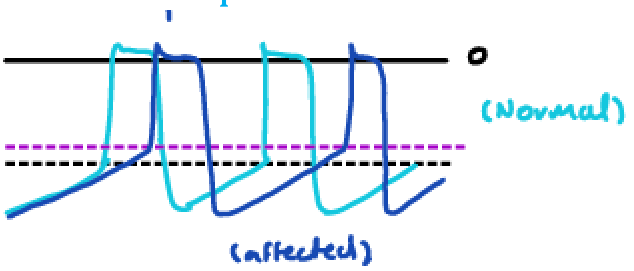


the normal heart operation can be altered by a variety of factor, here we review the main ones.

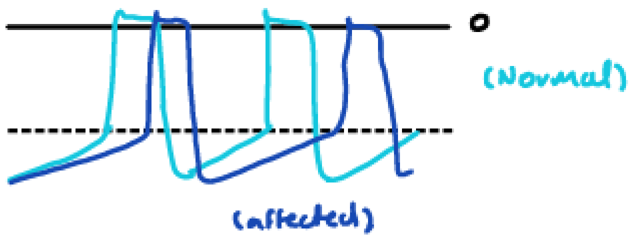
Diastolic potential moder negative



Threshold more positive



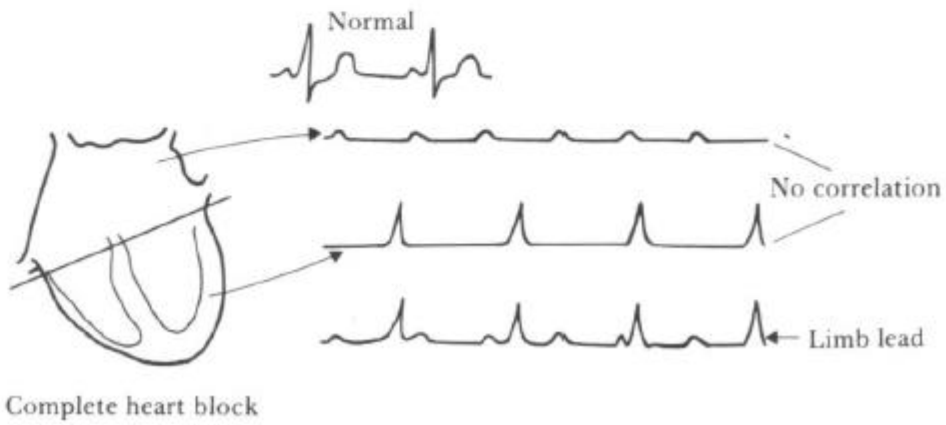
Lower depolarization slope



Examples of arrhythmias

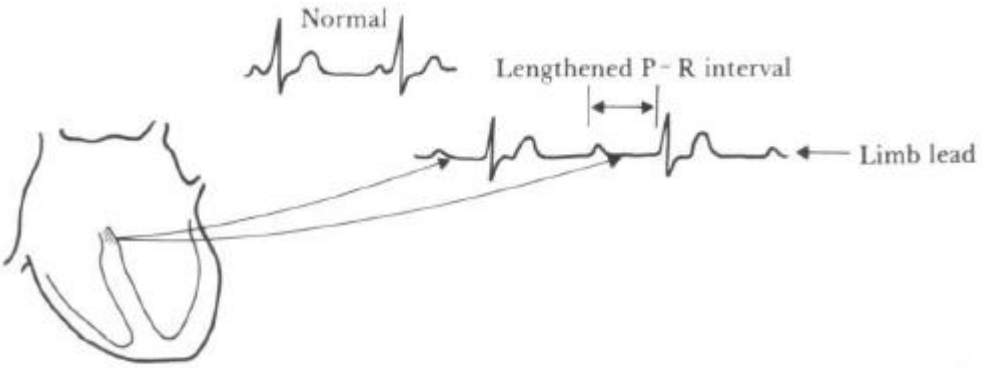
AV node inactive

In this case we have a cardiac block as there is no longer any correlation between the signals of the other nodes



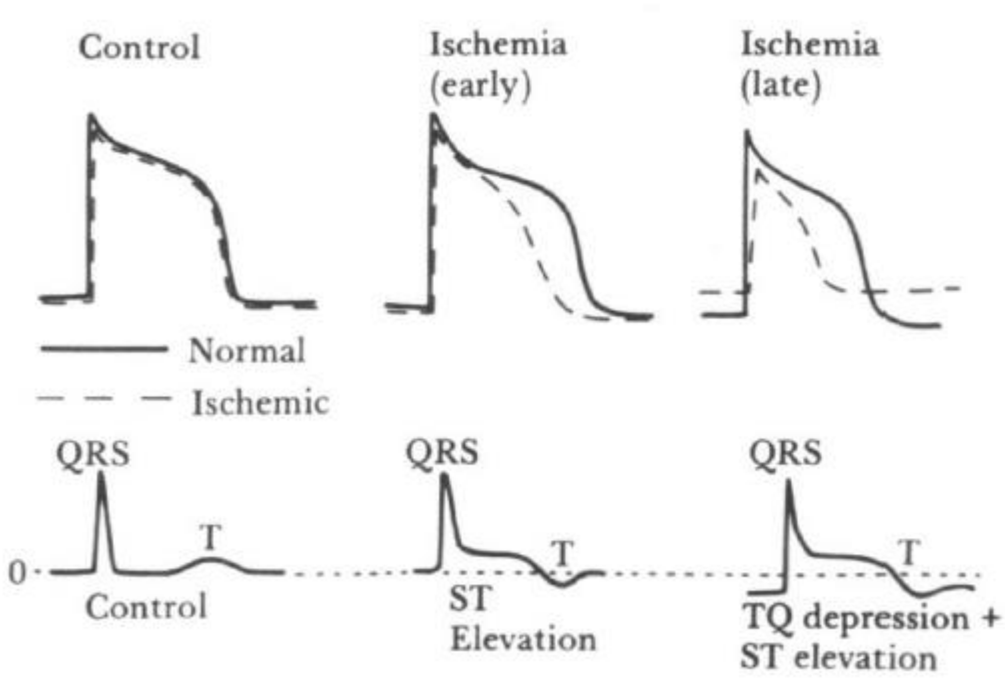
AV node partially operative

In this case we have a delayed transfer of stimulus between atria and ventricles



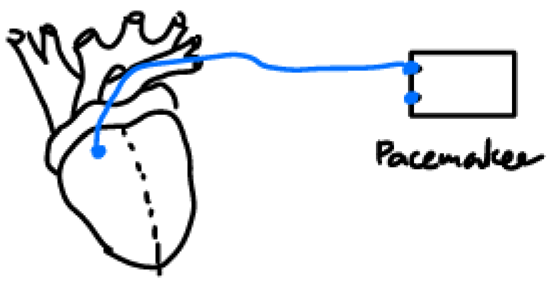
Ischemia

An ischemia is an occlusion of the coronary arteries which reduces the blood distribution in the heart, the consequent reduction in oxygenation changes in the electrochemical equilibrium of tissues as now the concentration of k^+ increases while that of Na^+ increases, thus the resting potential change shape as wehll as the action potential



The pacemaker

The peacemaker is an electronic device implanted to aid in the normal functions of the heard.



The pacemaker produces electrical pulse by electrodes on the surface of the muscle or inside the cavity, these pulses normalize the regular and periodic contractions of the heart.

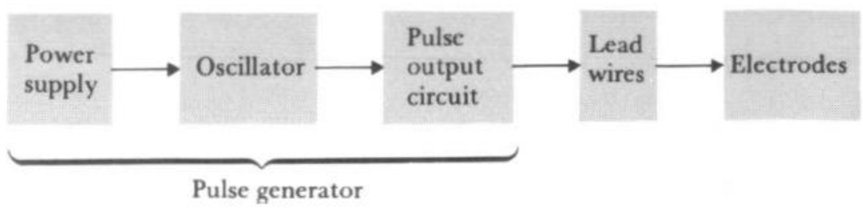
Classes of pacemaker

We can divide pacemakers into 3 categories

1) Asynchronous
2) Synchronous
3) Rate adaptive

Asynchronous pacemakers

There were the first kind produced, they supply pulses in a free running modality at a fixed rate without referring to the physiological parameters of the body.
Their structure is very simple



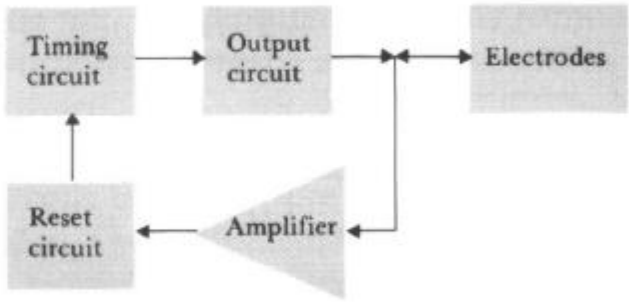
Synchronous pacemakers

We can distinguish 2 types of synchronous pacemakers.

- Demand type
- Atrial type

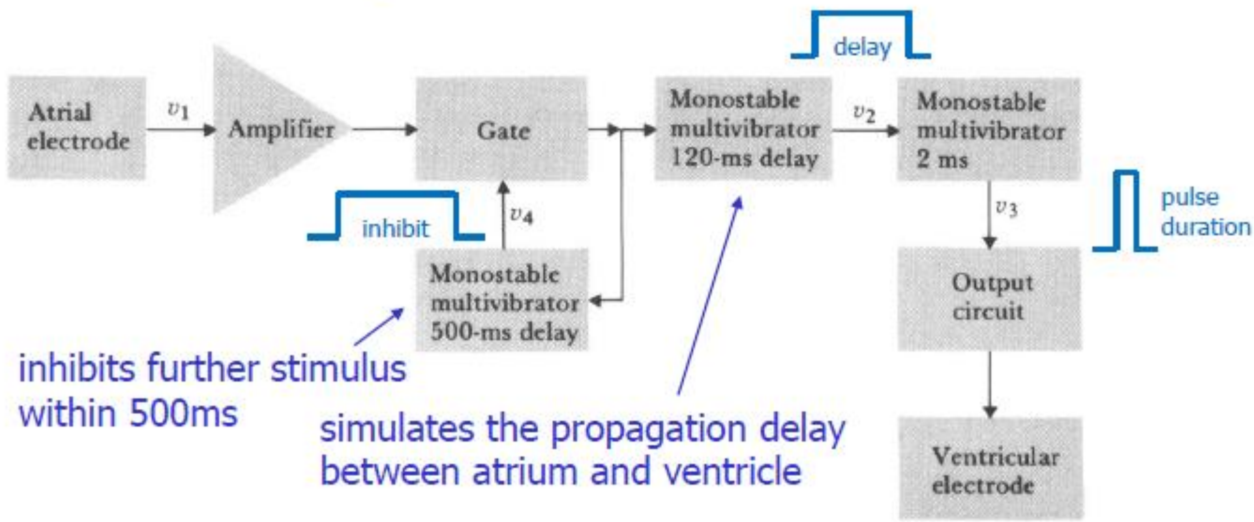
Demand type pacemakers

These pacemakers sample the heart activity and provide a pulse id one isn't detected, otherwise if one is present the timer is reset to its initial stage.
Electrodes serve both as a way to apply the stimulus and to detect if the natural electric signal is present.
We can see that in this case we have a control included in the block diagram.



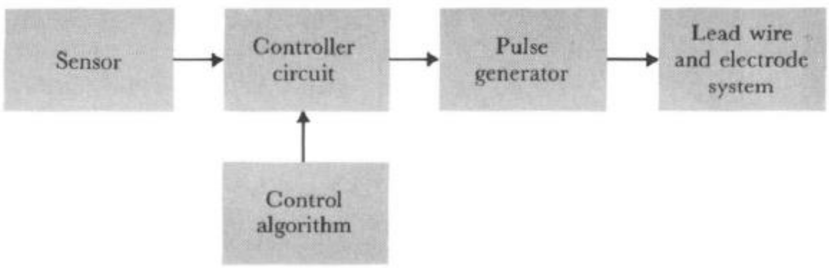
Atrial type pacemakers

These pacemakers substitute a not correct transmission of the stimulus from the atrium to the ventricle, taking the role of the AV node.
What they do is detect electric signals corresponding to the contraction of the atria and the uses appropriate delay to activate a stimulus pulse for the ventricles

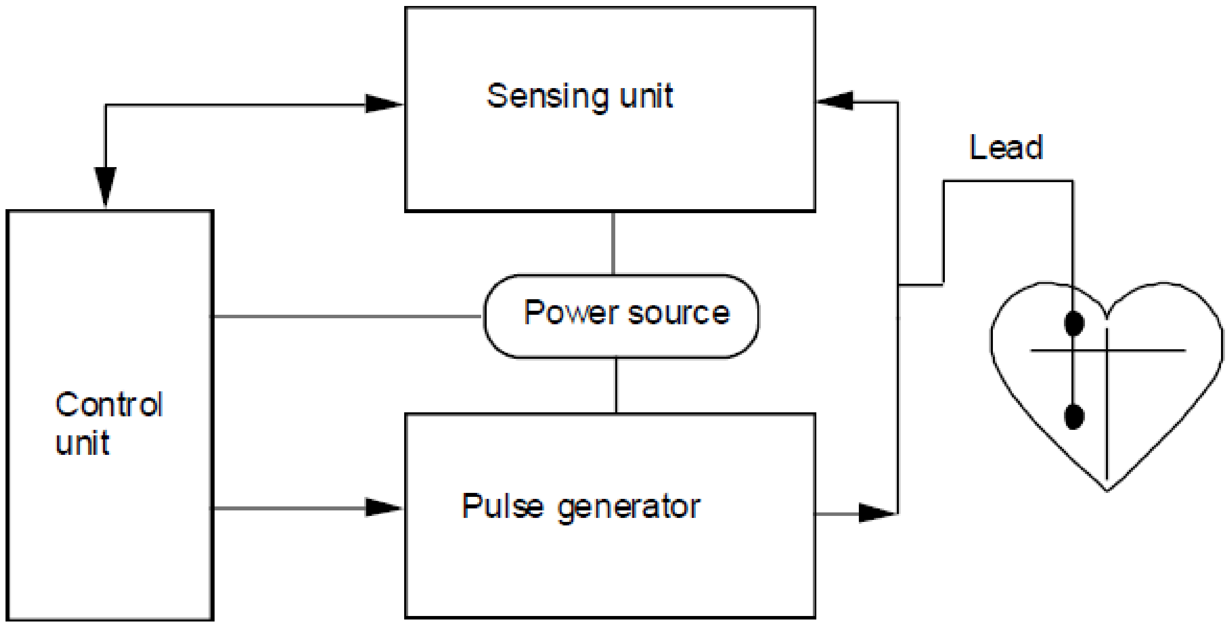


Rate adaptive

These pacemakers operate according to the acquisition by sensors of physiological parameters of the patient adjusting its own stimulus to the actual status of such parameters by means of a control system.



Basic components of a pacemaker



Electrodes of a pacemaker

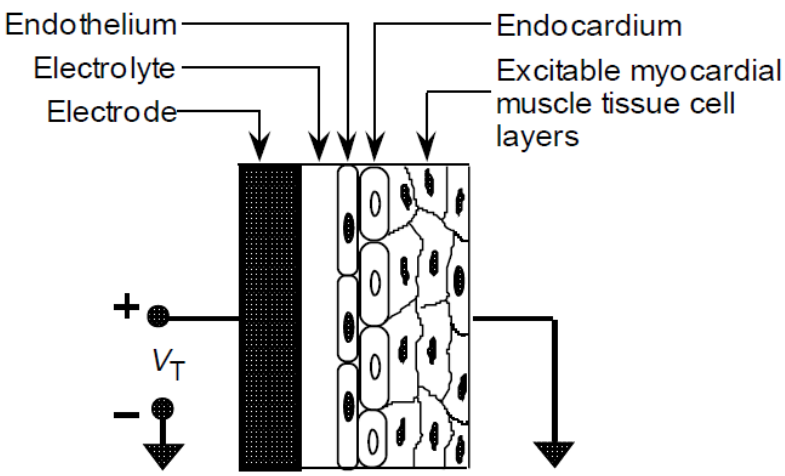
The electrodes are responsible for

- The **application** of the electric stimulus to the cardiac tissue
- The **measurement** of the electric potential of the tissue

When designing the electrodes we need to take into consideration 4 factors

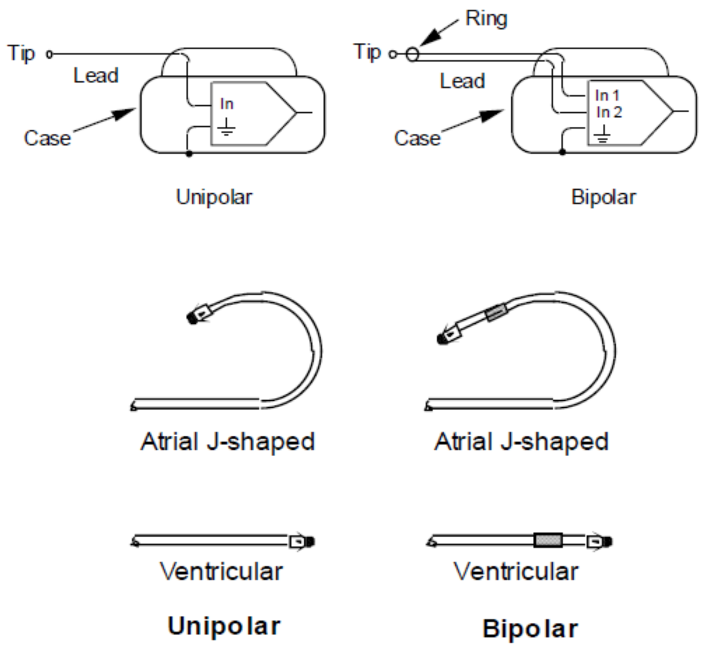
1) **Safety**: the cardiac stimulation should be safe for the patient
2) **Energy**: the losses of energy in the electrode should be minimized
3) **Biocompatibility**:
4) **Dimensions**: should not be invasive and easily implantable

Electrode and myocardium interface



Unipolar/bipolar electrodes

Electrodes can be unipolar or bipolar: unipolar electrodes only have a cathode while bipolar designs include a ring anode approximately 10mm from the cathode to also provide the negative voltage.



Battery

Power budged computation

Average current available

The pacemaker should be able to operate within the patient for a long time without interference.
the main power draw will be that of the pulse generation, this meant hat if we consider an operating life of 10 years and a battery with a capacity of $2000mAh$ the average current draw over the time period should be

$$I_{avg} = \frac{capacity}{duration} \simeq 23\mu A$$

Pulse current

To provide a $10mV$ pulse over a 1Ω load we need a pulse of about $10mA$ thus we obtain an average current drawn for the pulse generation of

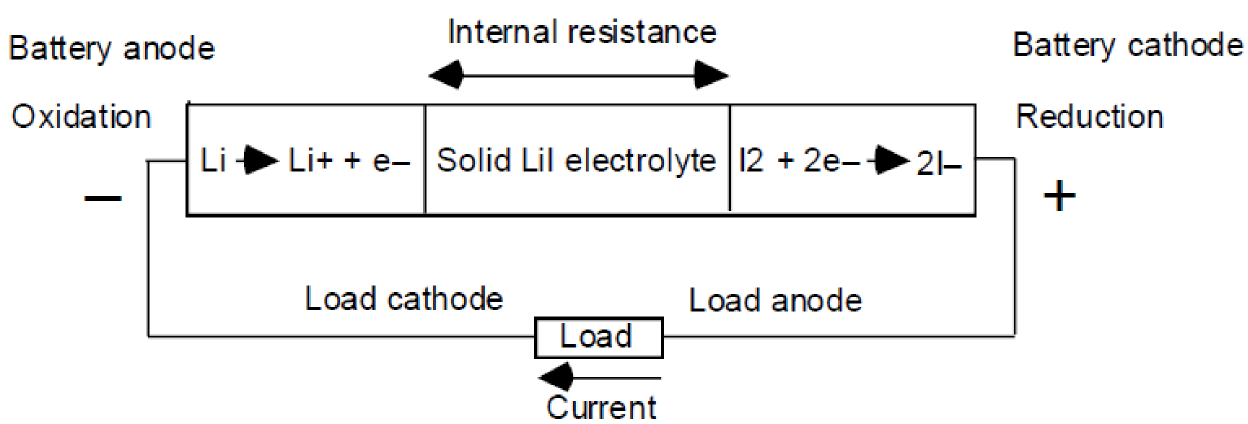
$$I_{avg,puls} = 10mA \cdot \frac{1ms}{0{,}7s} = 14\mu A$$

Remaining power

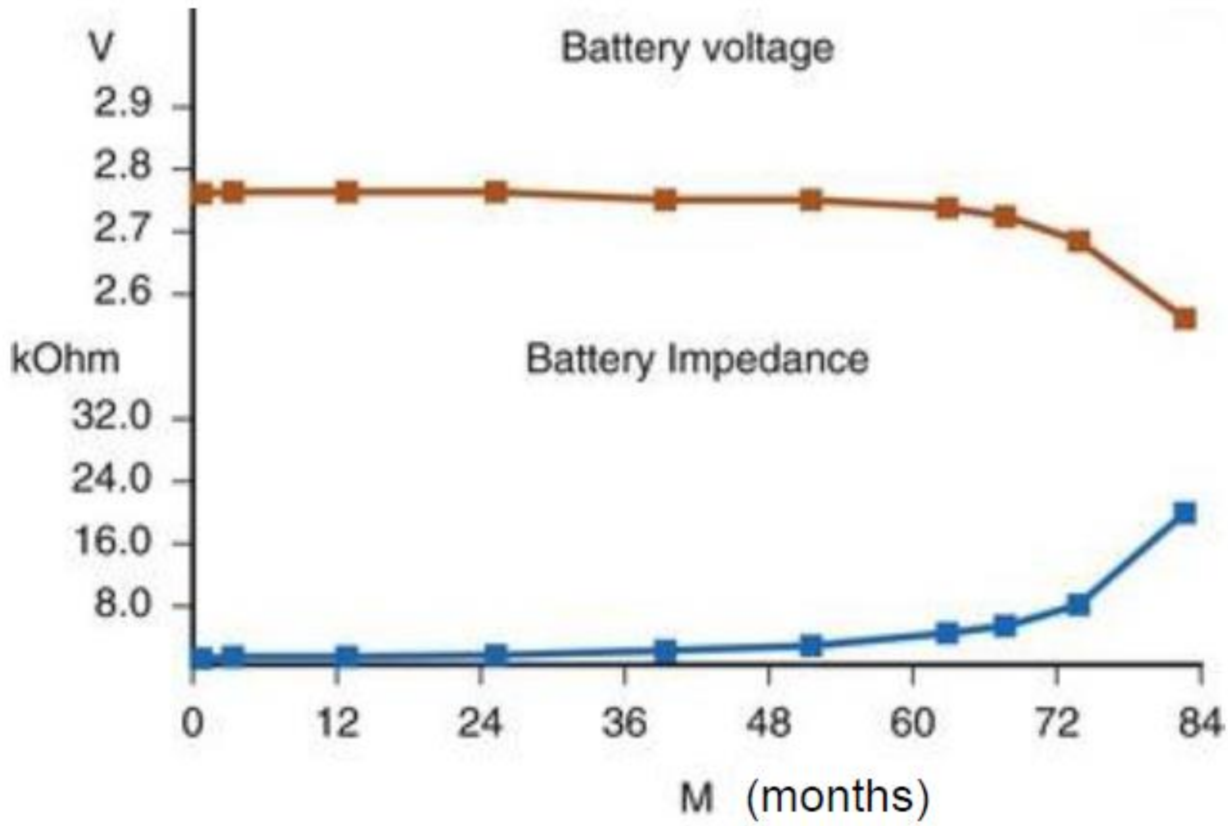
This means that the current available for the remaining electronics is about $9\mu A$, thus the pacemaker is an **extremely low power application**.

Structure of a battery

The typical battery for electronic application is made of lithium iodide



It is a solid state battery with a typical voltage of $2{,}8V$, as time progresses the electrons are depleted causing an increase of the solid lithium electrolyte as well as of the internal resistance, as a consequence the output voltage will be smaller since a larger portion will fall on this internal resistance.

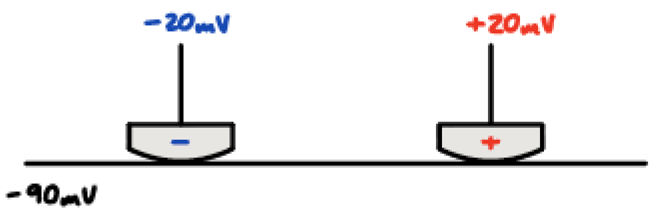


Pulse generation

The pulse is fundamental to the operation of the pacemaker, so we are going to study in detail how it is generate applied and its effect on tissue.

Effect of the pulse on the tissue

let's consider the application of a $\pm 20mV$ bipolar pulse to a tissue with a $-90mV$ resting potential.

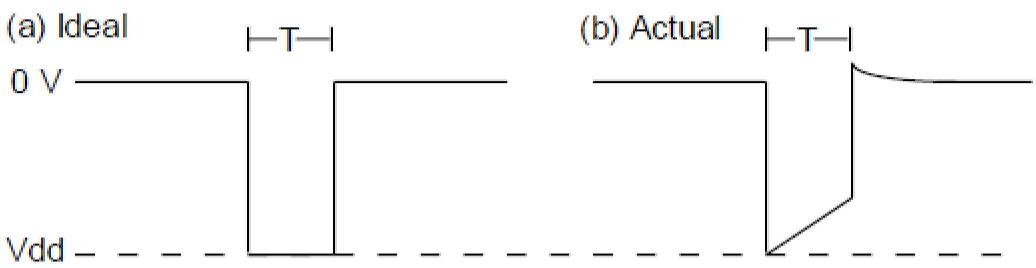


The overall potential difference is

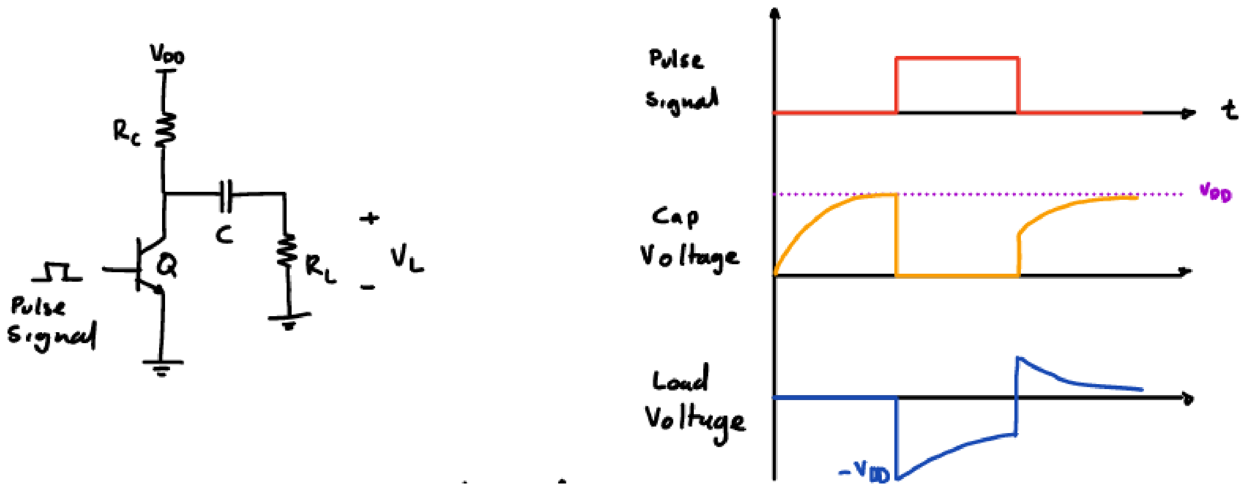
- On the positive contact
$$\Delta V^{+} = -90mV - 20mV = -110mV$$
this is even further away from the threshold and action potential is not triggered
- On the negative contact
$$\Delta V^{-} = -90mV - (-20mV) = -70mV$$
This is past the threshold so we trigger the action potential

Shape of the pulse

Note that the pulse shape will not be and ideal square



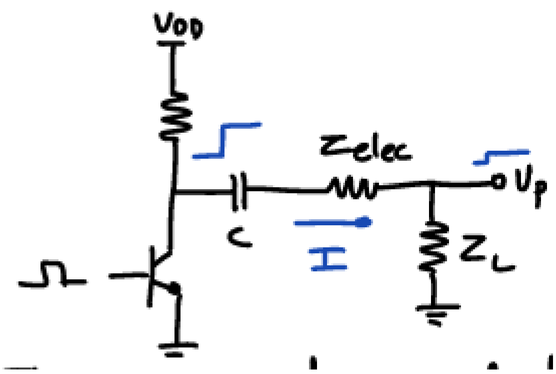
Pulse generation circuit



Let's assume that C starts discharged so it has $0V$ across it, when the BJT is off the capacitor starts to charge exponentially with a time constant $\tau = C(R_C + R_L)$ towards V_{DD}.

Once we turn on the BJT the capacitor is connected to ground and since its voltage can not change instantly the voltage across R_L will instantly decrease.
Then the capacitor will start to discharge with a time constant $\tau = C \cdot R_L$

Effect of the electrode impedance



the electrode impedance greatly attenuates the pulse before it arrives to the tissue because $Z_{EL} \gg Z_L$ so the majority of the voltage drop will occur on the electrode rather than the tissue.

We can see that the current needed to provide the pulse canbe found as

$$I = \frac{V_{DD}}{R_L + Z_{elec}}$$

Since as we said $Z_{EL} \gg Z_L$ we can write

$$I \simeq \frac{V_{DD}}{Z_{elec}}$$

This means that **the current is essentially independent form the load and we can increase it by modifying the value of the supply voltage**.

Optimization of the pulse amplitude/duration

Due to the attenuation we will need a large V_{DD} to generate the pulse of $20mV$ we desire.
The membrane can be described as a load with an RC characteristic so the voltage variation is equal to

$$V = IR_L\left(1 - e^{\frac{t}{RC}}\right)$$

Note that we indicate the current rather than the voltage V_{DD}
The asymptotic value is

$$V = I \cdot R_L$$

We have that the minimum current value to reach $20mV$ is thus going to be

$$I_0 = \frac{20mV}{R_L} = 20mA$$

Under this condition however the threshold will be reached after infinite time so we need first to select a higher value $I = \alpha I_0$, there is a trade off

An higher current will allow for a faster pulse

What is the best compromise between current amplitude and pulse duration?

$$V_0 = \alpha I_0 R_L\left(1 - e^{\frac{t}{R_L C}}\right)$$

$$\frac{V_0}{\alpha I_0 R_L} = 1 - e^{\frac{t}{R_L C}}$$

$$-\frac{t}{RC} = ln\left(1 - \frac{V_O}{\alpha I_0 R}\right)$$

We now approximate $ln(1 + x) \simeq x$ and write

$$\frac{t}{RC} = \frac{V_O}{\alpha I_0 R}$$

We have that $\alpha I_0 = I$ so we can write

$$t = C\frac{V_O}{I}$$

Solving by I we get

$$I = C\frac{V_O}{t}$$

Then we multiply and divide by I_0

$$I = C\frac{V_O}{t} \cdot \frac{I_0}{I_0}$$

We define $t_c = C\frac{V_0}{I_0}$ and obtain

$$\boldsymbol{I = I_0\frac{t_c}{t}}$$

This result is correct for t small, while for larger t we use

$$\boldsymbol{I = I_0\left(1 + \frac{t_c}{t}\right)}$$

Now we can compute the result which gives us the minimum energy consumption by writing the energy formula and computing the derivative

$$E = I^2 \cdot Z \cdot t$$

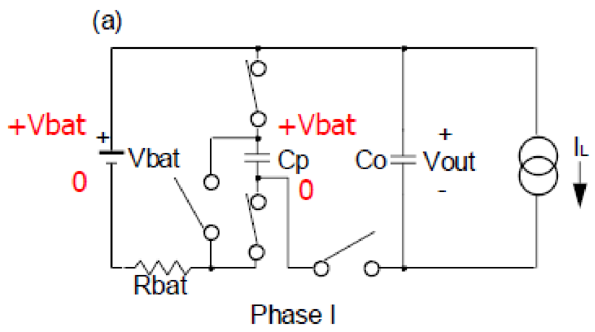
The time derivative gives us

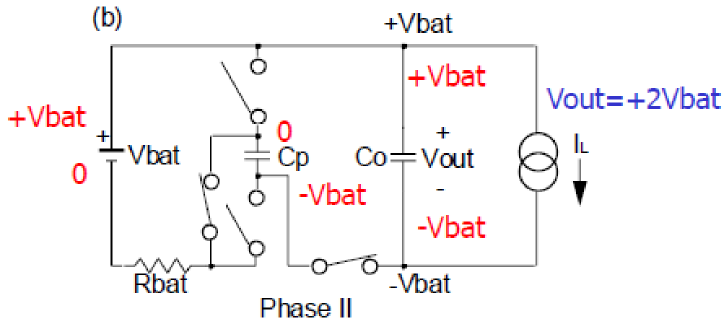
$$\frac{dE}{dt} = I_0^2 Z\left(1 - \frac{t_c^2}{t^2}\right)$$

This is equal to zero for $t = t_C$ which gives us $I = 2I_0$.

Creating a voltage larger then the power supply

We might be in a condition in which the battery voltage is not high enough for us, for this reason we need to create a voltage higher than the power supply



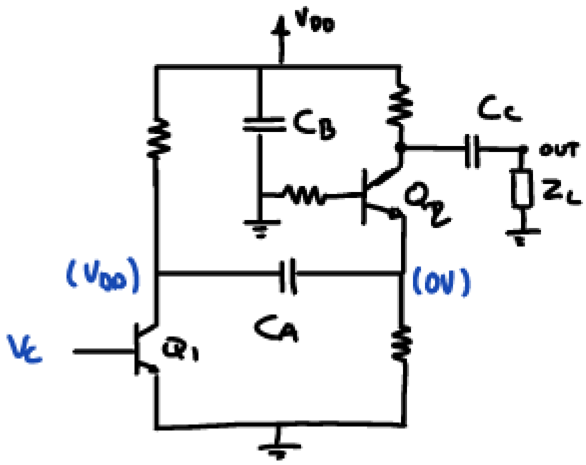


In this circuit we have 2 phases.

In the first phase the capacitance C_P is charged up to V_{Bat} and the output is also connected to V_{BAT}

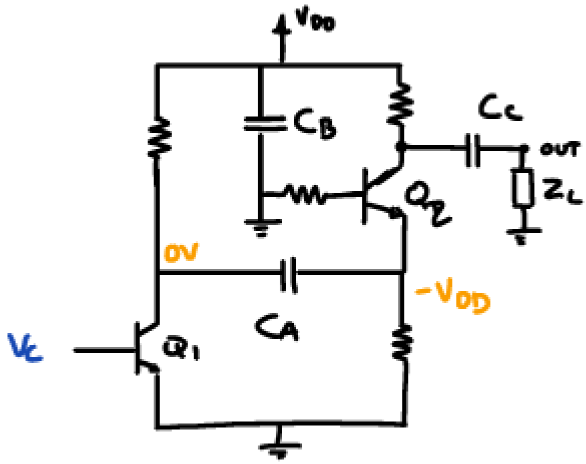
In the second phase the switches commute and now the battery and the capacitor C_P are in series so that the total votlage drop across the output is equal to the sum of their voltage and thus to $2V_{bat}$.

Unipolar pulse generator



OFF PHASE Q_1 off

In this condition all capacitors will charge up to V_{DD}, the output will be $0V$ and the switch Q2 will also be off as both its base and its emitter are at $0V$.

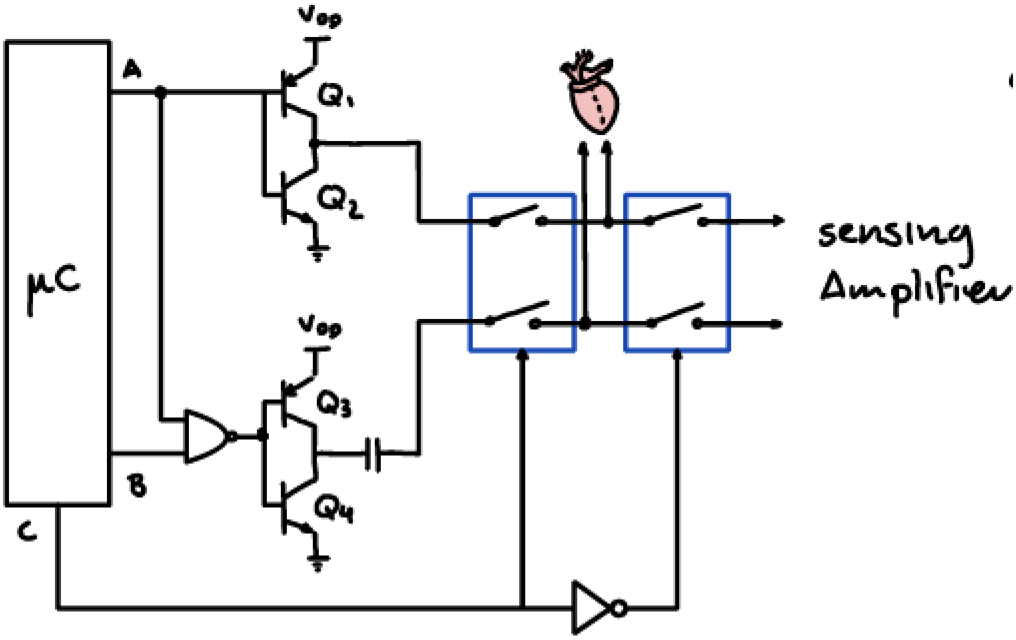


On phase Q_1 on

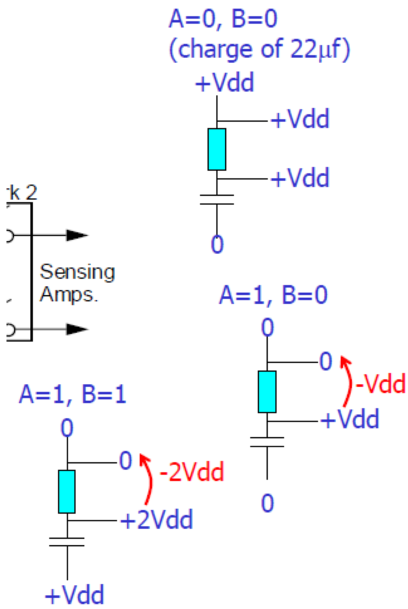
Now C_A is connected directly to ground forcing the emitter of Q_2 to $-V_{DD}$.

this means that now Q_2 is turned on and thus forces the collector voltage to $-V_{DD}$.
Since the capacitance C_C is still charged to V_{DD} the resulting output voltage will be $-2V_{DD}$.

Bipolar pulse generator



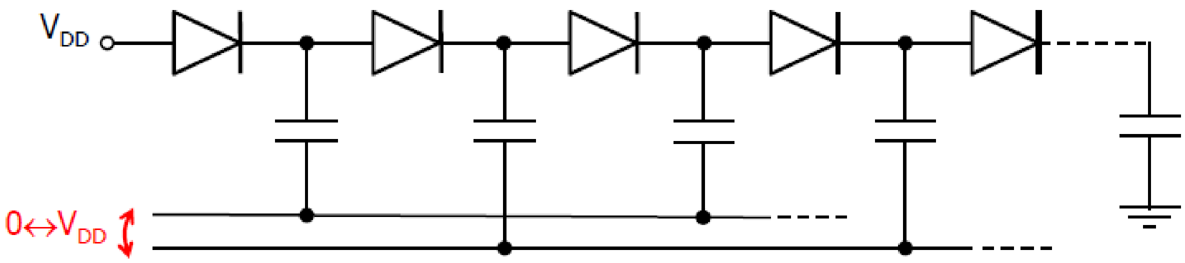
We have 2 input inverters and 2 switch networks which we use to connect to either one or both inverter in the case where the battery voltage decreases too much.



Charge pumps

We have seen how it is important to generate voltages larger then the power supply, to do so we utilize charge pumps, one example of charge pumo is the Dickson charge pump

Dickson charge pump

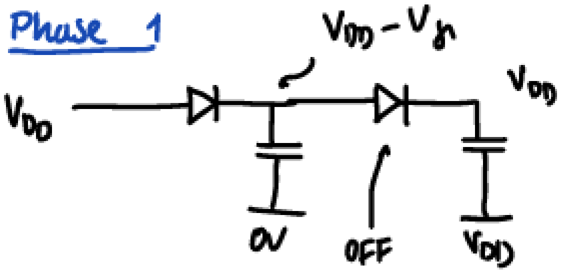


We have a chain of diodes and capacitors.
The capacitors are connected alternatively to 2 different lines which are switched form 0V to V_{DD} periodically.

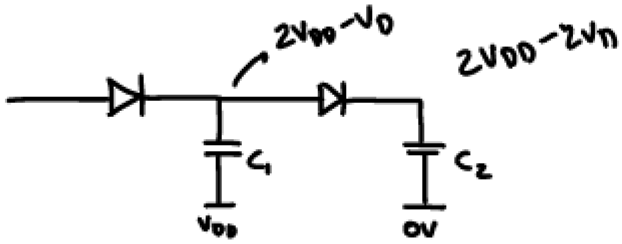
The circuit can be studied by considering the basic cell of 2 diodes and 2 capacitors operating in 2 different phases

Phase 1

The first diode is turned on, so current flows through it charging up the capacitor, instead the second diode is off. After a transient the first capacitor is charged to a value equal to $V_{DD} - V_D$



Phase 2



Now the voltages at the ends of the capacitors are switched.
The first diode is off a d the total voltage at the end of the first capacitor includes now an additional term V_{DD} so we have $2V_{DD} - V_D$.

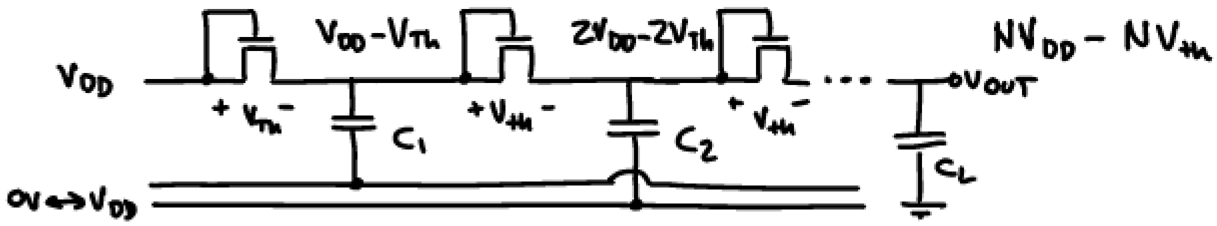
The second diode is now turned on and current flows charging the second capacitor to the voltage $2V_{DD} - 2V_D$.

If we chain multiple cells or we repeat the process multiple time the voltage at the output increases with the formula

$$\boldsymbol{V_{out} = NV_{DD} - NV_D}$$

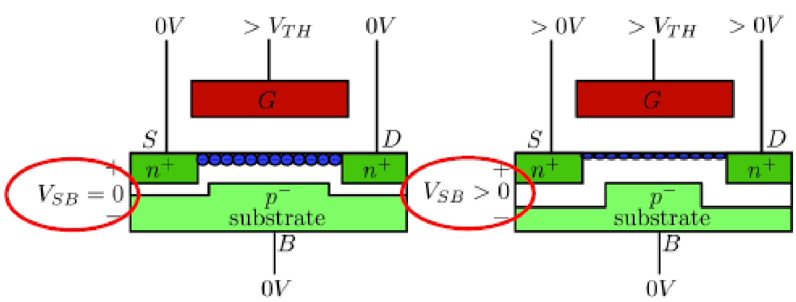
Dickson charge pump using MOS transdiodes

Rather then using diodes we utilize MOSFETs in transdiode configuration as this allows us to use less area and lose less voltage with each stage



Body effect

The source terminal of the MOSFETs s connected to progressively higher voltages, this increases the width of the depletion layer in the MOSFET making it more difficult to create a channel with the same V_{GS} and effectively increasing the threshold voltage.

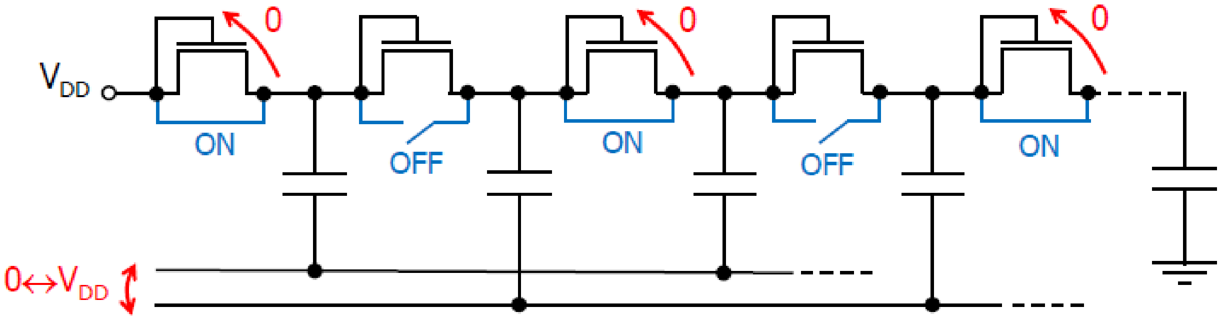


This phenomenon is called body effect and can be computed with the formula

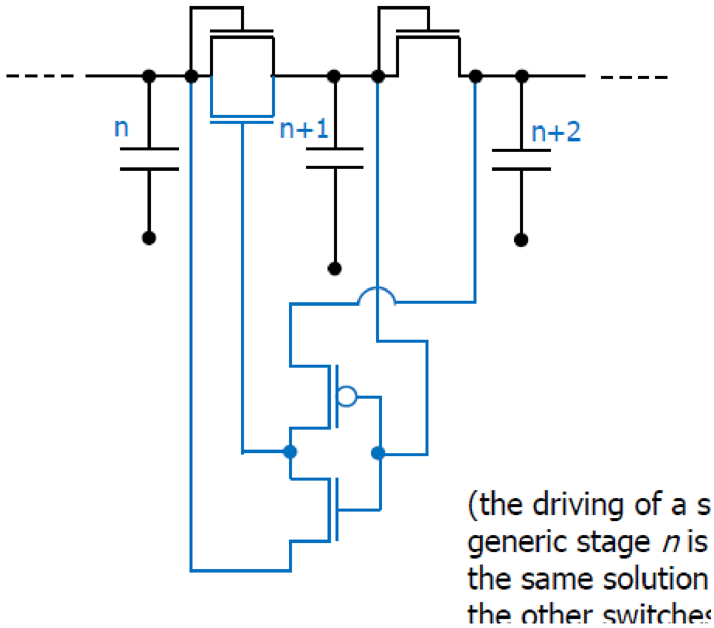
$$V_{TH} = V_{T0} + \gamma\left(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}\right)$$

Solution

We add a switch in parallel with the transdiode to short drain and source eliminating the body effect.
Note that these switches must be closed when the transdiode is on and open when the transdiode is off



How to drive the switches



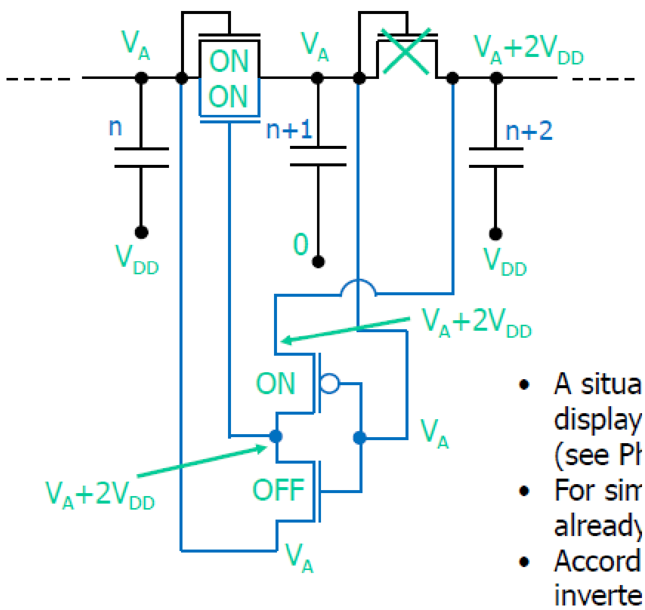
To drive these switches with the correct timing we utilize an inverter connected as shown

On state

With this solution the switches will act as desired without the need for an external control.
As we can see in this condition we have that

- The positive voltage of the inverter is $V_A + 2V_{DD}$
- The negative voltage of the inverter is V_A
- The input voltage is equal to V_A and thus to the low value
- The output voltage will be equal to $V_A + 2V_{DD}$

So the switch will be driven with an high gate input and turn on

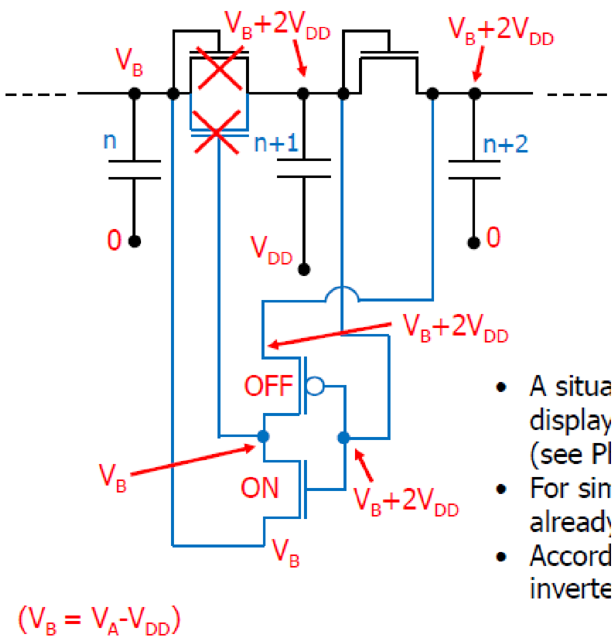


Off state

in this condition we have that the new voltage V_B is equal to $V_A - V_{DD}$ since the voltages at the terminals of the capacitors have switched.

This means that now the voltage at the input of the inverter is no longer the low voltage but the high voltage.

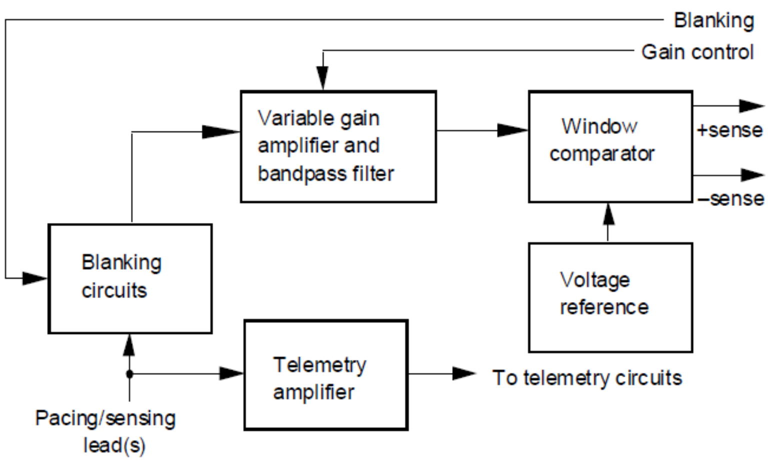
Now its output will be low and thus the switch in parallel with the transdiode will be off.



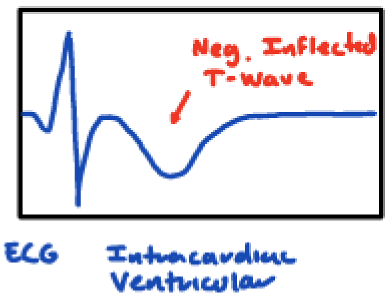
Important

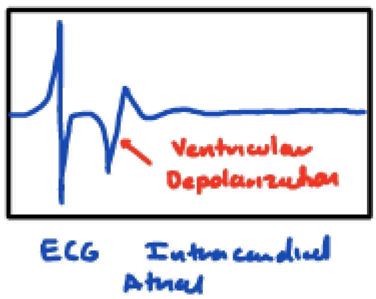
We can not remove the transdiodes and use only the switches as we require the inverters are able to operate only as long as the chain is already supplying the voltages. Without the diodes there would not be the voltages required to drive the inverters.

Sensing amplifiers and pacemakers



The sensing amplifier measures the cardiac activity using the same electrodes we use to send the signal.

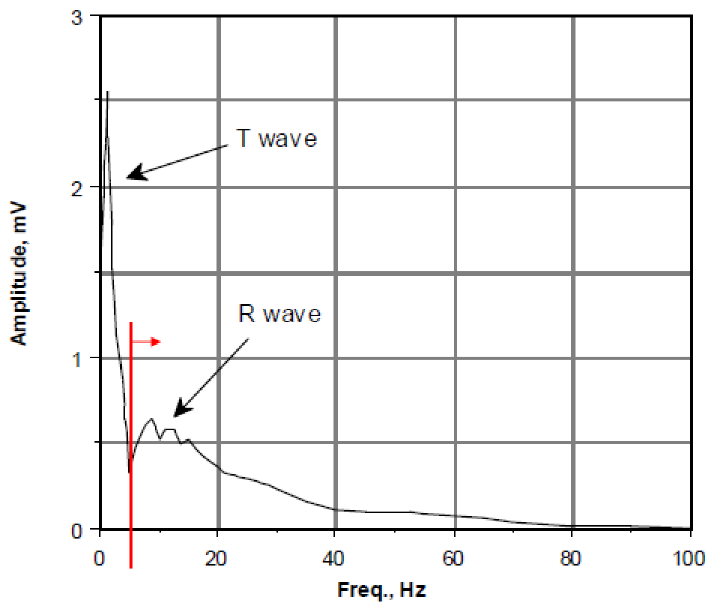




ECG intracardiac ventricular frequency spectrum

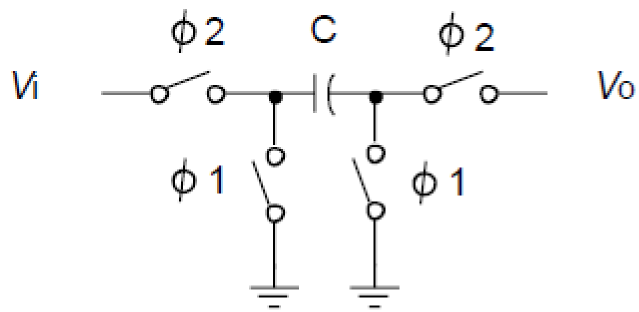
As we can see the frequency range in which we are interested in is very narrow, this means that we are going to need huge resistors to implement a correct filter.

This is problematic because resistances with high value require also a lot of space, so we need to find an alternative implementation.



Switched capacitor amplifiers

In this solution we charge and discharge a capacitor at high frequency so that its behavior on the slower time scale of our signal appears to be that of big resistor.



For this solution to be applicable we need 2 prerequisites

1) The 2 clocks we use to drive the switches do not overlap
2) The signal is much slower than the clock frequency

Operation

The switches are driven 2ith 2 clocks when the switches 1 are on the capacitor is discharged while when the switches 2 are on the capacitor is connected in series between input and output.

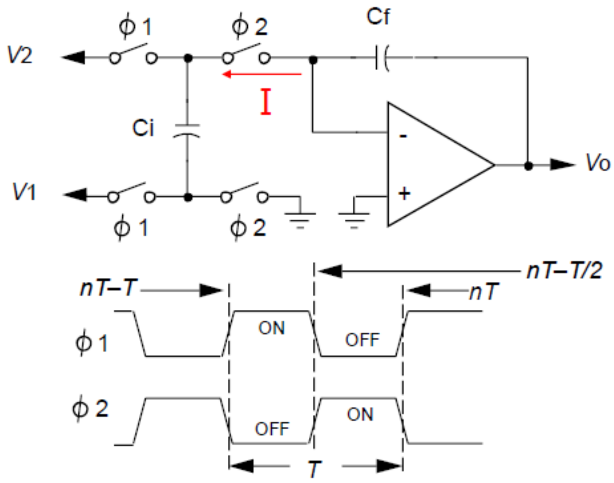
Under this conditions the effective resistance seen from a signal with slower frequency than the clock is

$$R_{eff} = \frac{V_I - V_O}{I} = \frac{V_I - V_O}{fC(V_I - V_O)} =$$

$$\boldsymbol{R_{eff} = \frac{1}{fC}}$$

This means that with a small capacitance $C = 2pF$ and a clock frequency $f = 10kHz$ we get an effective resistance $R_{eff} = 50M\Omega$

Switched capacitor integrator



the differential input $\Delta V = V_2 - V_1$ sees an effective resistance

$$R_{eff} = \frac{1}{fC_i}$$

The output outage will be equal to

$$V_O = \Delta V_i \frac{1}{sR_{eff}C_F}$$

So the total gain is

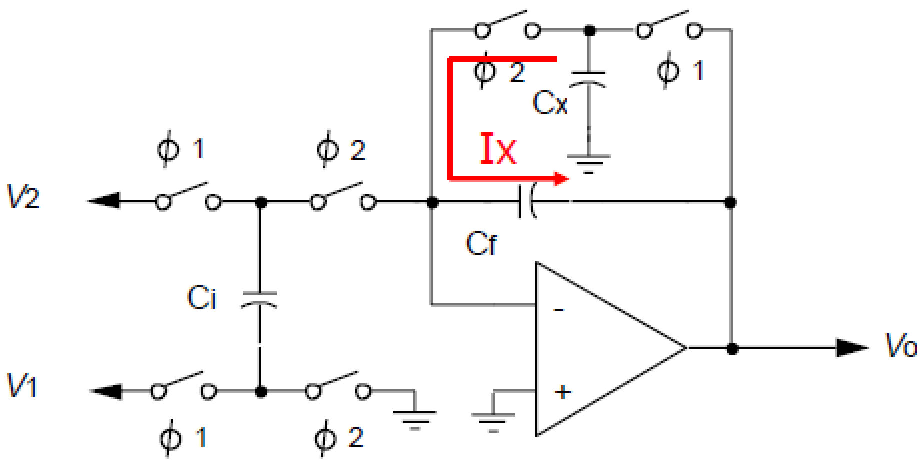
$$\frac{V_O}{\Delta V_i} = \frac{1}{sR_{eff}C_F} = \frac{fC_i}{j\omega C_f}$$

So we get

$$\boldsymbol{\frac{V_O}{\Delta V_i} = \frac{C_i}{C_F}\frac{f}{j\omega} = \frac{C_i}{C_F}\frac{1}{j\omega t}}$$

Real integrator with switched capacitor

To avoid saturation a real integrator requires a resistor to be placed in parallel with the feedback capacitance to allow passage to the bias currents, this resistor will also be implemented as a switched capacitor



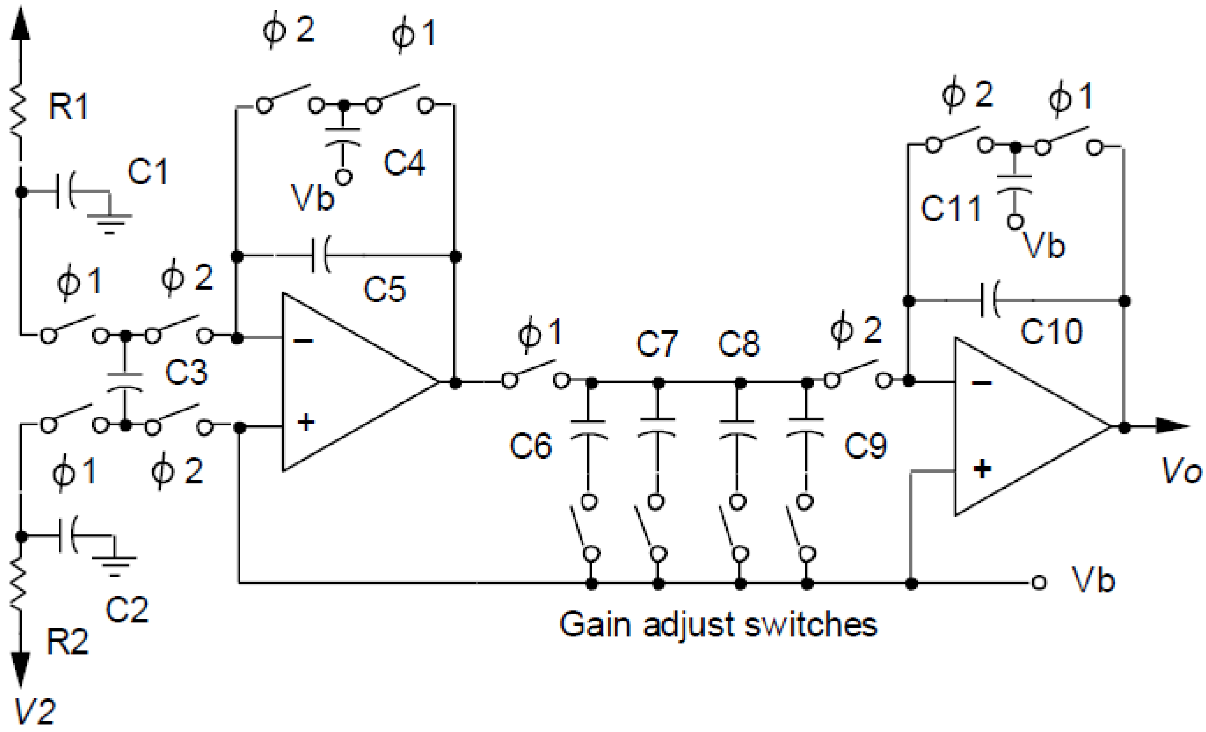
Again the equivalent resistance of the switched capacitor is equal to

$$R_F = \frac{1}{fC_X}$$

The transfer function will now present a pole not at zero but at a slightly higher frequency $\frac{1}{2\pi R_F C_F}$

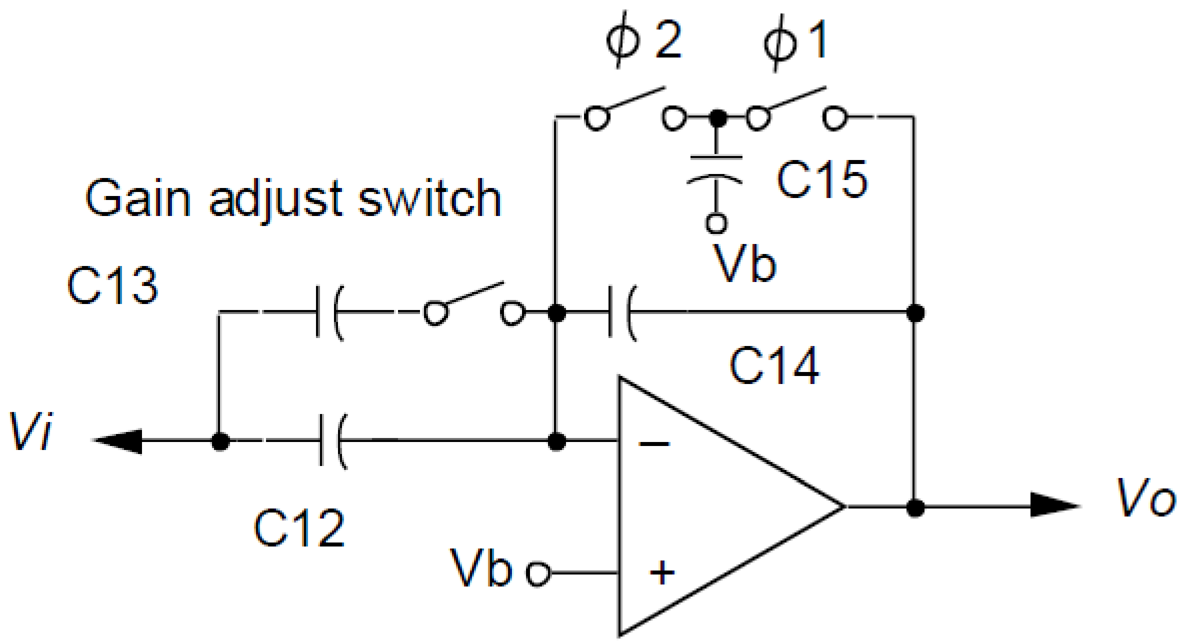
$$\boldsymbol{\frac{V_O}{V_i} = \frac{C_i}{C_x}\frac{1}{1 + sC_F R_F} = \frac{C_i}{C_x}\frac{1}{1 + sC_F \frac{1}{fC_x}}}$$

Switched capacitor differential input amplifier and low pass filter



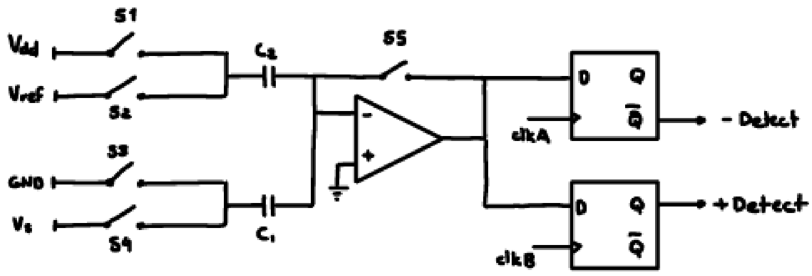
(a) Differential–input amplifier and low–pass filter

Adjustable gain high pass filter



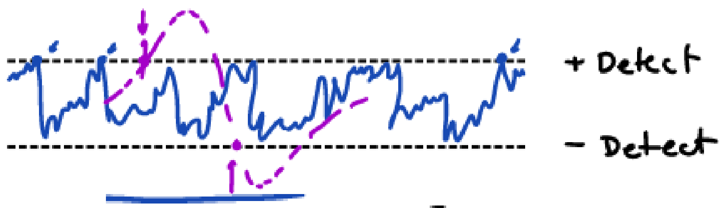
Comparator

The comparator has to be capable of detecting both the positive and the negative crossing,
This detection makes the comparator robust against noise



Operation

The operation of the comparator is divided in 4 phases



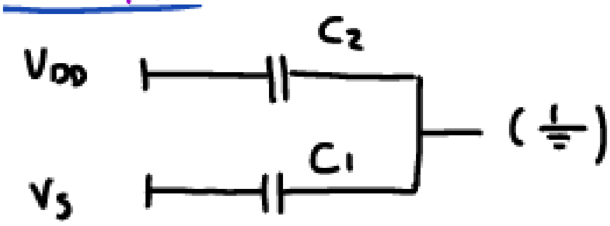
Phase 0

In this phase the switches

- 1,4,5 are on
- 2,3 are off

we have that

- The capacitance C_2 charges up to V_{DD} storing a total charge of

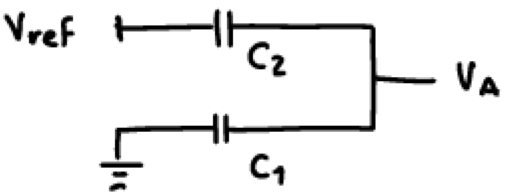


$$q_2 = V_{DD} \cdot C_2$$

- The capacitance C_1 charges to V_S, however since it is connected in the opposite way with respect to C_2 we consider the voltage and thus the charge negative

$$q_1 = -V_S C_1$$

Phase 1



In this phase we have broken the feedback so the 2 capacitances are now connected in series and thanks to the switches 2 and 3 the configuration is the one on the left.

The total amount of charge remains the same but it will redistribute to make so that the total sum of the voltages across the capacitances is V_{ref} we thus have

$$V_{ref} = V_{C1} + V_{C2}$$

So the new charges at equilibrium must be

$$q_2' = (V_{ref} - V_A)C_2$$
$$q_1' = V_A C_1$$

We can use the charge balance to find V_A

$$q_1 + q_2 = q_1' + q_2'$$

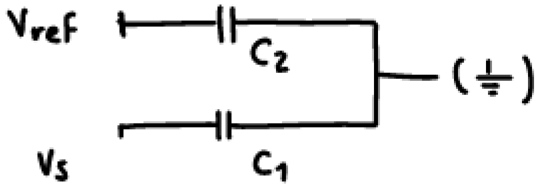
We obtain

$$V_A = \frac{C_2(V_{ref} - V_A)}{C_1 + C_2} - C_1 \frac{V_S}{C_1 + C_2}$$

We can conclude that the voltage will be positive only if the signal voltage will be lower than the negative threshold

$$V_S < -\frac{C_2(V_{dd} - V_{ref})}{C_1}$$

Phase 2



analog to phase 0 but with the capacitor C_2 being charged to V_{ref} rather than V_{DD}

Phase 3

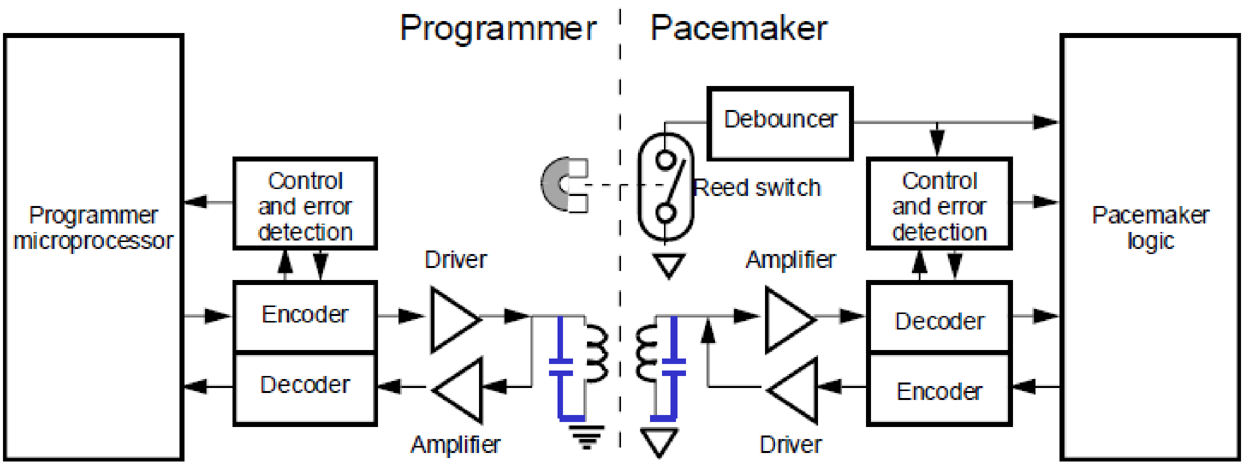
Analog 2 phase 1 but with the voltage now being V_{DD}, the result is that the term in the charge of C_2 is the opposite of before

$$q_2' = (V_{DD} - V_A)C_2$$
$$q_1' = V_A C_1$$

This way we obtain that $V_A < 0$ only if the signal is past the positive threshold

$$V_S > \frac{C_2(V_{dd} - V_{ref})}{C_1}$$

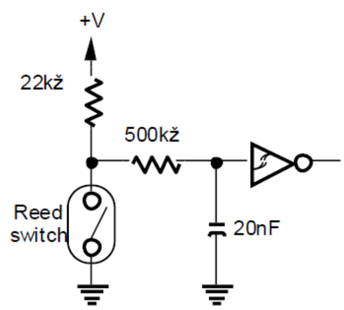
Telemetry for external programming



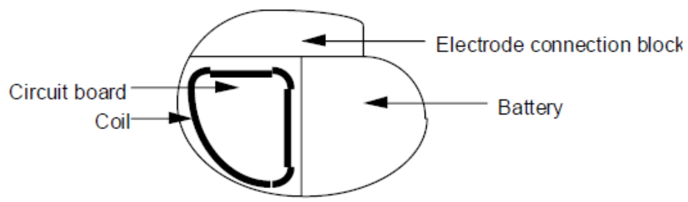
Once the pacemaker is installed we need a way to be able to program it to adjusts its parameters wirelessly.

Magnetic switch with de bouncing circuit

The first thing we need is a way to transition the pacemaker into a programmable mode, to do this we utilize a magnetic switch which will be turned on during the programming phase.



In this component a coil is used to drive switch, note that we have a low pass filter ad a Schmitt trigger to prevent any bouncing of the switch to transfer further into the pacemaker



Then we need a RC circuit to transfer the data packets to program the pacemaker, this coil will take up much space inside the pacemaker as we can see almost half of the pacemaker is occupied by the battery and the coil takes up a lot of the remaining space.

Data packets

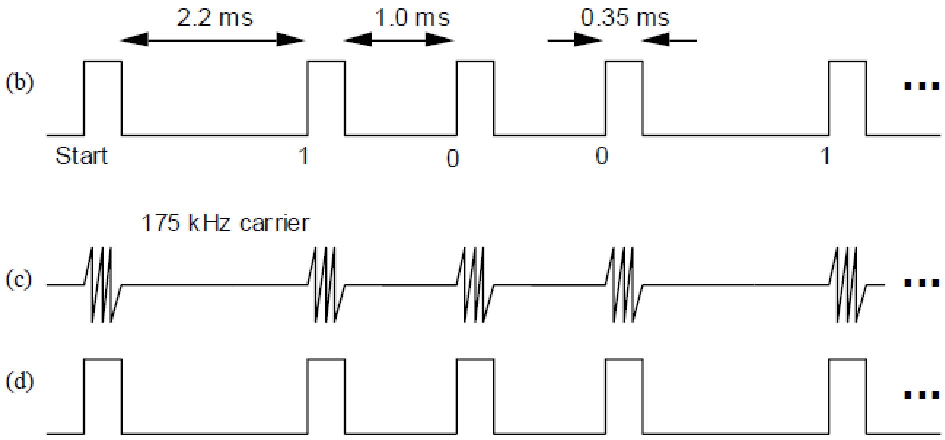
(a)

for timing	pulse rate	80 beats/min	code	errors detector
Start bit 1 bit 1	Parameter no. 8 bits 10010000 MSB LSB	Parameter value 8 bits 00101100 MSB LSB	Access code 8 bits 10010111 MSB LSB	Parity 8 bits 10111100 MSB LSB

The data needed to program the circuit are divided into packets and are transferred using different protocols which determines the packet subdivision and how the transmission happens.

Data transmission

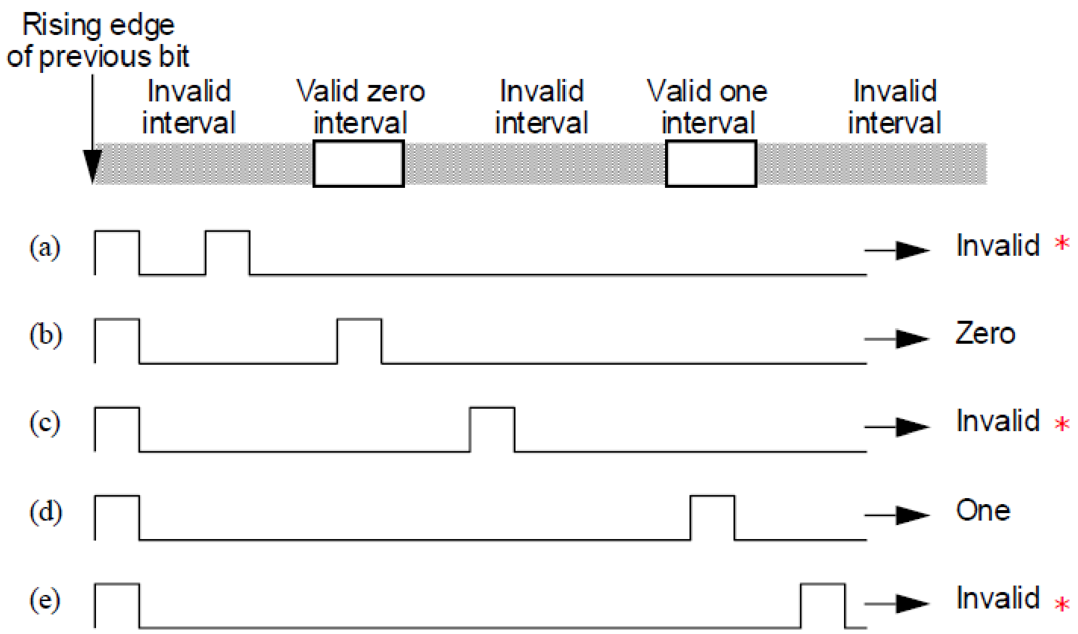
Data is coded not in the signal amplitude but in the timing of the high fluctuations. We transfer the signal using an oscillator which is turned on or off generating peaks of signal.



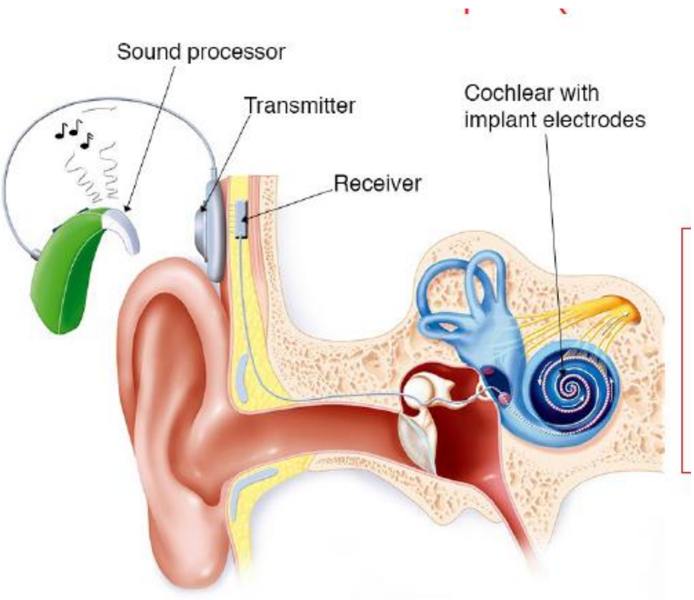
After the first peak is perceived we have that

- If the carrier wave is high after $2{,}2ms$ than we are receiving a high signal
- If the carrier is high after $1ms$ then we are receiving a zero

This solution is much more resistant to noise since the pacemaker will need to consider valid only the inputs coming during the time windows of $1ms$ and $2{,}2ms$ while al other peaks which may be caused by interferences can be discarded.



Cochlear implant



The cochlear implant is a prosthetic device implanted in the inner ear with the goal to restore partial hearing by stimulating directly the auditory nerve.

The ear

The ear has a frequency range between $20Hz$ and $20kHz$.

Operation

- The external ear picks up acoustic pressure.
- The middle ear converts them to mechanical vibration by a series off small bones
- The inner ear (cochlea) transforms the mechanical vibrations into fluid vibrations
- The vibrations cause the displacement of a flexible membrane **basilar membrane**, these variations cause some hair cells to bend and release an electrochemical substance that causes neurons to fire

Frequency encoding

We have that different frequency cause maximum vibration at different points along the basilar membrane

- Low frequencies at the apex
- High frequencies at the base

This difference is what allows the brain to differentiate between different frequencies.

To recreate the same effect we need to utilize an electrode array to stimulate the nerve fibers in different places, additionally each frequency range should be transferred with the appropriate gain to each electrode, so a high frequency input will be transferred with a low gain to the first part of the electrode and with a high gain to the final part.

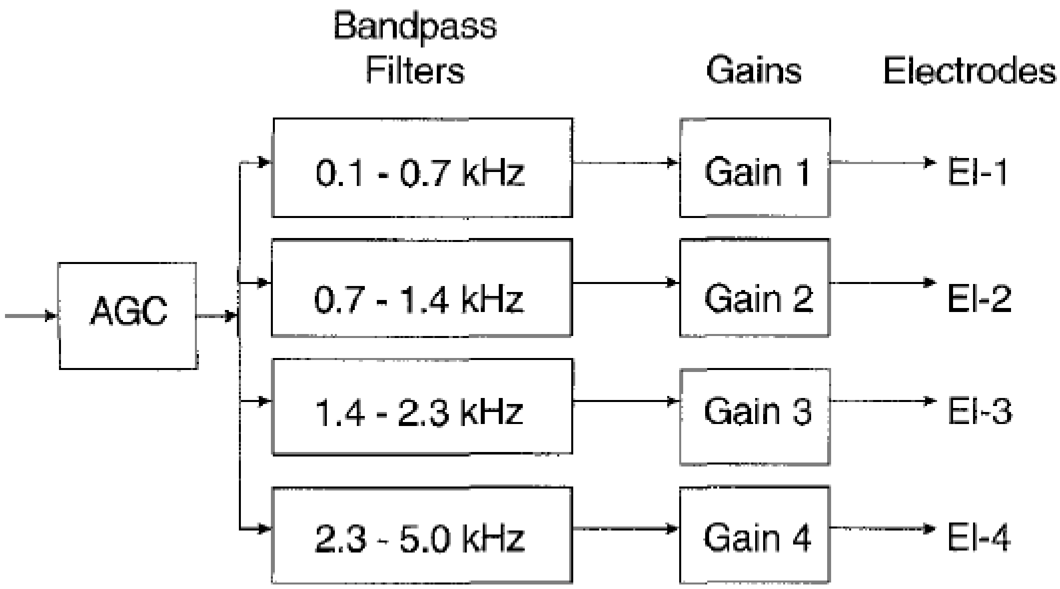
Compression of acoustical signal amplitudes

The amplitude range of a microphone can not be translated 1 to 1 into current signals to be transmitted to the nerves, we are going to need to apply a compression.
This compression can be applied in different phases during the signal elaboration chain, with different trade offs

- Applying the compression as late as possible allows us to neglect the effect of electrical noise added by the circuit
- Applying the compression too late however may cause the saturation of some of the electronics thus compromising correct operation

Typically we apply the modulation in 2 stages at the input to avoid saturation and then finally at the output with a logarithmic modulation so that we are not affected too much by the noise

Compressed analog CA

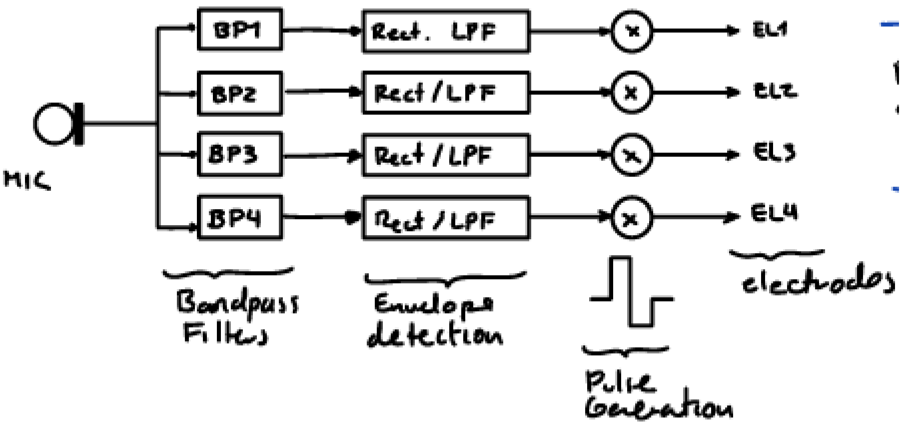


In this implementation the analog signals picked up by the microphone are elaborated

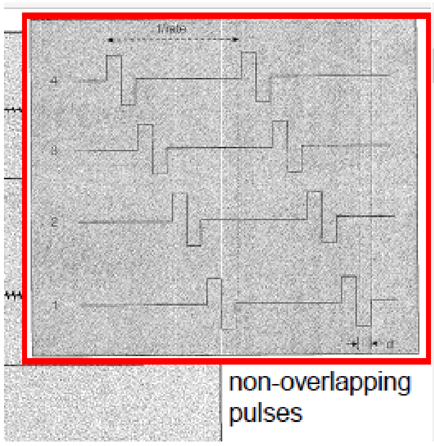
- We have first a general attenuation.
- Then a series of band pass filters to separate the frequencies
- Then an amplifier for each specific frequency to correctly assign its value for the corresponding electron

Then the signals are used to drive directly the electrodes

Continuous interleaved sampling CIS



- In this solution the signals are rectified and converted into a train of digital impulses. The digital signal is then sent to the electrode.

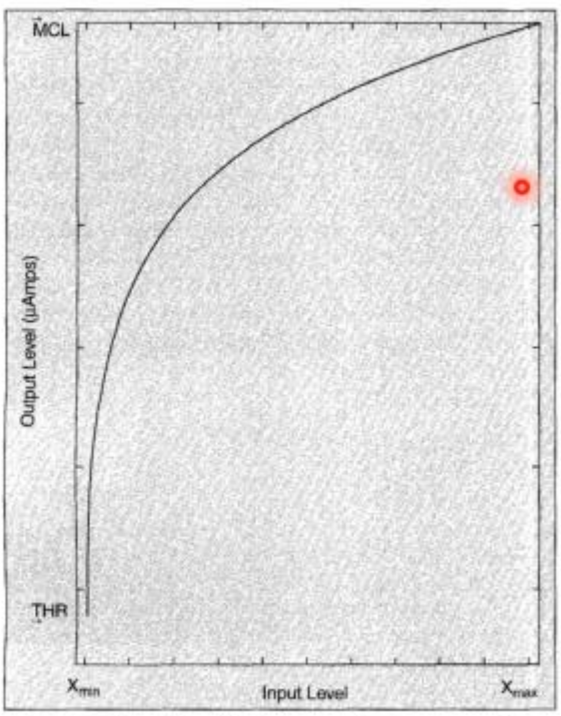


It is important to note that the transmission is not simultaneous over the channels but instead the signals are sent one after the other to avoid interferences between the electrodes because of overlapping pulses.

Gain compression and non linear mapping

We do not utilize a single gain because this would either make the circuit saturate in the case of signal with high amplitude or the value to be too small in case of small signals.
The resulting circuit will present a variable gain depending on the amplitude of the signals so that smaller gain will have a bigger gain while bigger signals will have a smaller gain.

Logarithmic compression



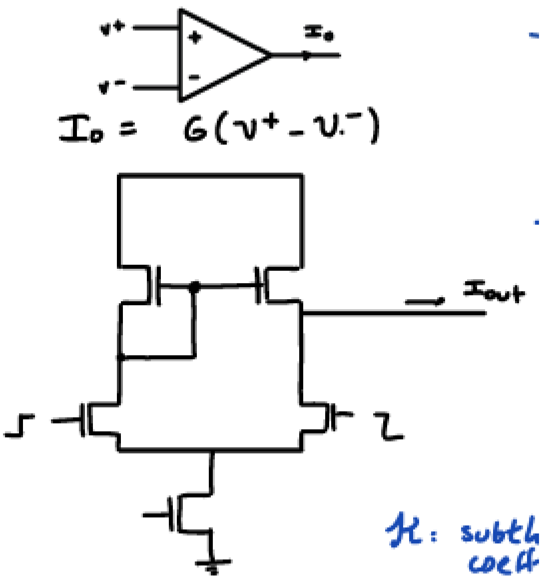
On the horizontal we have the range of the sound while on the vertical axis we have the output current.

As we can see the gain is progressively reduced according to a logarithmic curve.

This logarithmic compression will be obtained utilizing a diode.

Basic electronic blocks

OTA operational transconductance amplifier

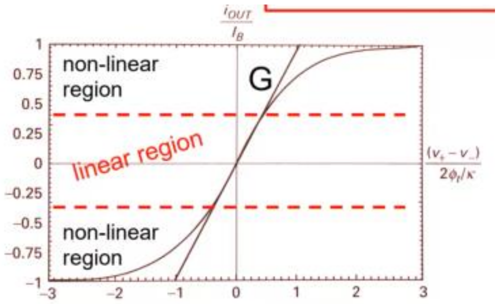


This circuit is based on a differential couple of transistor biased with a current mirror.
When a differential signal is applied to the inputs the bias current of the transistor changes, this difference is then copied by the current mirror so that at the output we have a current proportional to the input signal and amplified by a gain dependent on the transistor transconductance if the **transistor operate in saturation**

$$\frac{I_O}{V_{in}} = G = g_m = 2K(V_{GS} - V_T)$$

Overall the transfer function is an hyperbolic tangent as the current saturates at the bias current value.

$$i_{OUT} = I_B \tanh\left(\frac{V^+ - V^-}{\frac{2}{k}\frac{kT}{q}}\right)$$



Note that

- $\frac{kT}{q}$ is the thermal voltage
- k is called the subthreshold coefficient, this is required because the transistor may be operating below threshold

We are in the linear region if

$$|V^+ - V^-| < \frac{I_B}{\frac{2kT}{qk}}$$

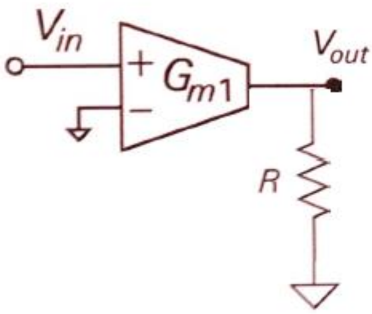
Then we can approximate and write

$$G = \frac{I_B}{\frac{2kT}{qk}} = \frac{I_B}{V_L}$$

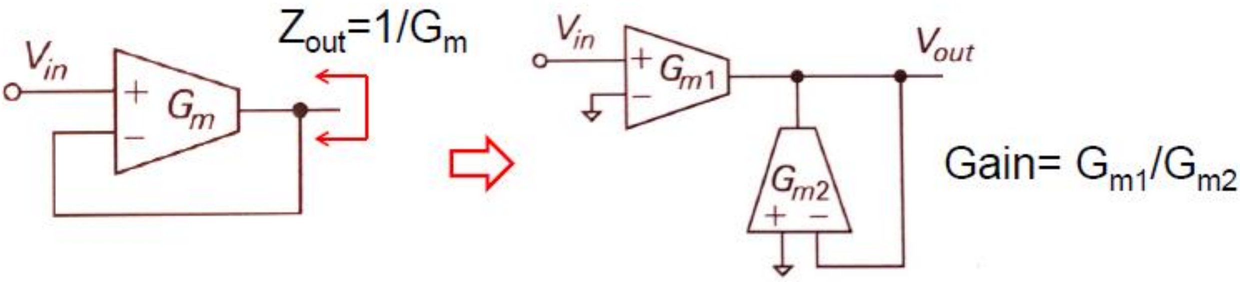
$G_m - R$ amplifiers

We place a resistance at the output of the OTA giving us an output voltage gain equal to

$$G = G_{m1} \cdot R$$



zRather than using a resistance we utilize another OTA connected with a negative feedback in a buffer configuration, this vies us an output impedance $Z_{out} = \frac{1}{G_m}$



$G_m - C$ filter

This configuration gives us a ow pass filter.

Fast study

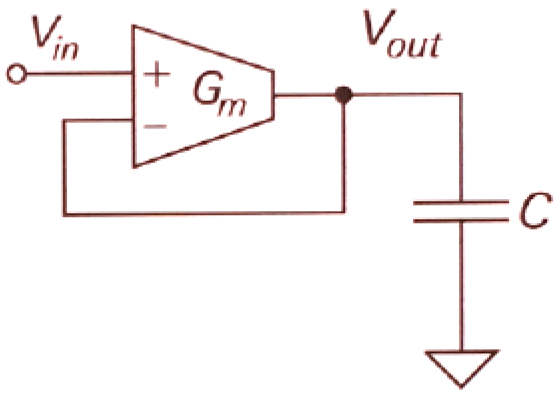
Because of the feedback we must have

$$G_m(V_{in} - V_{out}) = sCV_{out}$$

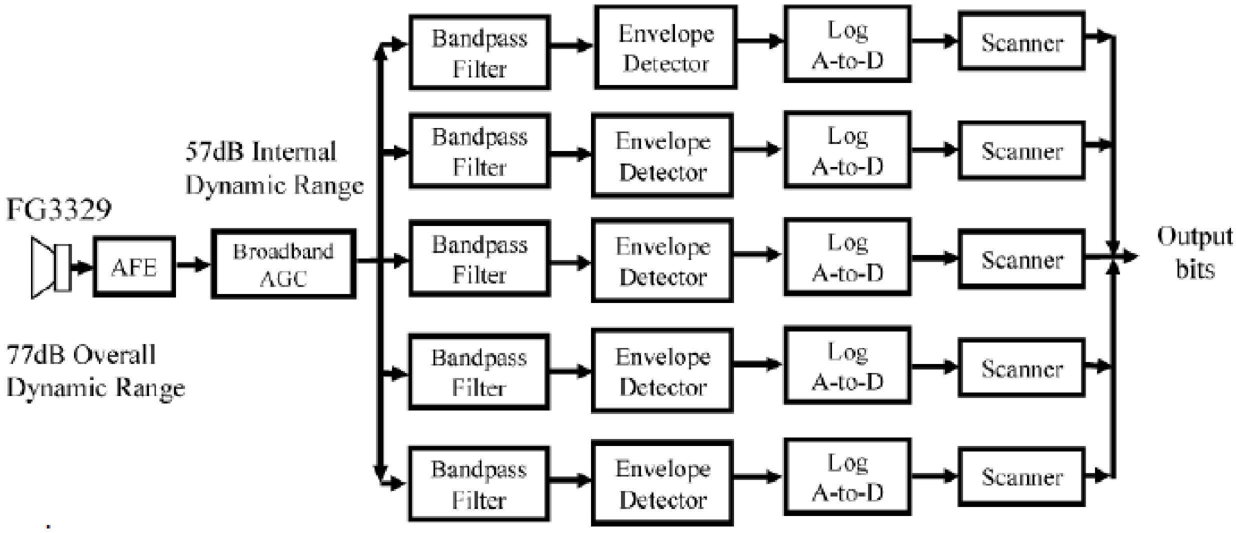
So we can solve

$$\frac{V_{Out}}{V_{in}} = G_m \frac{1}{1 + s\frac{C}{G_m}}$$

The time constant is $\tau = \frac{C}{G_m}$



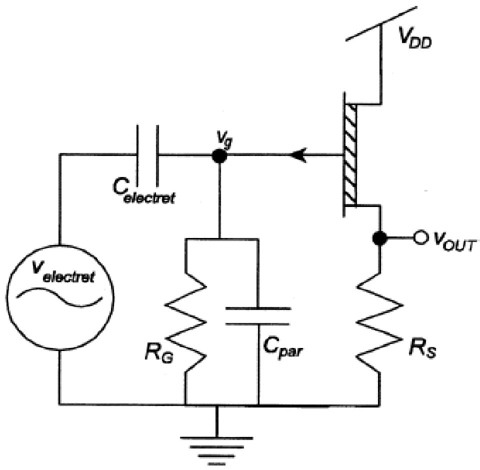
Example of CIS cochlear implant processor



Let's now study all the components

Microphone preamplifier

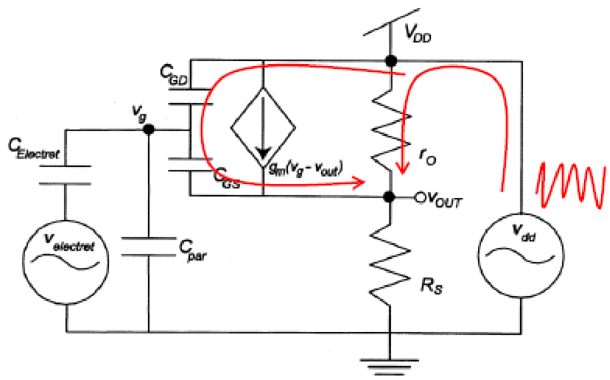
The circuit an electret microphone (a condenser) with a permanent charge built into it, sound wave oscillates the capacitors plate changing the voltage across it and generating the voltage signal.
A JFET in source follower configuration to act as a buffer is used to provide a low impedance voltage output.



We can see that a resistor R_G,is utilized to bias the gate of the JFET providing the current and additionally setting the voltage bias for the gate (ideally if the resistor isn't too large and the current is small the bias point will be at zero), note that the resistor cannot be too small otherwise we kill the signal transfer.

We are going to have noise because we bias the JFET utilizing V_{DD}, the general power supply will be noisy as a consequence of the interaction with all other components in the system.
This noise will transfer to the output through

- the output resistance
- the gate source and gate drain capacitances of the transistor



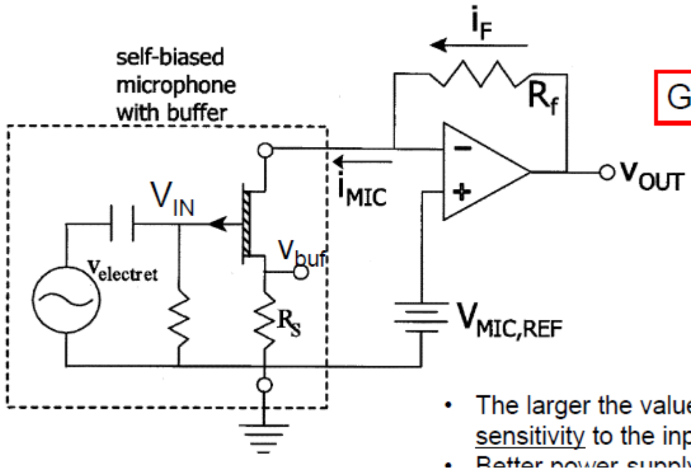
Drain current readout by sense amplifier

This is an alternative front end, we read out the charge in the drain current by means of a sense amplifier.

This way we can include a gain the larger the value of R_F the larger the gain

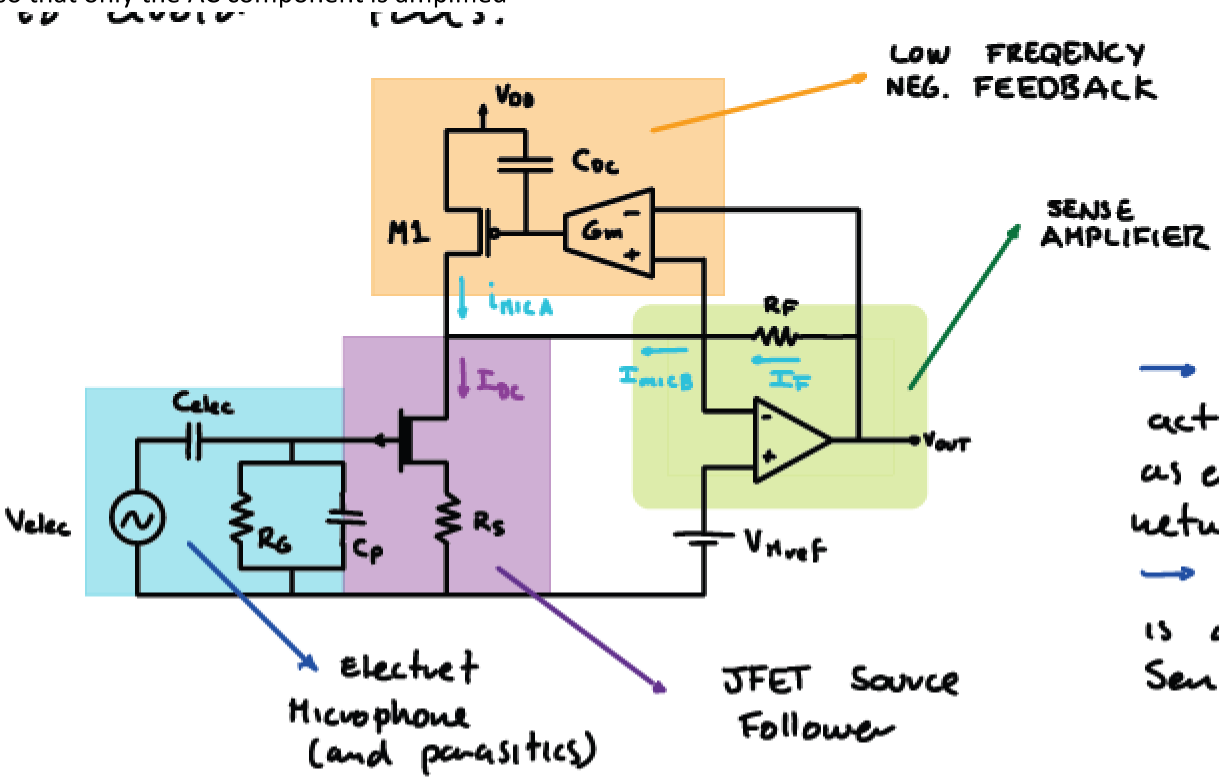
$$G = \frac{R_F}{R_S}$$

This solution giver better bias supply rejection because we bias the drain terminal through a reference voltage which can be filtered and because we do not draw current from it, rather then from the main V_{DD}.



Amplifier saturation

A large R_F is derisible to achieve a rage gain and a low thermal noise of the resistor, this however would cause the amplifier to saturate because of the simple bias current of the JFET.
To avoid this we add a low frequency negative feedback which subtract from the input the DC bias current so that only the AC component is amplified



The feedback is only active as low frequency tanks to the $G_m - C$ low pass filter, this way M1 provides only current at low frequency while the high frequency components need to flow through R_F and thus will be amplified.

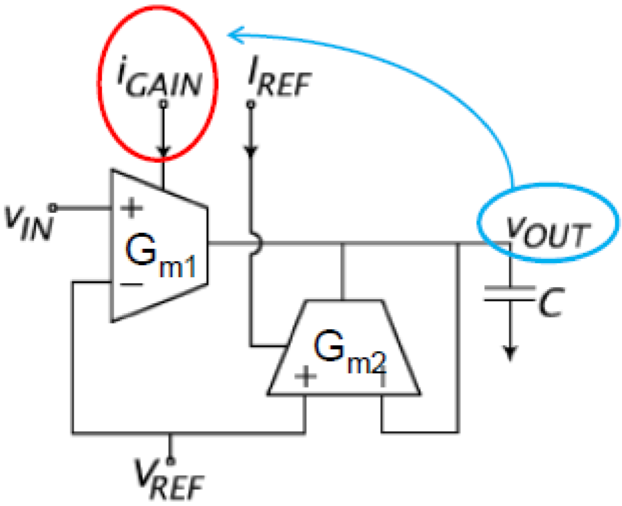
Automatic gain control AGC

The role of the AGC cell is to compress the $77dB$ input dynamic range into the $57dB$ internal dynamic range, we do this by varying the gain so that soft sounds are amplified by a large gain while lord sounds are amplified weakly.

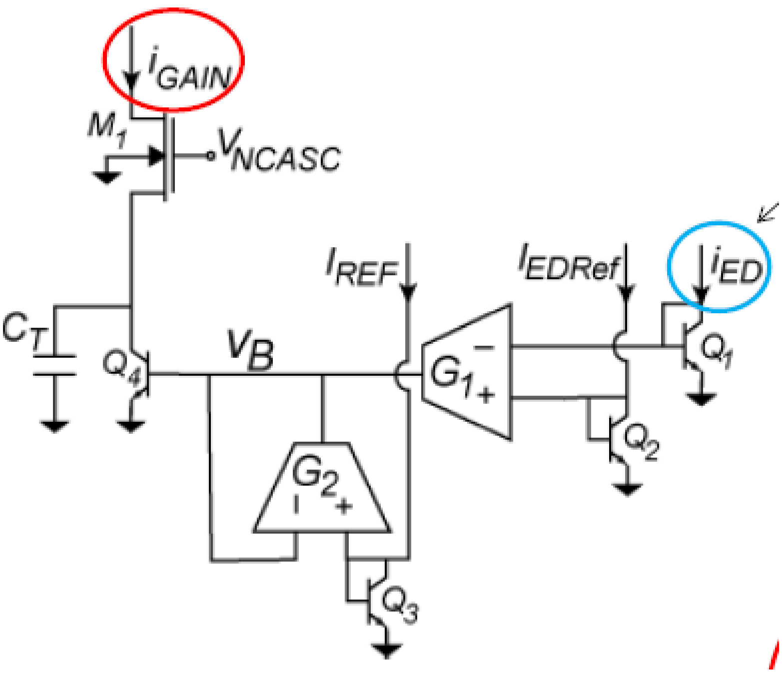
Operation principle

AGC senses the output envelope of the amplifier and uses it to control the gain of the amplifier itself, to implement it we utilize a **variable gain amplifier** based on **regulating the transconductance G_{m1} of a $\frac{G_{m1}}{G_{m2}}$ configuration**.

In this configuration the OTA are biased to work in weak inversion so that the gain will depend on the bias currents



How to modulate I_{gain}



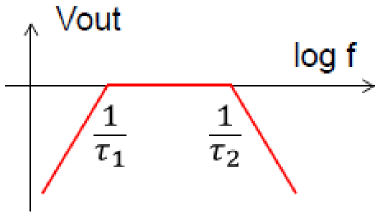
V_{out} is converted into a current I_{ED}, we then use an OTA to compare this current to a reference current $I_{ED,ref}$ utilizing an OTA in a $G_m - R$ configuration.

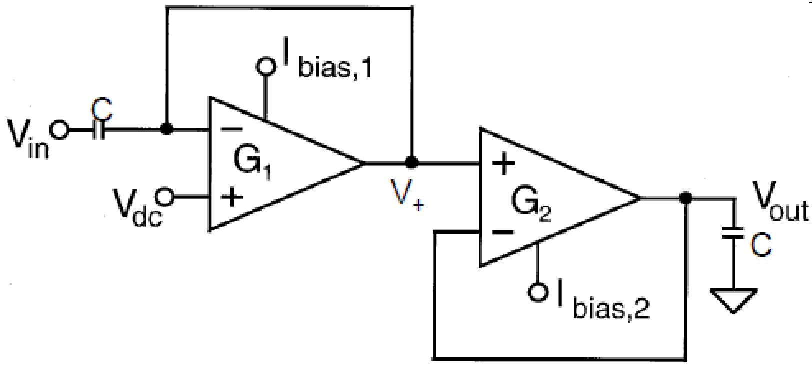
The resulting voltage V_B is used to drive a BJTT which in turn determines the, we have that the higher I_{ED} the lower the I_{gain}

$$I_{GAIN} = I_{ref}\left(\frac{I_{ED,ref}}{I_{ED}}\right)^{\frac{G_1}{G_2}}$$

The band pass filter

A band pass filter is created by a cascade of a high pass and a low pass filter





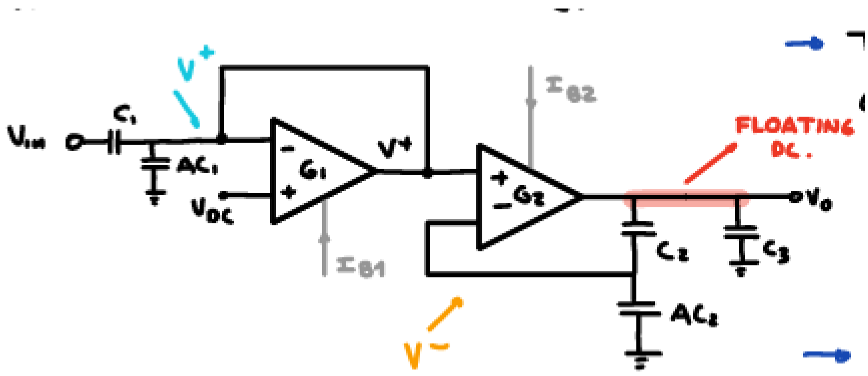
$$V_+ = V_{in}\frac{sC/G_1}{1+sC/G_1}$$

$$V_{out} = V_+\frac{1}{1+sC/G_2}$$

The major limitation of this design is its small liner range as the signal amplitude is limited to the range of the transconductors (OTA), remember that the signal is large enough to require a compression.

Increasing the linear range

Rather then feeding the signal directly in the filter we feed it through a capacitive partition, then in the feedback of the low pass stage we add instead a partition which increases the gain by the same factor for which we attenuated it before so the overall gain is unchanged.



Analysis

At the input the partition is

$$V^+ = V_{in}\cdot\frac{C_1}{C_1+AC_1} = \frac{V_{IN}}{1+A}$$

We can se that we have an attenuation of $1+A$.

when we consider the second stage we can just consider that because of the feedback we must have

$$V^- = V^+ = V_{OUT}\frac{C_2}{C_2+AC_2} = \frac{V_{OUT}}{1+A}$$

We invert to obtain the transfer function and we get

$$V_{OUT} = V^+(1+A)$$

And since $V^+ = \frac{V_{IN}}{1+A}$ we have that the gain remains unchanged

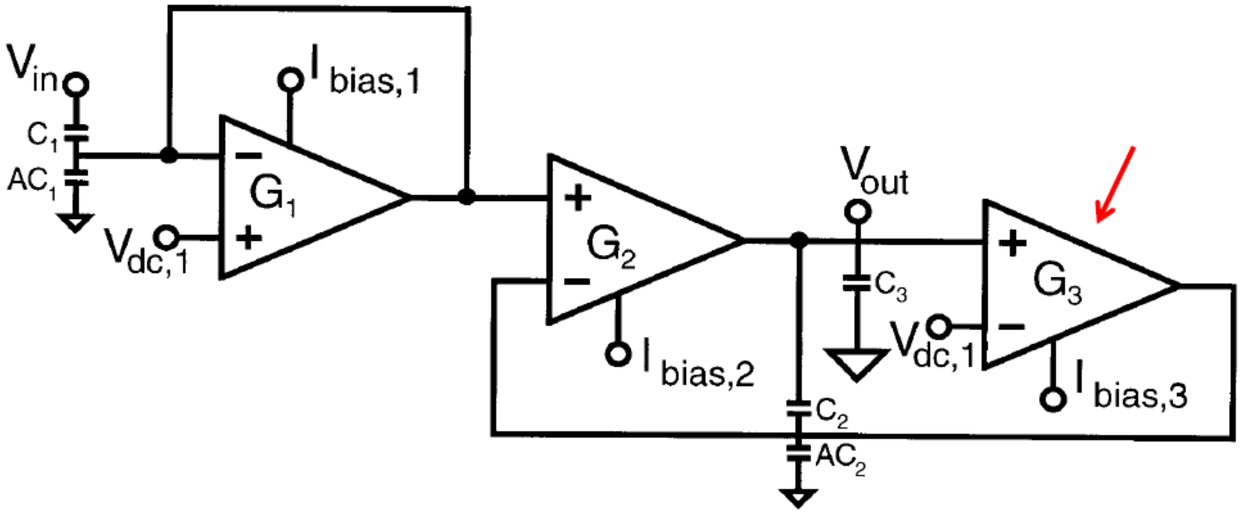
Floating DC problem

With this solution the bias point of the output state is completely decoupled from the power supply or ground.

This means hat the DC bias point is not stable

Band pass filter with DC stabilization

We add a third stage which provides a low frequency path between V_{out} and G_2, fixing the DC operating point of the circuit to the DC value $V_{DC,1}$.
Note that the loop gain of the circuit is small so the impedance seen at the input is not zero
Additionally, a capacitance C_3 is added in parallel with the attenuating capacitances to increase the filter capacitance and thus moving to lower frequency the pole.

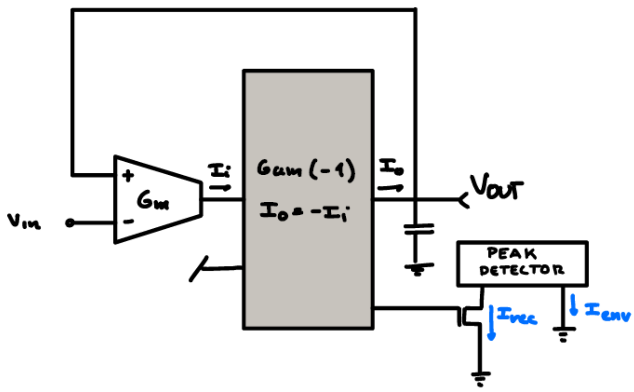


Note that typically this structure is replicated 2 times to that we can have a more selective band pass filter.

The envelope detector

We want to rectify the signal before transmitting it to the ADC.
To do this we utilize a device is based on a $G_m - C$, which however includes an intermediate block which inverts the output current of the OTA.

$$I_{OUT} = -I_{in}$$



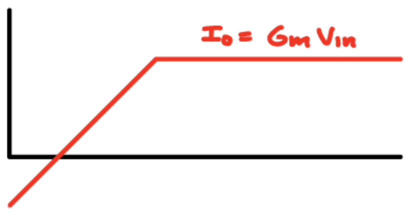
The transfer function is

$$V_{out} = V_{in} \cdot \frac{1}{1 + s\dfrac{C}{G_m}}$$

And the output current can be found by simply dividing the output voltage by the output capacitor impedance

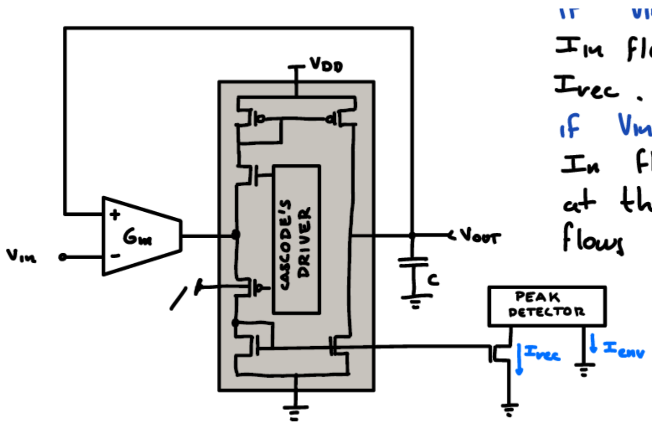
$$I_{out} = sC \cdot V_{in} \cdot \frac{1}{1 + s\dfrac{C}{G_m}}$$

The high pas filter effect is utilized to reject DC noise.



The intermediate block will replicate I_{out} rectifying it making so that $I_{rec} = |I_{out}|$

Intermediate stage (half wave rectifier)

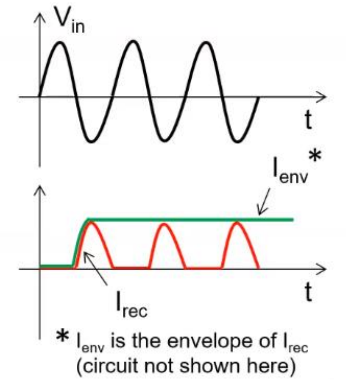


The stage is composed by

- 2 current mirrors
- A cascade configuration (current buffer)

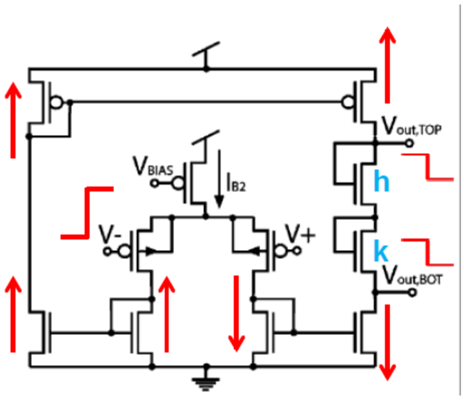
We can see that

- If $V_{IN} < 0$ than $I_{in} > 0$
  The additional current will flow in the bottom mirror while the upper mirror is going o be off. The current of the bottom mirror is copied and transferred in I_{rec}
- If $V_{IN} > 0$ than $I_{IN} < 0$
  The mirror at the bottom is OFF and thus no current flows in I_{rec}



We can obtain a full wave rectifying by sending the pMOS current into an nMOS mirror and add this current to I_{rec}

Cascodes driver



The drivers must bias the gates and switch on or off the transistors depending on the signals.
To do this we utilize the following circuit, where we use a differential pair to receive the input voltage and compare it to the reference voltage.
The difference between the 2 creates a differential current changing the bias of the output transistors and thus generating a voltage.

As we can see we have 2 outputs separated by transistor in transdiodes configuration which allow us to have a difference between the 2 voltages and thus bias the cascaded gats

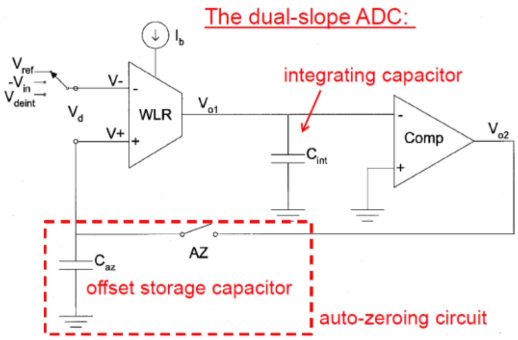
Logarithmic ADC

Now that we have the rectified signal we need to translate it into a digital signal while also applying a logarithmic compression.

To obtain this result we utilize an ADC composed by

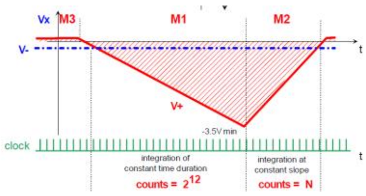
- A diode to obtain he logarithmic transfer function
- An OTA to perform a voltage to current conversion
- A dual slope ADC to convert the signal into digital data

Dual slope ADC



A dual slope ADC operates integrating a current in a capacitor, we utilize the OTA to convert the input voltage as well as the reference voltage into a current then

1. First we integrate the input current for a fixed amount of time the integration lasts for a fixed amount of time equal to $2^n T_{clock}$
2. Then we discharge the capacitor using the reference current



The time required to discharge the capacitor will be proportional to the ratio between the input voltage and the reference voltage, from it we can find the equivalent digital signal.

Computations

Total charge must be zero

$$Q_{in} = Q_{out}$$

$$I_{in} \times T_{int} = I_{ref} \times T_{measured}$$

$$T_{measured} = \frac{I_{in}}{I_{ref}} \times T_{int}$$

Autozeroing

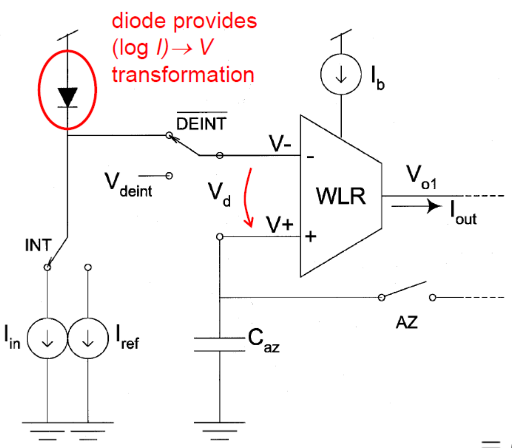
A real circuit will be subjected to offset variations in the OTA and the capacitor, the capacitor C_{az} has the objective to reduce these variations. Before measuring the loop is closed and the input are disconnected, in this condition C_{az} charges up with a value equal to the offset. Once the input are connected than they will no longer be referred to ground but to the voltage across C_{az} so the offset will be cancelled.

Important

- Since we are using the same physical components for the same operations we have that their value is simplified in the ratio and so the result is not affected by divergences from their nominal value.
- Since we integrate the signal for a fixed amount of time we reject all harmonics with a period that is a submultiple of the integration time

Adding a logarithmic conversion to the ADC

We know that a diode presents a logarithmic relationship between its current and the voltage across it, so what we can do is rather than utilizing directly the current generated by the input voltage and the reference voltage we feed it first through a diode and then pass the resulting voltage through another OTA to obtain the currents we will effectively integrate



Computation

The voltage across the diode is going to be

$$V_d = \frac{kT}{q} ln \frac{I_d}{I_s} = \phi_T \, ln \frac{I_d}{I_s}$$

During the first phase the capacitor is charged with a current

$$\phi_T \, ln \left(\frac{I_{in}}{I_S}\right)$$

During the second phase the capacitor discharges with a current

$$\phi_T \, ln \left(\frac{I_{ref}}{I_S}\right)$$

The differential voltage at the input will be

$$V_{diff} = \phi_T \, ln \left(\frac{I_{in}}{I_S}\right) - \phi_T \, ln \left(\frac{I_{ref}}{I_S}\right) = \phi_T \, ln \left(\frac{I_{in}}{I_{ref}}\right)$$

So the differential voltage across the OTA has a logarithmic dependance on I_{in}

Independent from temperature

The output current will be given by

$$I_{out} = G_m \cdot V_{Diff}$$

Where $G_m = \frac{I_{bias}}{V_L}$ so we get

$$I_{out} = \frac{I_{bias}}{V_L} \cdot \phi_T \, ln \left(\frac{I_{in}}{I_{ref}}\right)$$

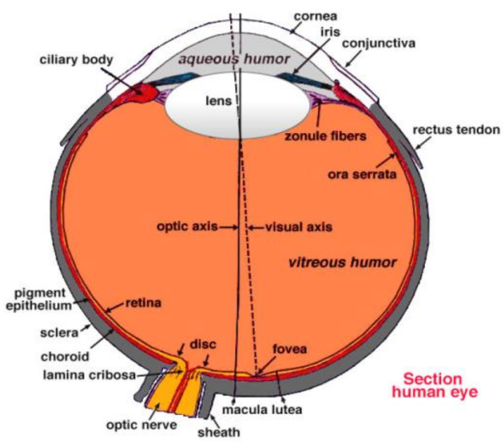
we need to remember that $V_L = \frac{2\phi_T}{k}$ so the term ϕ_T can simplify, meaning that **the final result will be independent form temperature variations**.

Electronics for artificial vision

Anatomy of the eye

The eye has semispherical shape and contains over 70% of the sensory receptors in the body.

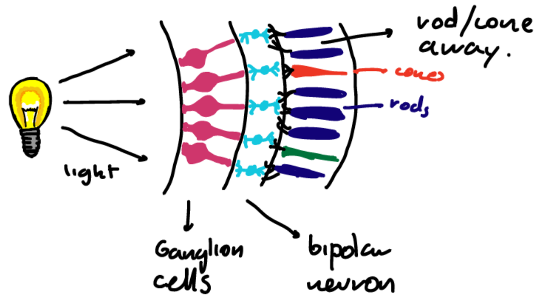
- The eye is filled up with fluid to help it retain its shape
- The cornea and pupil help focus and regulate incoming light
- Light is converted into electrical at the retina
- The optic nerve sends the data out to the brain



The retina

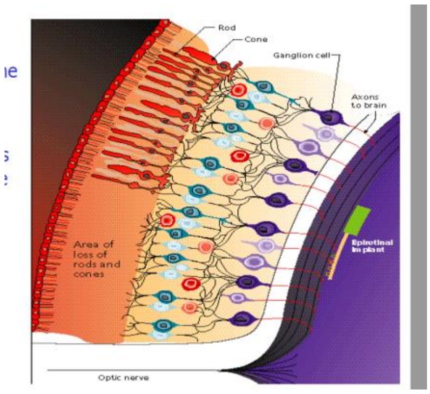
This is the back of the eye and where the light is converted into electrical signals

- **Cones** are all equipped with an individual **ganglion cells** which give highly detailed information about color.
- **Rods** are mor numerous but reveal no information on color only on light



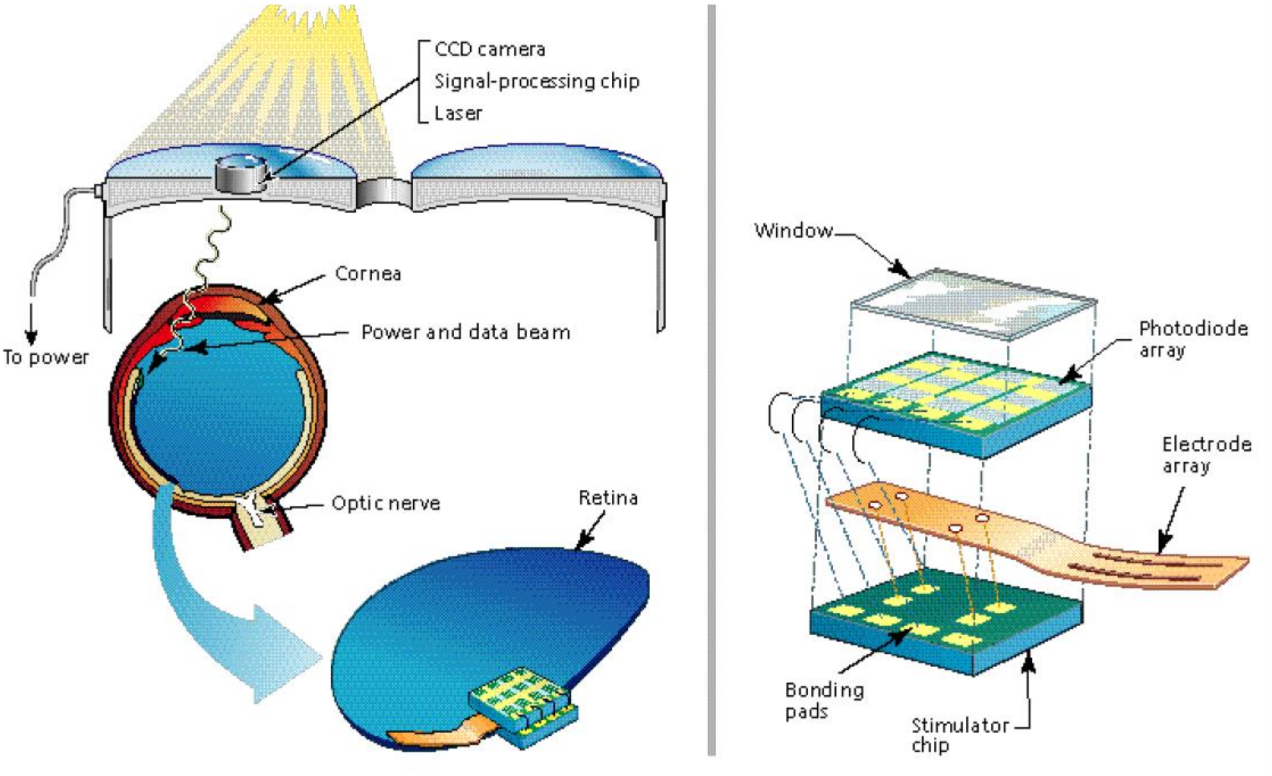
Epiretinal prothesis

This prothesis directly stimulates the ganglion cells, the drawback is that we lose some signal elaboration performed in the preceding layer so an external processing device is needed.



Operation

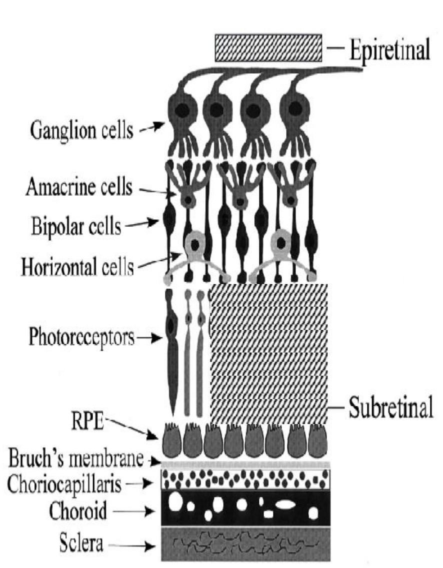
- A CCD camera receives the image
- The external video processor processes the image
- The signal is sent to the internal chip by either radiofrequency transmission or optical transmission
- The stimulator chip generates the current pulses and sends them to the microelectrodes



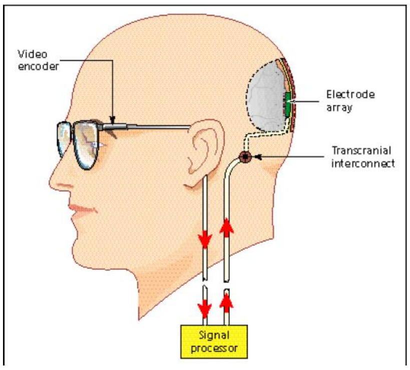
Subretinal prothesis

This retina substitute the layer of degenerated photoreceptors the pre processing role of the following cellular layers is maintained so the prothesis just needs a much simpler circuit: a photodiodes array which converts the incomping light into electrical pulses which are transmitted to the healthy cells through gold microelectrodes.

The application of this prosthesis is more complex as we need to apply it behind



Cortical prosthesis

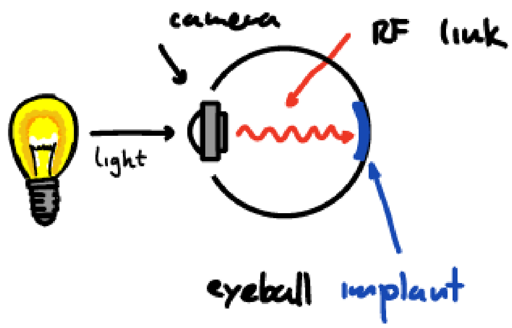


The prosthesis is made by an array of microelectrodes which is placed in contact with the cortical tissue.
A portable computer processes the images and correct them for the non-linearity of the retina cortex map.
The external unit is connected to the internal one by an RF system or by transcranial interconnection.
These prosthesis allow us to cope with pathologies which affect the optic nerves like glaucoma

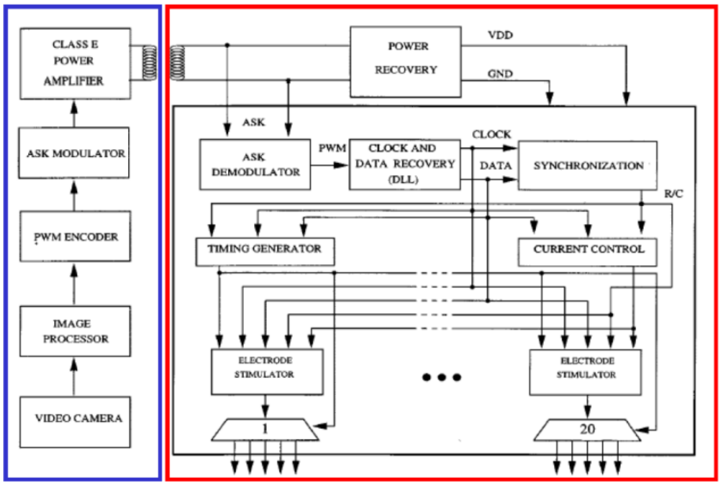
Multiple unit artificial retinal chipset MARC

This device is mad form 2 components

- An external part which acquires the images
- An internal part that stimulates the photoreceptors



Block diagram



IMPLANTED STIMULATOR

Each of these components is composed by several parts

External circuit

- Video camera and image processing hardware
- PWM data encoder
- ASK modulator
- Power amplifier and coils

Internal circuit

- Power recovery (secondary coil rectifier etc.)
- ASK demodulator
- Clock and data recovery
- Stimulus pulses generator
- Electrode array

ASK encoding

The modulator is an oscillator whose amplitude is determined by the value of the PWM so we have 2 different amplitudes depending on whether the PWM signal is high or low.
The signal is coded in the value of the duty cycle, we have that

- 0 is coded with a duty cycle of 50%
- 1 is coded with a duty cycle higher than 50%

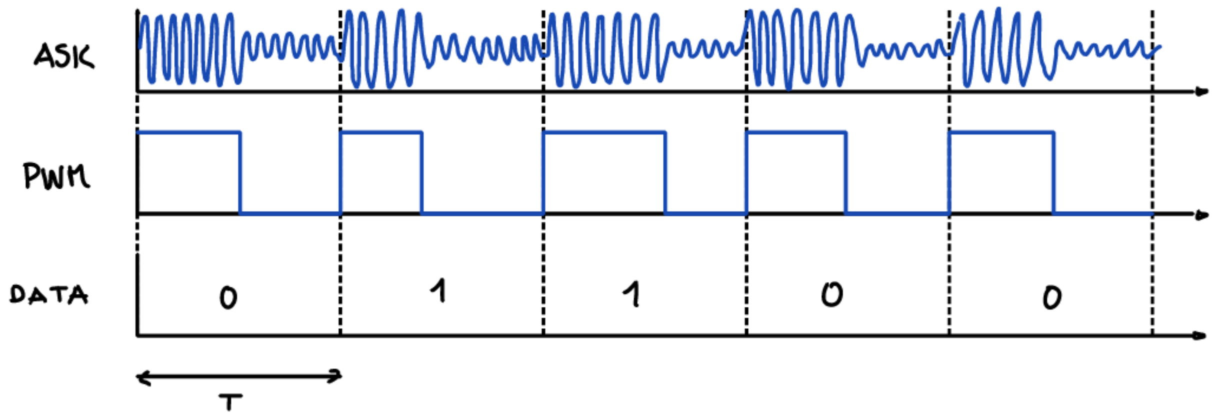
This solution has 2 advantages

Clock recovery

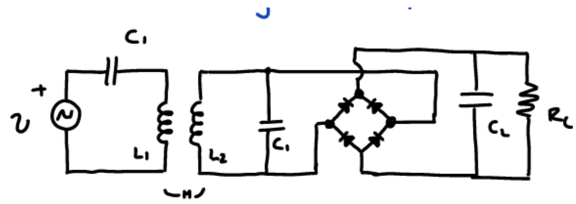
While the duty cycle change and thus the falling edge of the PWM happens at different moments the rising edge will happen always at the same time so we can recover the clock from the incoming signal.

Power transmission uniformity

The power to the internal circuit is also transmitted through the signal wave, since now even with zeroes we have a non-null duty cycle the variation in power transmitted with respect to the data is negligible, while if we were to directly modulate the carrier without the PWM protocol a series of zeroes would cause a much lower power transfer than a series of ones.

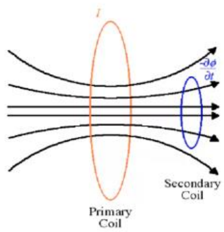


Wireless powering of implanted unit



Power to the internal circuit is provided by a RF link.
The internal antenna resonates at a frequency f_0 and absorbs power from the external antenna via mutual coupling.
Inside the circuit the AC signal is rectified to a DC voltage which is then used to drive the internal circuits represented above by the equivalent R_L, C_L

Coupling

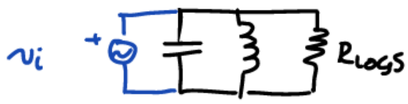


A current running in the primary coil creates electromagnetic fluxes, the secondary coil intercepts such lines, the amount of lines intercepted depends

- On the geometry of the coil
- On the distance between the coils

Generating the waveform

Basic LC resonator



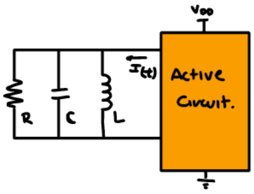
Without losses a LC resonator would resonate indefinitely at its resonance frequency

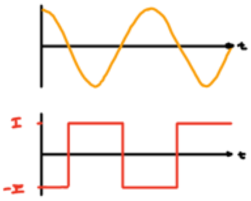
$$\omega_0 = \frac{1}{\sqrt{LC}}$$

However losses due to resistive elements cause the amplitude of the oscillations to decay.
Small values of R cause a faster decay as they draw more current.

Active circuit for loss compensation

To sustain oscillation despite the losses we need to compensate for the power lost in the resistor.



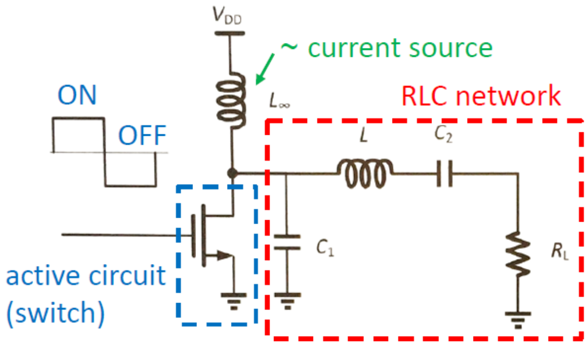


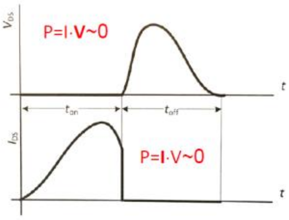
To do this we utilize an active circuit which will provide the current required by the resistor during each phase:

- When the voltage across the resistor is positive we need to inject current
- When the voltage across the resistor is negative we need to absorb current

Class E amplifier

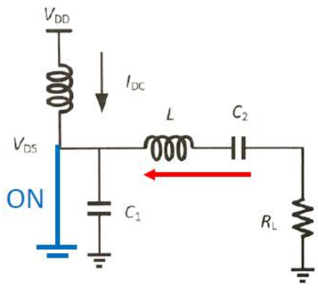
The circuit operates by first charging the inductor connected to V_{DD} during the on state and by then connecting it in series with eh RLC network so that it can provide energy to it.

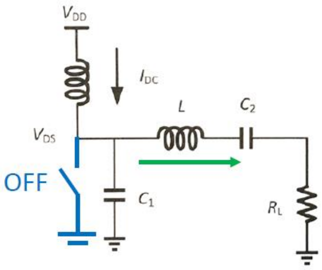




On state

During the on state both inductors are connected to ground, the current in the inductor L_∞ increases as more energy is stored in it.





Off state

During this state the current of the inductor is forced into the RLC circuit providing the energy that would otherwise be lost through the resistor.

Power dissipated

If we consider and ideal switch we have that

- During the on phase we have current flowing in the switch but no voltage across it
- During the off phase we have a voltage but no current

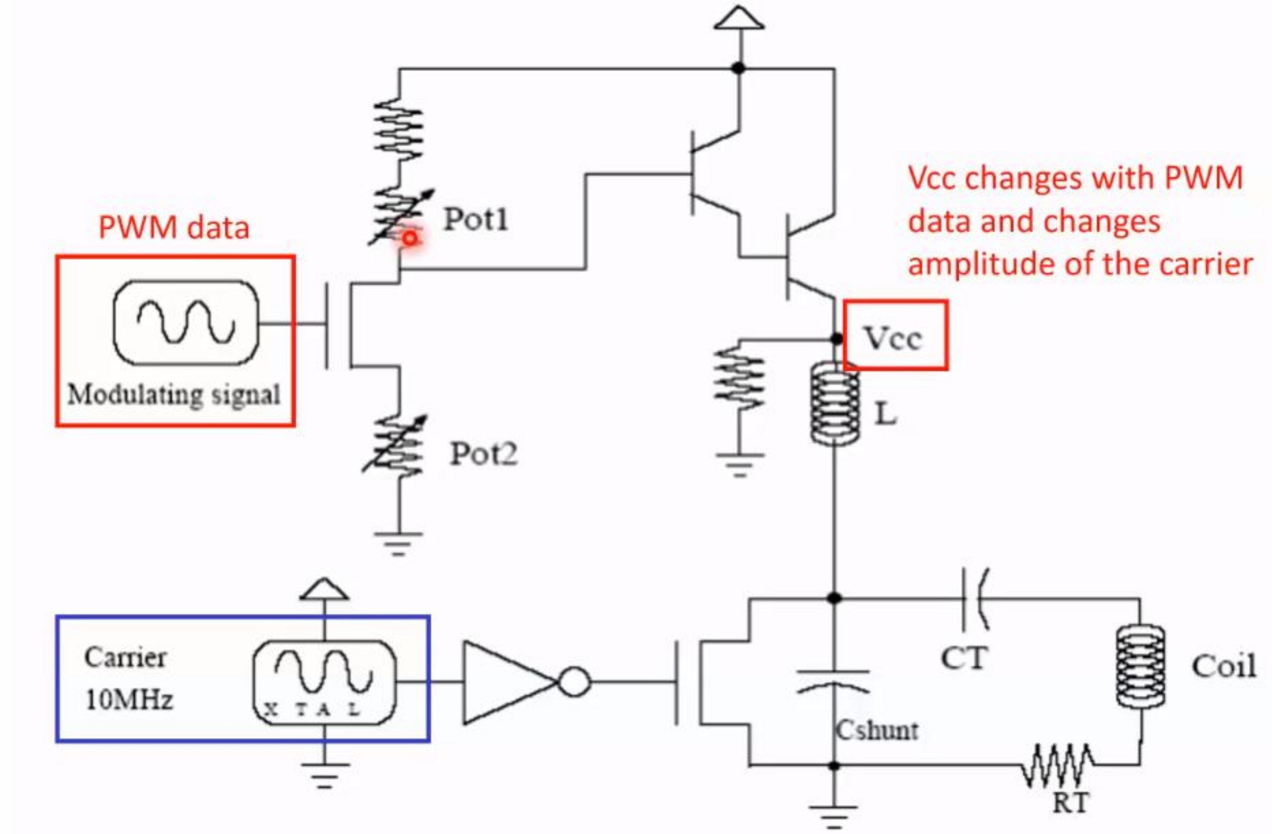
Since the power dissipated is equal to the product of voltage and current the total power dissipated with an ideal switch is zero.

Important

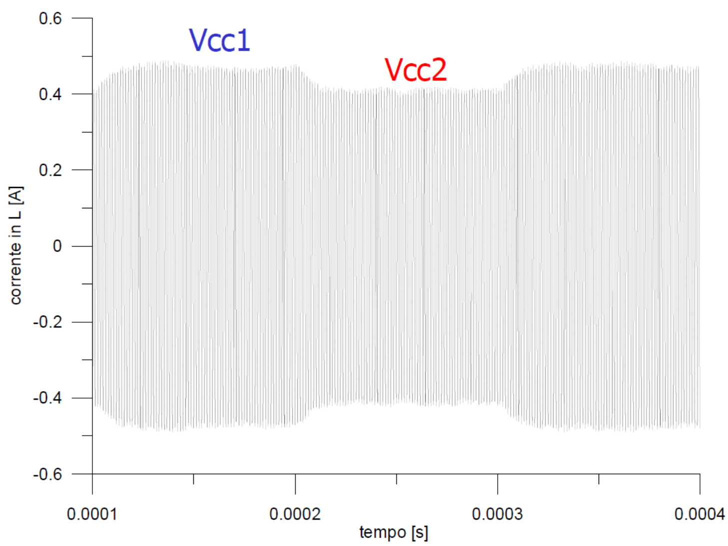
The gate of the switch must be driven with the same frequency as the resonant frequency of the RCL network.

Adding the PWM

In the MARC system we do not need only the oscillator but also we need to be able to modulate the waveform according to the PWM behavior. To do so we need to be able to change the value of the V_{DD} of the oscillator.



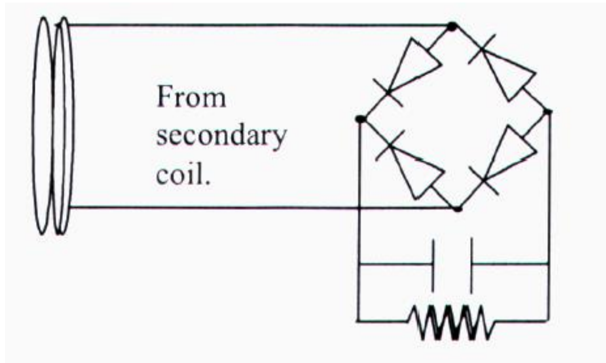
The resulting wave is the following



Envelope detector

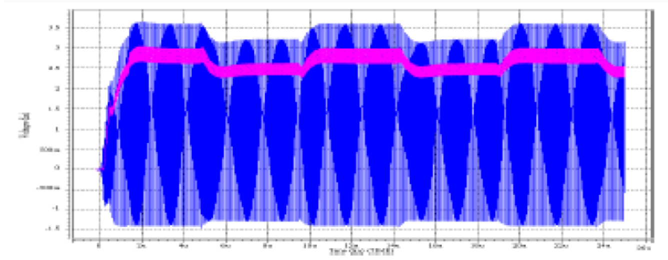
We now move inside the eye in the receiver, the first circuit we need to study is the envelope detector and which recovers the data transferred as well as obtaining the power transmitted.

We utilize a full bridge rectifier to rectify the current then an RC circuit is used to introduce a time constant and smooth out the signal.



Note

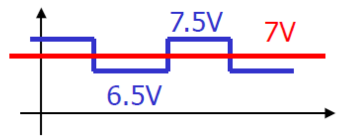
We need to balance the RC time constant as a slower time constant will gives a smoother power signal but will also make it harder to read the incoming data.

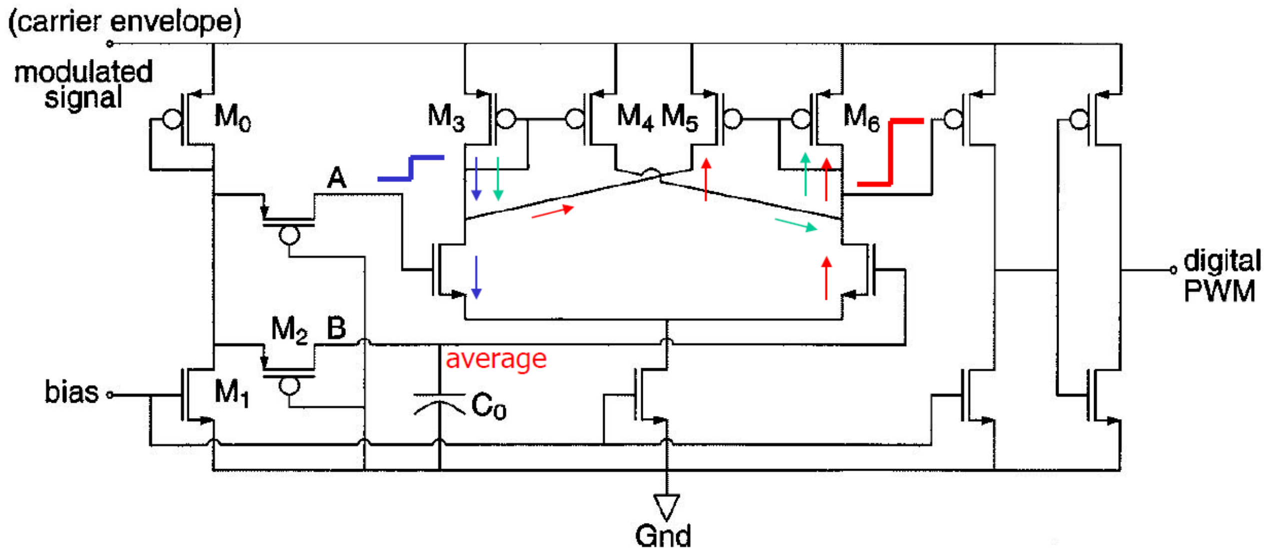


In blue we have the received signal while in pink we have the rectified signal.

ASK demodulator

We need to extract the signal from the rectified signal, the amplitude variation is very small so what we would like to do is transform the oscillation around the average value into a rail to rail signal, this is the function of the ASK demodulator.





This circuit is a comparator which compares the modulated signal with its average giving at the output a full swing signal.

Average computation

To compute the average, we have that the signal is integrated over the capacitance C_0 throughput the transistor M_2 which operates in ohmic regime.

Operation of the comparator

We can see that the differential stage has a criss-ross configuration this creates positive feedback in the configuration which will speed up the transients that go to the output.

1. M6 receives the current unbalance which is copied by M5
2. M3 provides the new current required form M5
3. M4 copies the current of M3 and delivers it back to M6

The same happens for the other side,

Output stage

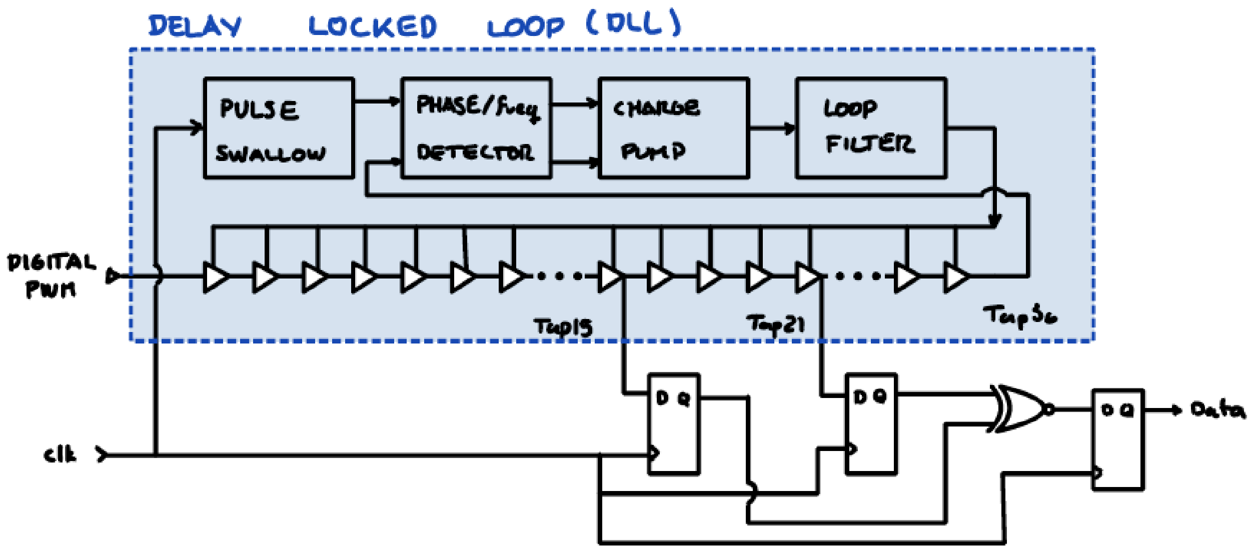
Note that the output might not still be rail to rail so we have an additional common source configuration and then finally an inverter to assure that the output is rail to rail.

Note on timing and power supply variation

The amplifier switches at the same time during which the power supply switches so we do not expect the circuit to fail because the circuit is not switching when it is doing an amplification we simply make comparisons and we have 2 possible power supply values.

Clock and data recovery unit

This circuit has to retrieve from the PWM the clock as well as the data.



Clock

The clock line is easy to retrieve as it matches with eh PWM rising edge which already has a constant period so we have a stable clock.

We just need to utilize components sensitive to the positive rise so we do not need to care about the falling edge.

Data extraction

To extract the data we need to determine the duty cycle. This is done by running the circuit into a series of digital buffers which will all introduce an equal delay.

In our example we have a chain of 36 buffers, we have that the delay after the 36th buffer is exactly equal to 1 clock period.

To do this we utilize a phase detector which takes as inputs

- The incoming pulse entering the chain
- The output pulse of the chain

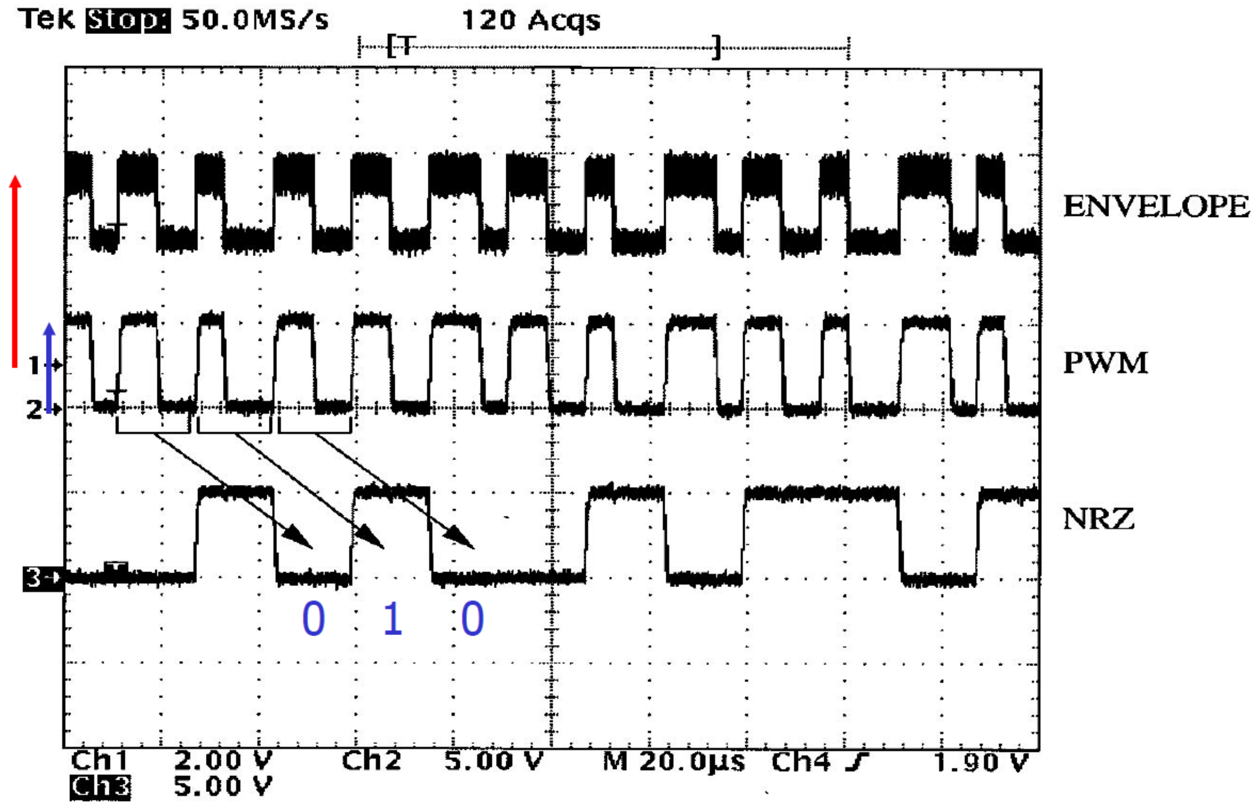
It checks the phase difference between the 2 and then through a charge pump changes the V_{DD} of the buffers so that their propagation delay can be adjusted.

Once the delay is exactly equal to 1 clock period the delay is locked **DLL delay locked loop**.

Since the chain include 36 buffers we have that each buffers stores the behavior of the pulse during each corresponding fraction of the period.
What we do is extract the values of buffer 15 and buffer 21

- If their value is 1 we have a duty cycle of 60% which correspond to an incoming 1
- If their value is 0 we have a duty cycle of 40% which correspond to an incoming 1
- If the value of 15 is 1 and the value of 21 is 0 we have a duty cycle of 50% which correspond to an incoming value of 0

We send the 2 outputs to a flip flop each to assure that we check the data only one time per clock cycle and only at the beginning of each new clock cycle.
The output of the flip flops and then XNOR port this way the output is 1 only if both outputs match otherwise, it is 0.

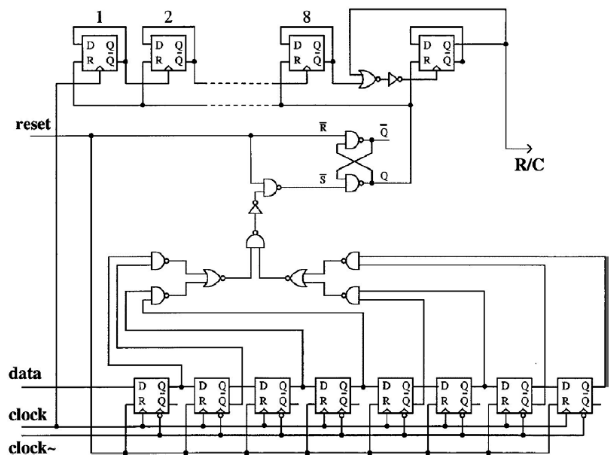


Synchronization circuit

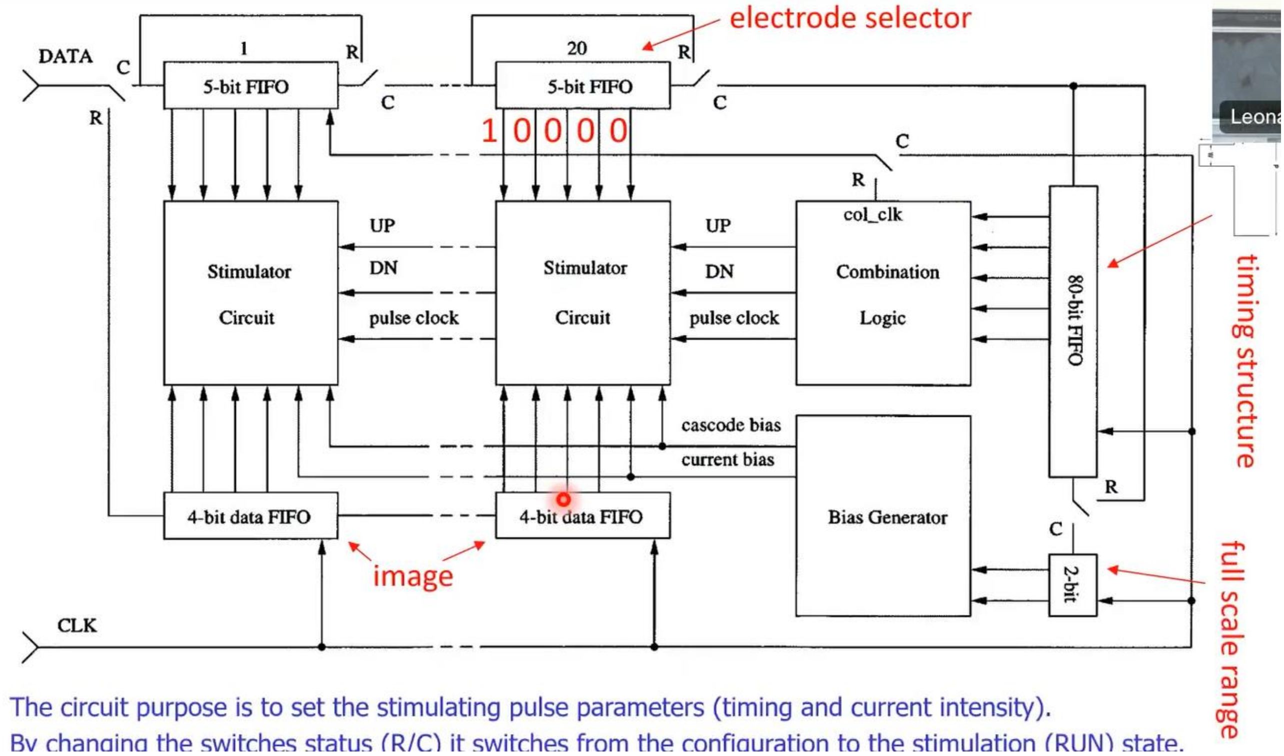
This circuit has 2 configurations

- One in which the incoming data are used to program the chip: configuration mode
- One in which the incoming data are used to run the chip: run mode

A specific key sequence is used to switch the chip from one configuration to the other, the synchronization circuit is what recognizes the key sequence and switches the ciruict

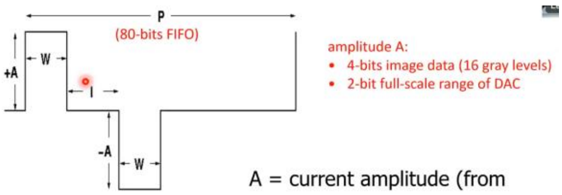


Pulse generation and definiton



The stimulator is the circuit is the key block in the device as it provides the pulse to trigger the action potential.

It is a biphasic pulse where only one lobe will stimulate the tissue, the other lobe is used to remove the charge injected with the first pulse so that there is no charge build up in the tissue.



Stimulator vs electordes

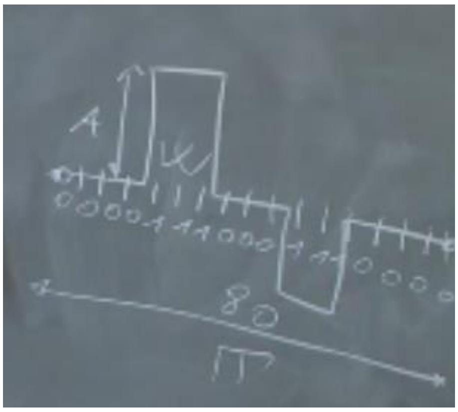
There is not one stimulator per electrode so the same stimulator is shared among 5 electrodes and it must be able to select to which to send the pulse.

Note that as a consequence we have that the electrodes are not stimulated at the same time but sequentially: this is not a problem because the speed of the stimulation is much higher than the one our eye can distinguish.

We have a 5 bits FIFO memory to select the electrode only 1 bit contains a 1 while the other contain zeroes, the stimulator will stimulator will act on the electrode corresponding to the 1.

The FIFO is a circular register so when we need to go to the next electrode we simply shift all data by 1 position.

Pulse timing



Each stimulator is also driven by and 80 bits FIFO which contains the time structure of the pulse (not the amplitude which will be indicated by the data) we need to know the width of the pulse

This FIFO il also closed in a loop because when we upload the time structure we continue to run it periodically for each period.

Important

We have a logic between the FIFO and the stimulator which digest the output of the FIFO and then gives to the stimulator just the commands to start and end the pulse.

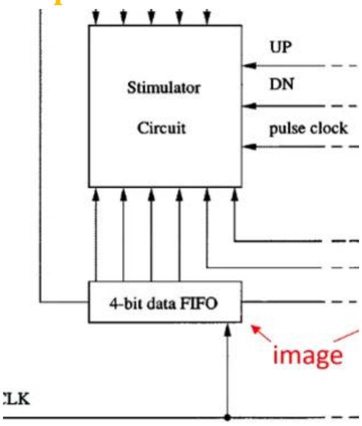
Timing

It is important to consider that we need to program the FIFO with significant speed.
Each stimulator requires for each frame to receive the 80 bits of the pulse timing description five times (one for each electrode)

$$Frame\ delivery\ time = 80\ clk \times 5 = 400clk$$

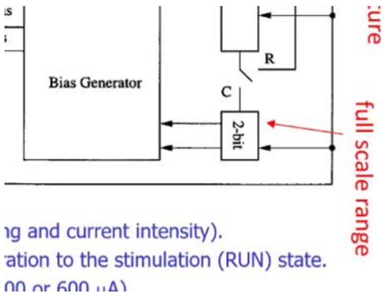
To be below the perception time we need to have that this can happen below $\frac{1}{50}s$

Amplitude

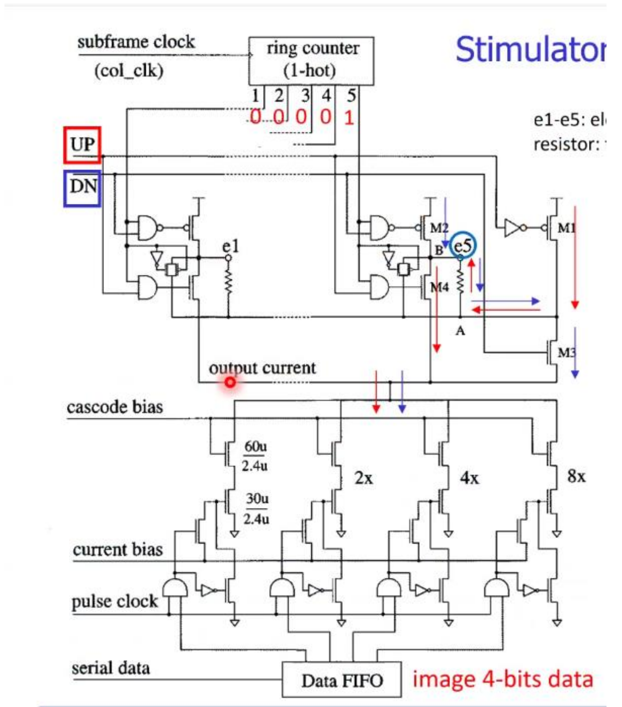


The amplitude is given by a FIFO of 4 bits as each stimulation is coded with 4 bits so that we have 16 gray levels. The content of the FIFO is provided by the received data.

We also need to provide the FULL SCALE RANGE which will be divided in the 16 levels.
This is selected with the 2 bits on the right, typical values could be 200,400,600,800 μA.



Stimulator circuit internal operation



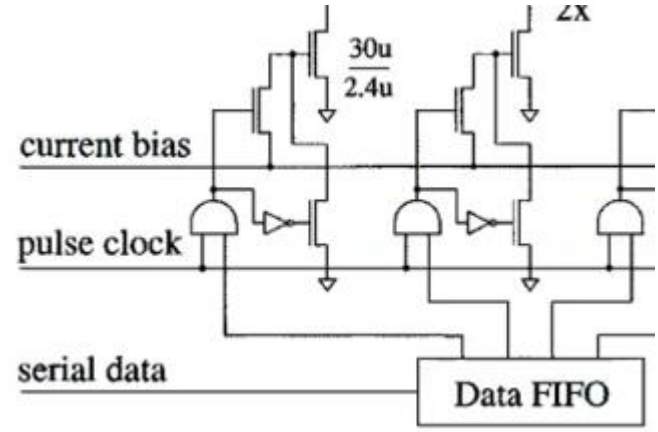
We know that the stimulator has to send a current into the tissue like for the pacemaker. However, we know that it took a long time to receive the directions for the stimulus.

We divide the stimulator in 2 branches an higher and a lower branch

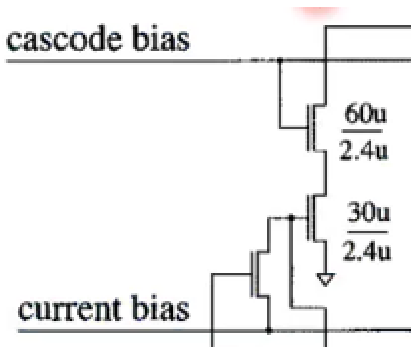
- The lower branch is a DAC which allows us to select an amount of current using the amplitude bits (16 possible values)
- The upper branch is a bi-directional current circulation circuit

DAC

The DAC is simple composed by a series of transistors of increasing size so that they each drive more current than the previous, we use the input bits to turn on or off the correct combination of transistors and obtain the desired output amplitude.



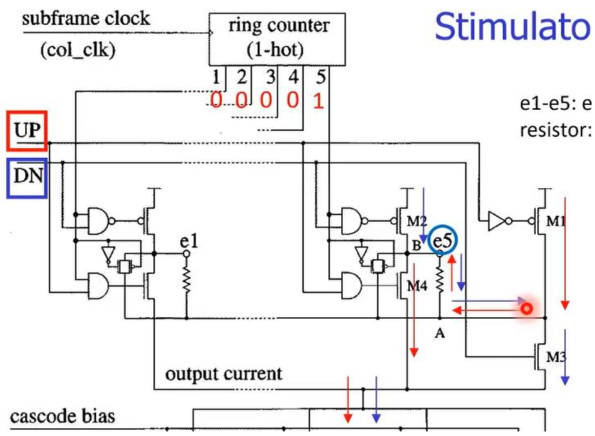
We can see that the AND port receives the data form the FIFO and then either connects the transistor gate to the bias voltage or to ground turning it on or off.
The bias is the voltage selected by the 2 bits we indicated before.

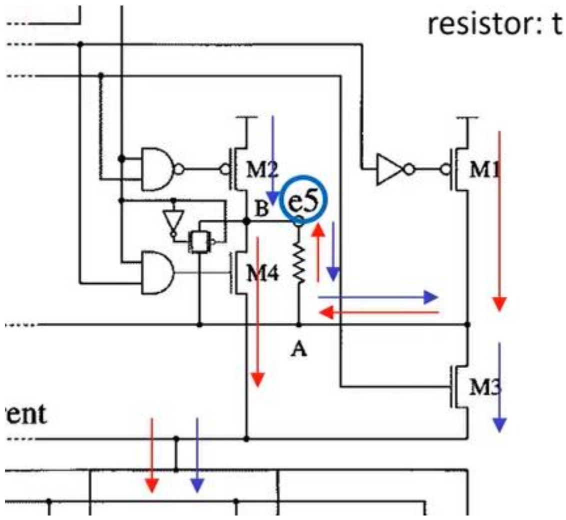


Additionally we can note that the output transistors are connected in a cascode configuration so that the output has a lower resistance.

Upper branch

Thanks to the lower branch we now have access to a current generator with the desired amplitude, however we need to be able to generate both a positive and a negative pulse.
This is made possible tanks to the upper branch: this circuit allows current to circulate inside it and we can switch on and off some switches to inverse the direction of this circulation





here we can see the example of the configuration of the electrode 5.

Up 1 Down 0

In this condition we have that

- $M1$ and $M4$ are on
- $M2$ and $M4$ are off

In this condition the current is called from the node B through the transistor M4.
So it passes in the electrode with direction BA and it is provided by M1

Up 0 Down 1

In this condition we have that

- $M1$ and $M4$ are off
- $M2$ and $M4$ are on

In this condition the current is called from the node A through the transistor M3.
So it passes in the electrode with direction AN and it is provided by M2.

Medical imaging

The purpose of medical imagining I sot look into the patient and identify the spatial distribution of parameters such as

- Morphology o tissues bones and organs
- Regions where pathologies are localized
- Physiological functionalities and their time evolution

Technologies for medical imaging are

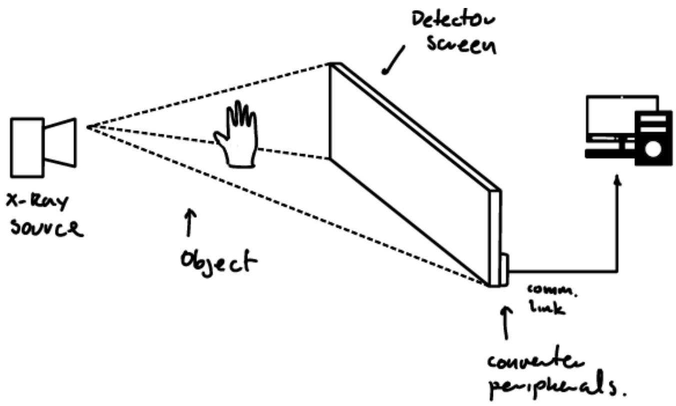
- **X-ray**: radiography (2 dimensional), X ray competed tomography CCT (3 dimensional, in Italiano TAC)
- **SPECT**: single photon emission computed tomography
- **PET**: positron emitted tomography

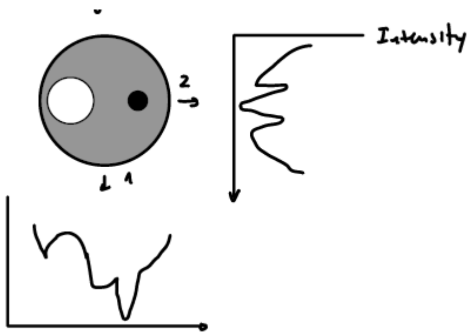
Other technologies non studied in this course are

- MRI magnetic resonance imaging
- Ultrasound (echography)

Radiography basic principles

This is the first technique invented, we shine an external X ray source on the patient and measure the amount of light which passes through them which will be inversely proportional to the density of the tissue encountered.

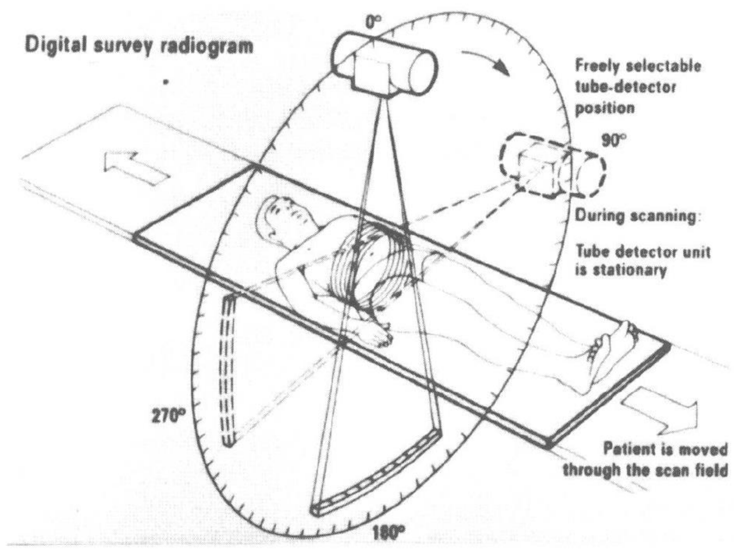




Computed tomography

With the process described above we obtain only a 2D description of the object to obtain 3D description we need to compare several 1D projections.

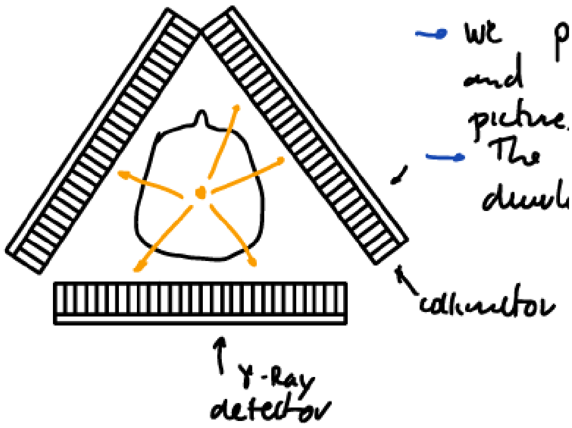
Spiral CT scan



The angle between each measure gives us the resolution of out final image analog to how frequency is relevant in the Shannon theorem

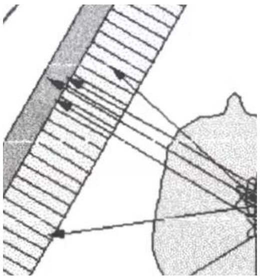
Single photon emission computed tomography SPECT

We place a radioactive substance producing γ rays in the patient and utilize an array of γ detectors to create an image.



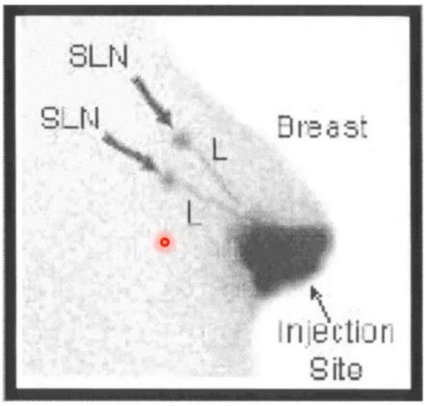
Detecting direction of the gamma ray

The main difference with radiography is that now the emission of rays is not form a specific point but isotopically inside the patient.
We detect the point of interaction of the gamma ray with the detector but we do not know its origin position.



Collimator

A collimator is a structure made form a very dense material capable of absorbing gamma rays.
The collimator will present holes to allow the gamma rays to pass through it an be detected, these holes create channels which make it so that only rays coming from certain direction will be able to pass through them, while others will be blocked by the walls of the collimator.

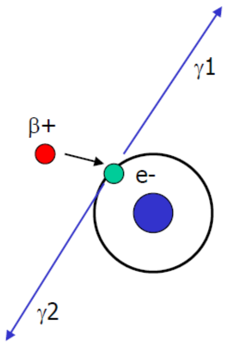


Scintigraphy

This is a technique utilized to identify sentinel lymphnodes for radio guided surgery.
We inject a tracer close to the tumor and take an image of the tracer, the darkest part indicate the closest lymphnodes.

Positron emission computed tomography PET

Positrons generation

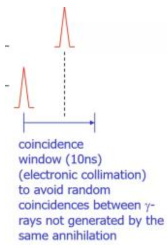


In this case the radiotracer emits a positron rather than a gamma ray.
Usually we utilize isotopes with too many protons with respect to the neutrons so what happens is that one of the protons become a neutron releasing a positron which will eventually interact with an electron.

Positron annihilation

When the 2 antiparticles interact they annihilate releasing 2 gamma rays in opposite directions as shown in the figure, the reason why the 2 gamma rays travel in opposite directions is because the momentum must be conserved so since the center of mass between the electron and the positron will have a total momentum of zero.

Electronic collimation



Since the gamma rays will be paired we can reconstruct the direction considering that simultaneous detection will be matched together.
This is important as we now have a higher detection efficiency than with SPECT since we do not lose photons because of the collimator absorption.

Detection efficiency comparison

In SPECT the sensitivity is low if we inject a radio tracer inside the patient we can expect a sensitivity of 10^{-4}, instead in PET we can expect a sensitivity of 10^{-2}.
This is important because we are limited in the amount of activity we can generate inside the patient for safety reasons.

The reason why SPECT is still used is because PET is more complex and the isotope utilized have a very short shelf life and need to be produced near the patient.

Key parameters of imaging techniques

Sensitivity

Efficiency of the imaging system to detect the parameter of interest

Selectivity (specificity)

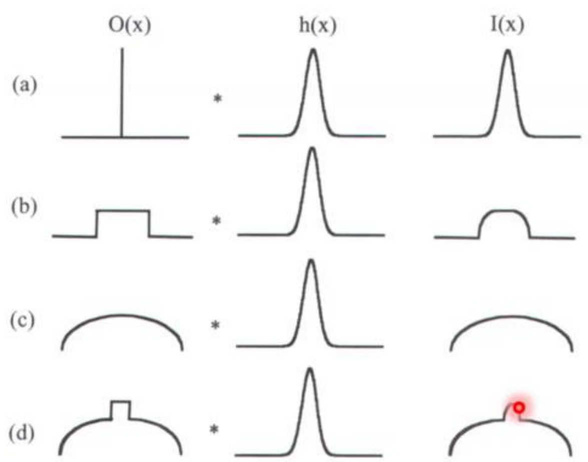
Capability of the system to distinguish the parameter of interest form other possible signals

Resolution and contrast

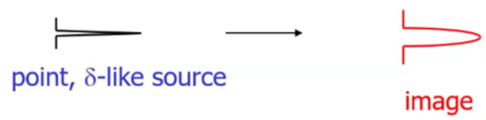
Capability of the system to distinguish details of the distribution very close to each other and separate regions with different concentrations

Spatial resolution: point spread function

$\boxed{I(x,y) = O(x,y)*h(x,y) \qquad h(x,y)\text{: PSF}}$

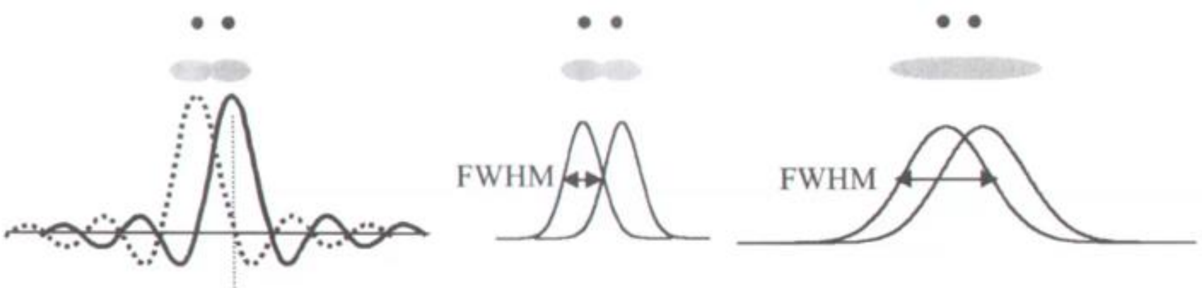


Like in a circuit we need to consider the impulse response of the circuit to understand how a waveform is transferred across it, in medical imaging we need to consider that different shapes will be transformed, the base of this transformation is the point spread function which indicates how a delta like point source is transformed, form this becomes possible to obtain the final shape for each shape by making the convolution of the original shape by the PSF.



Relationship between PSF and spatial resolution

The spatial resolution (minimum distance at which 2 object appear distinct) is defined as the full half width maximum of the PSF, meaning the width of the PSF curve at half of its amplitude.



The value of FWHM can be found considering the PSF as a gaussian curve

$$h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-x_0)^2}{\sigma^2}\right)$$

Gaussian PSF
(or Gaussian approx. of PSF)

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\sigma \cong 2.36\sigma$$

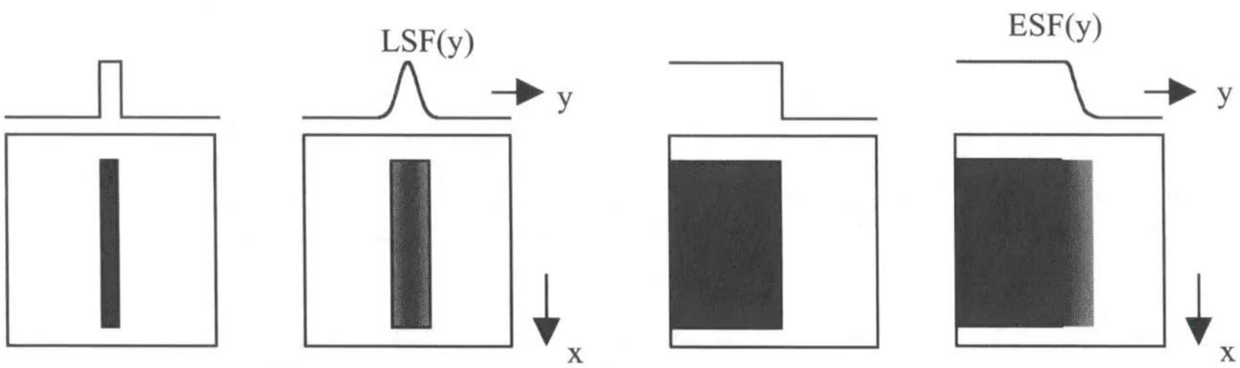
Full-Width-at-Half-Maximum

Note

If the shape of the PSF is not gaussian the FWHM does not completely describe the PSF but we are also going to need the MTF

Line spread function LSF and edge spread function

Sometimes it is not possible to have a point like source so instead we refer to a line object or to the edge of a square.



Modulation transfer function

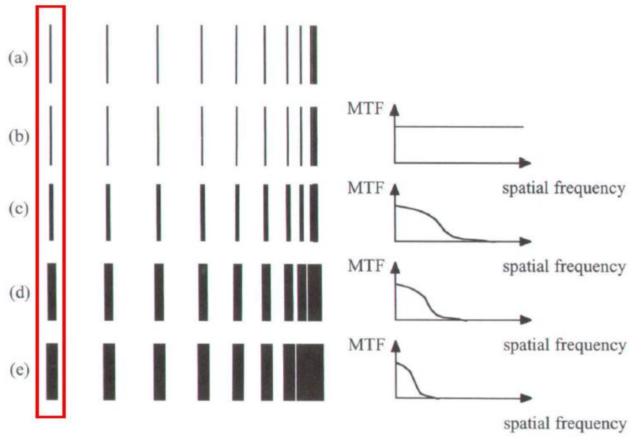
We know that in the case of an electronic system the impulse response of the system is the **inverse Fourier transform of its transfer function**.
Similarly in the imaging word we can define the MTF (modulation transfer function) which allows us to go from the space domain to the spatial frequency domain (analog of moving from time to frequency with Fourier transform).

$$\mathrm{MTF}(k_x, k_y, k_z) = \int\int\int \mathrm{PSF}(x, y, z) e^{-j2\pi k_x x} e^{-j2\pi k_y y} e^{-j2\pi k_z z}\, dx\, dy\, dz$$

We will be able to obtain the output of our system by simply multiplying the MTF of the system by its input if we operate in the spatial frequency domain.
In the same way the total transfer function of multiple blocks is equal to the product of the MTF transfer functions of each block.

Example

Let's expand on the concept of spatial frequency domain.
A series of images repeated in space can be equated to a signal in time.
The distance between the images represents the frequency of the image signal.
On the right we have the representation of the same image signal and how it is resolved with different MTF.
We can see that n the case a) since the MTF is constant we can resolve precisely every line even when they get close together, instead in the other cases where the MTF falls after a certain spatial frequency the lines are resolved with more and more inaccuracy becoming wider at higher frequencies until eventually we can no longer distinguish line too close together.

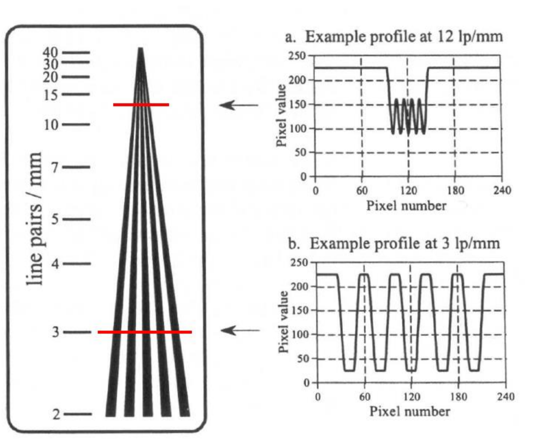


MTF computation

We can calculate ethe MTF either mathematically using the Fourier transform or utilizing instrumentation which tests for different spatial frequencies inputs

Example

As we get closer to the top the lines become closer and thus their spatial frequency increases.



SNR

The content of a pixel of our image is the number of radioactive cases coming form that volume, if we repeat the acquisition multiple times the average value should be constant however we are going to have variations between the single values introduced by Poisson statistics

Poisson statistics

When counting events, the probability $P(N)$ to count N events given an average μ of counts acquired in the time T is

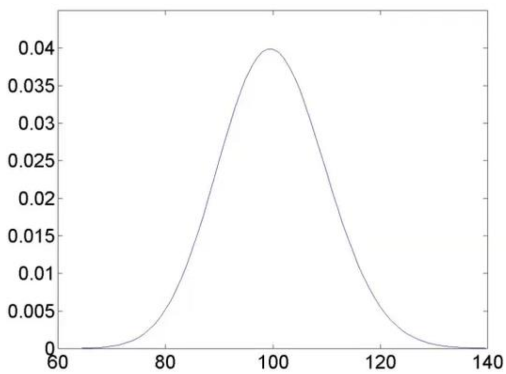
$$P(N) = \frac{\mu^N e^{-\mu}}{N!}$$

Where the variance is equal to the average count

$$\sigma^2 \simeq \mu$$

So the larger is the signal the lower is the influence of the noise as the RMS is

$$\sigma \simeq \sqrt{\mu} = rms$$

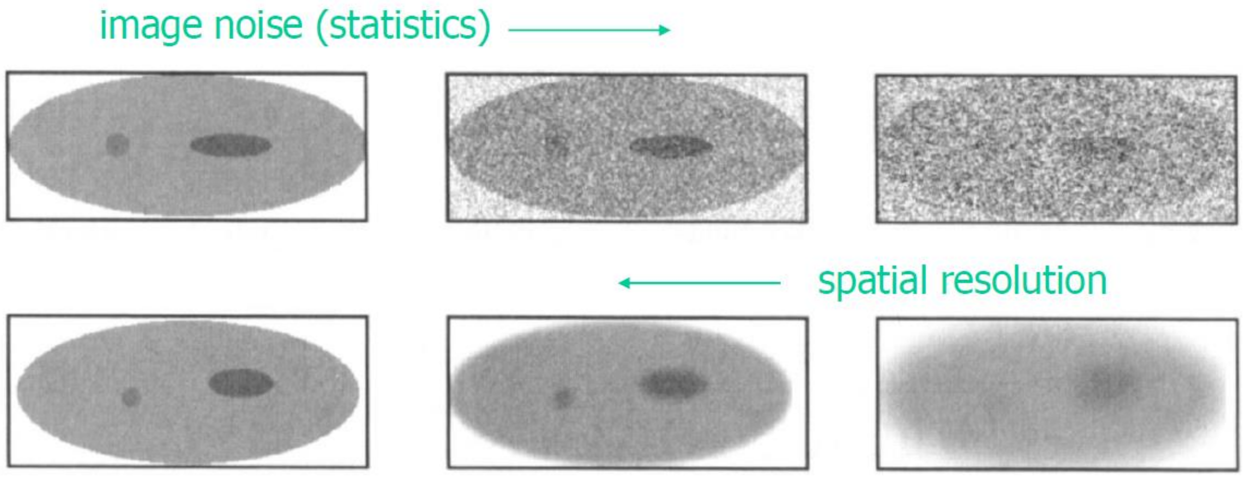


And thus the SNR is

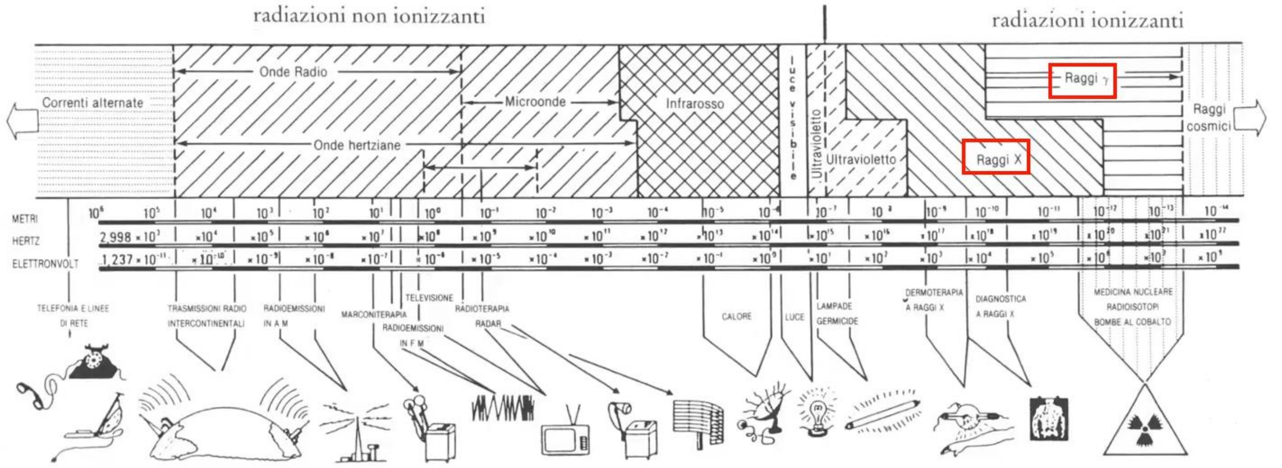
$$SNR = \frac{N}{\sigma} = \frac{\mu}{\sigma} = \sqrt{\mu}$$

Image contrast

Below we can see the effects of increasing the statistical noise (acquiring less events) on top
Decreasing the spatial resolution on the bottom



Electromagnetic spectrum

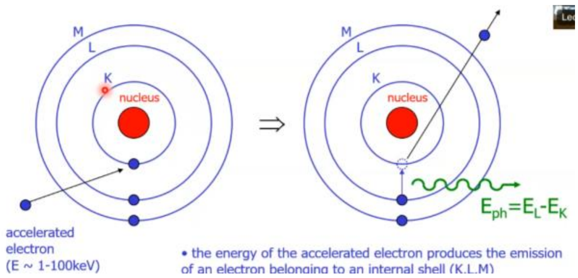


X rays	γ rays
Energy $$1keV \dots 100-300keV$$ **Origin** They are generated form atoms through fluorescence or bremsstrahlung, synchrotron light. **Applications** Radiography	**Energy** $$100keV \dots 10MeV$$ **Origin** They are generated form nucleus through nuclear emission or annihilation of positrons **Application** SPECT and PET

X rays

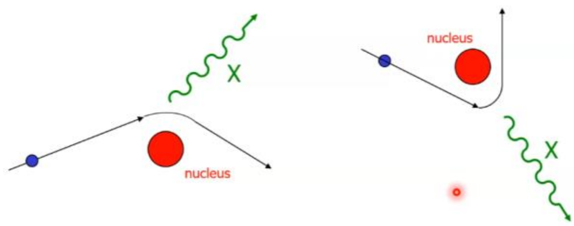
Origin of X rays

Florescence



An accelerated electron with energy between $1keV$ and $100keV$ hits an electron of the inner shells of the atom.
The electron is released leaving an empty space in the shell so one of the electrons form the outer shells decays taking its place.
A photon with energy equal to the difference between the 2 states is released, thus the X ray released has a single frequency equal to the energy gap between the 2 states

Bremsstrahlung



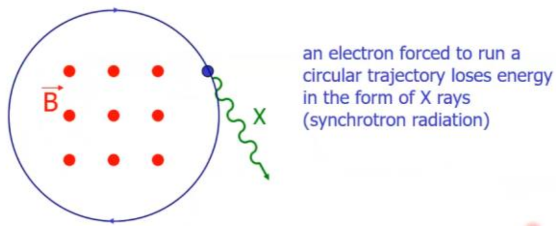
In this case the radiation is produced when an electron changes its velocity because of a Columbian interaction with the atomic nucleus.
The energy releases is proportional to

- The electron energy
- The Z of the absorbing material (higher deflection means higher energy)

The emitted X ray has a continuous spectrum from 0 to E_{MAX} which is the kinetic energy of the impacting electron in the case it is fully absorbed in the interaction

Synchrotron light

An electron is forced to run a circular trajectory an thus loses energy in the form of X rays because it loses the normal component of its velocity



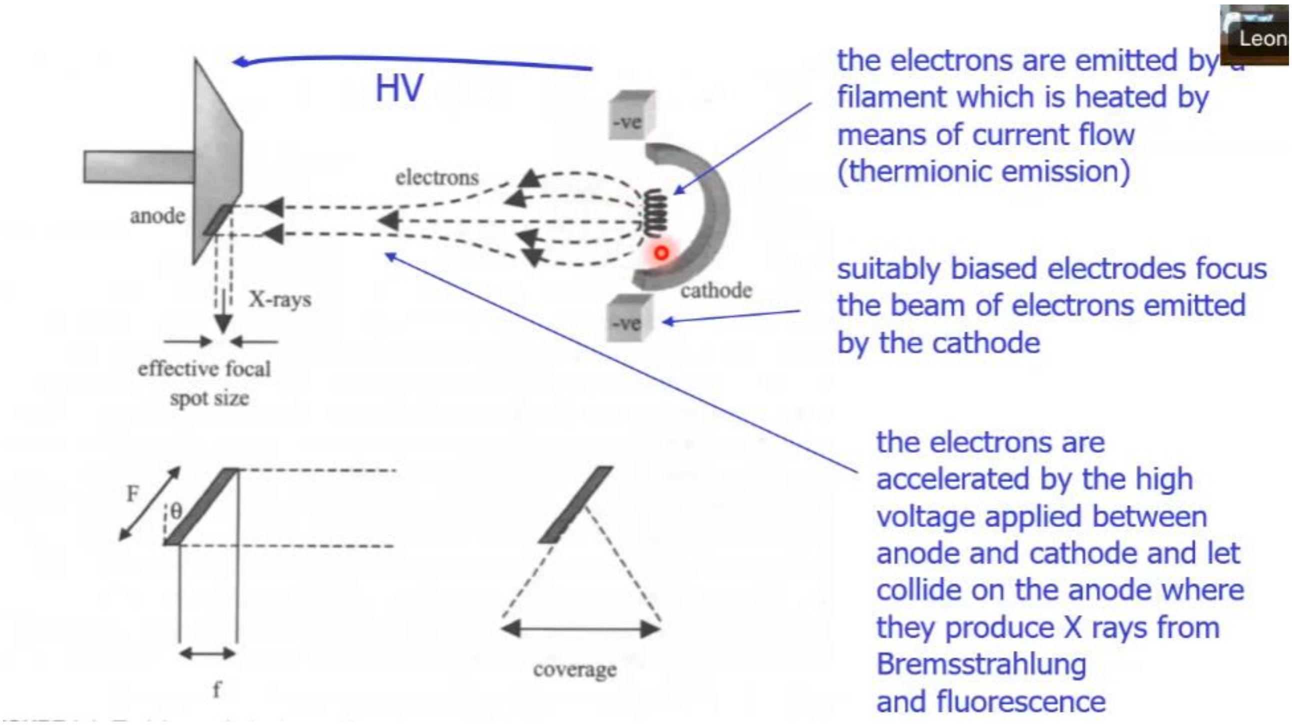
In summary

- With bremsstrahlung: the loss of tangential velocity releases energy in the form of X rays
- with synchrotron light: the loss of normal velocity releases energy in the form of X rays

X ray tube

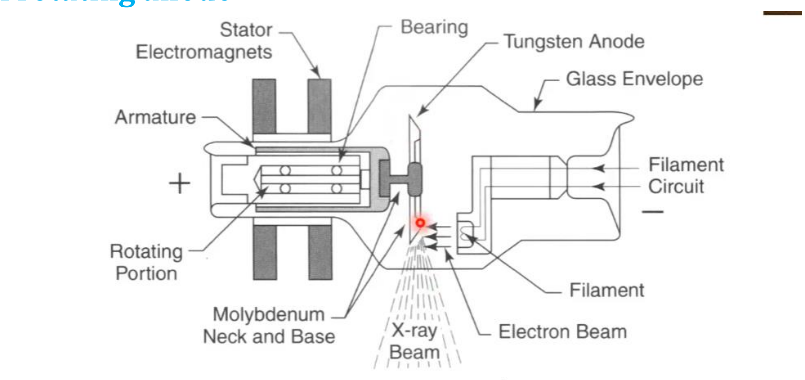
This is a machine which allows to generate X rays both from florescence and bremsstrahlung.

We create high energy electrons accelerate them and make them it a target to slow them down and thus generate the X rays.



Architecture of an X ray tube with rotating anode

The anode is a rotating piece of metal, it needs to rotate because the conversion of the electron kinetic energy into X rays is very inefficient, and most of the energy is converted in thermal energy. We need to change the position of interaction between anode and electrons to avoid damaging the anode.

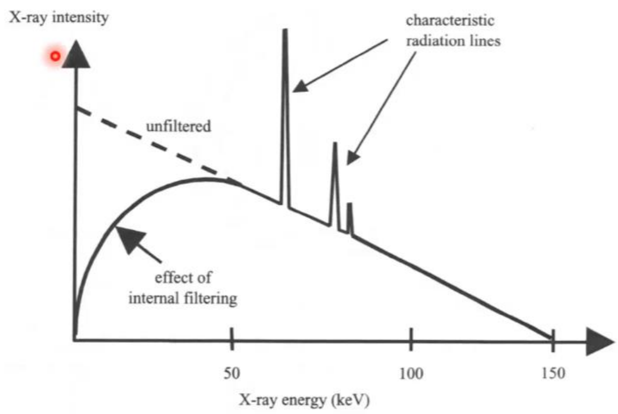


Around the X ray tube we have shielding to prevent the X ray to exit in unwanted directions.
Note also that the tube is a vacuum to avoid

- slowing down the electrons
- accelerating other ions together with the electrons

Typical spectrum of an X ray tube

- The low energy X ray can not exit the X ray tube so we have a filtering effect for low energy.
- The continuous line in the spectrum is caused by the Bremsstrahlung effect
- The peaks in the curve are instead caused by fluorescence



Which energies we want to use

In X ray imaging we typically use energies in the range from a few tens keV up to a few hundreds.
The smaller and the softer is the object the less energy we want to use while the opposite is true for thicker and denser objects.

γ rays

Origins of γ ray

Gamma ray originate form the radioactive decay of isotopes.

Radioactive isotopes

Atoms are defined by an atomic number which indicates the number of protons in their nucleus.
For each atom there can be multiple isotopes: atoms with the correct number of protons but a different number of neutrons, these isotopes may be stable or unstable.
Unstable isotopes can decay through different mechanism

- A neutron can become a proton
- A proton can become a neutron
- We might eject neutrons or protons

Some of the main mechanism are

Alpha decay

We have the ejection of an helium nuclei: so 2 protons and 2 neutrons, this is not typically used in medical imaging as they are very dangerous.

Beta plus decay

In this case we have that a proton becomes a neutron releasing a positron, this is what is used in PET

Beta minus decay

In this case a neutron becomes a proton and an electron is created.

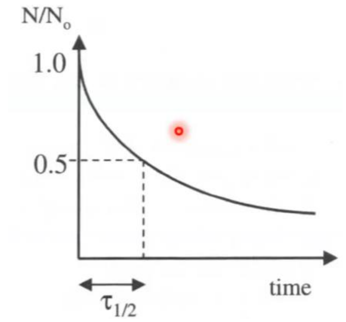
Radioactive decay ratio

The number of decays per unit of time is

$$A = -\frac{dN}{dt} = \gamma N$$

Where

- A is the activity of the radionuclide (becquerel = disintegrations/second)
- N is the number of nuclei
- λ is the decay constant



The number of remaining atoms for a certain isotopes after a certain amount of time is given by the exponential formula

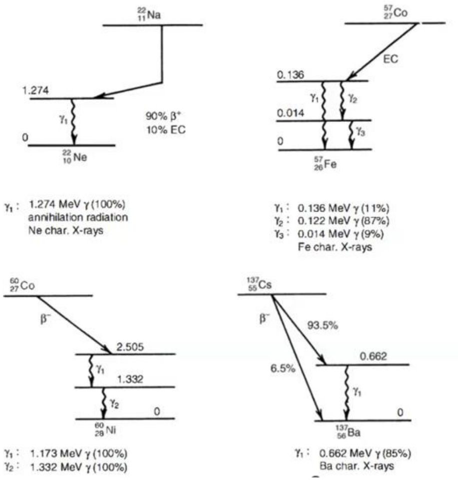
$$N = N_0 e^{-\lambda t}$$

Where N_0 is the initial number of isotopes.

Half life constant

We define the half life constant as the time required for the isotope population to be cut in half

$$\tau_{1/2} = \frac{ln\,2}{\lambda}$$



Summary

Gamma rays are emitted by excited nuclei in their transition to lower lying nuclear levels.

The excited nuclear states are created in the decay of a parent radionuclide.

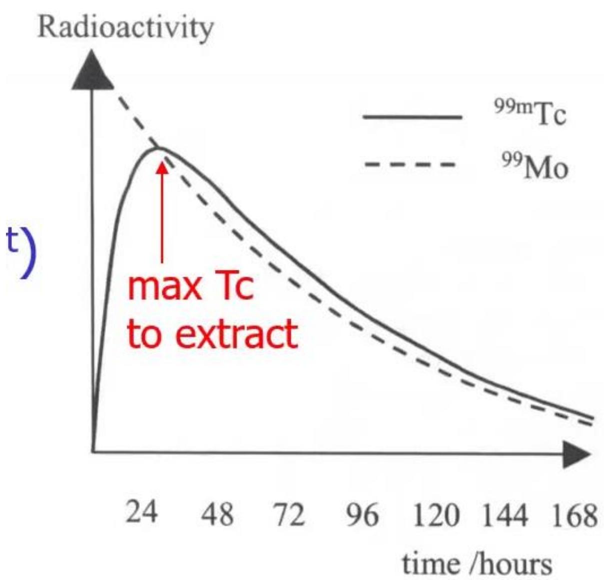
Technetium 99

This is one of the most popular γ ray emitting isotopes, its decay follows this scheme

$$^{99}_{42}\text{Mo} \xrightarrow{\tau_{1/2}\ 66\,\text{h}} \beta + {}^{99\text{m}}_{43}\text{Tc} \xrightarrow{\tau_{1/2}\ 6\,\text{h}} {}^{99\text{g}}_{43}\text{Tc} + \gamma$$

What makes it so popular is that the middle state (which is the exited state of the decayed isotope) is metastable and has a half life period of 6 hours which makes it particularly useful for diagnostics, as opposed to most other materials whose decay is much faster so much that almost all the isotope would decay before we could actually perform any tests.

Starting form the initial isotope we obtain the maximum about of technetium 99 after about 24 hours.



Additionally also the energy of the obtained gamma rays is good for diagnostic (140keV)

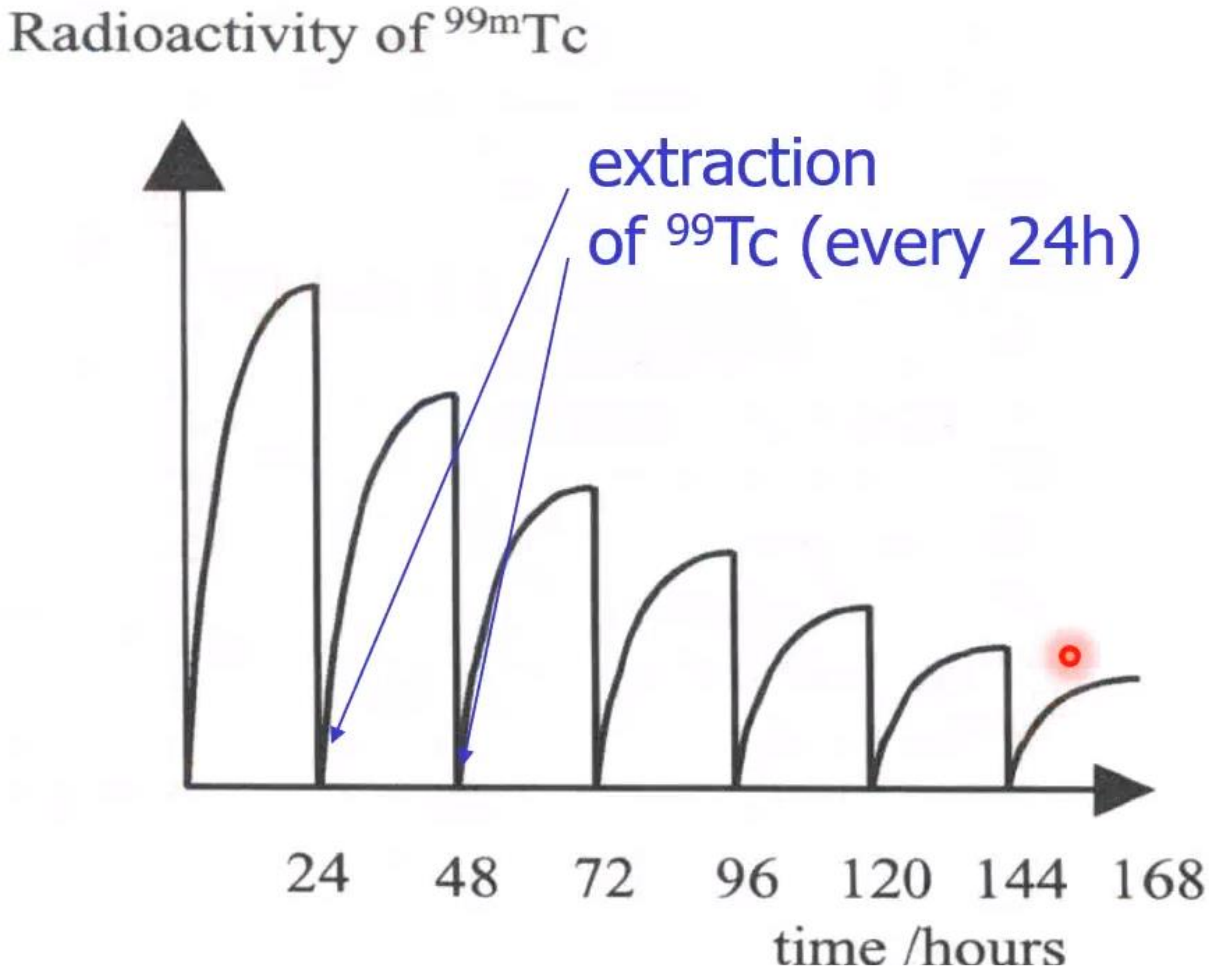
Requirements for radionuclides for diagnostics

Few general notes on radionuclides for diagnostics:

- $\tau_{1/2}$ too short $\Rightarrow$ short time between injection and diagnostics
- $\tau_{1/2}$ too long $\Rightarrow$ low activity, patient radioactive for too long
- $E\gamma$ too low $\Rightarrow$ few rays reach the detector
- $E\gamma$ too high $\Rightarrow$ patient 'transparent', difficult detection

Production

We start from Mo, we have that once this decays the technetium dissociate form the original crystal so we can simply wash it away, thus to collect the technetium we can simply keep a box of Mo and wash away the produced technetium with daily intervals



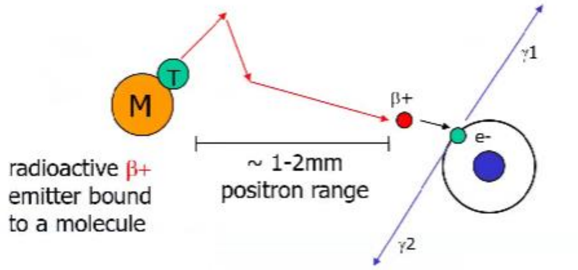
Note that Molybdenum is generated in nuclear reactors by either nuclear fission of by enriching base Molybdenum.

RADIONUCLIDES $\beta +$ emitter for PET

We also obtain gamma rays utilizing an element with a $\beta +$ decay.

$\beta +$ emission

This is the case of a isotope whit too many protons, in this case a proton becomes a neutron and releases a positron.



When the positron hits an electron the 2 analyte and 2 photons with frequency in the order of gamma rays are generated and propagate with opposite directions.

1) β+ emission:

$$^{A}_{Z}X \rightarrow {}^{A}_{Z-1}X + \beta^{+} + \upsilon$$

2) annihilation:

$$\beta^{+} + e^{-} \rightarrow \gamma + \gamma$$

β+ emitter isotopes typically used:
^{11}C, ^{15}O, ^{18}F, ^{13}N

Resolution limit

Since the gamma ray is not generated exactly where the isotope decay happens we have a minimum resolution of about 1 or 2 mm distance travelled by the positron before hitting an electron.

Limit of the half time

The half time is very short in the **order of tens of minutes,** so we need to produce the isotopes right before utilization

TABLE 2.3. Properties of the Most Common Radionuclides Used for PET

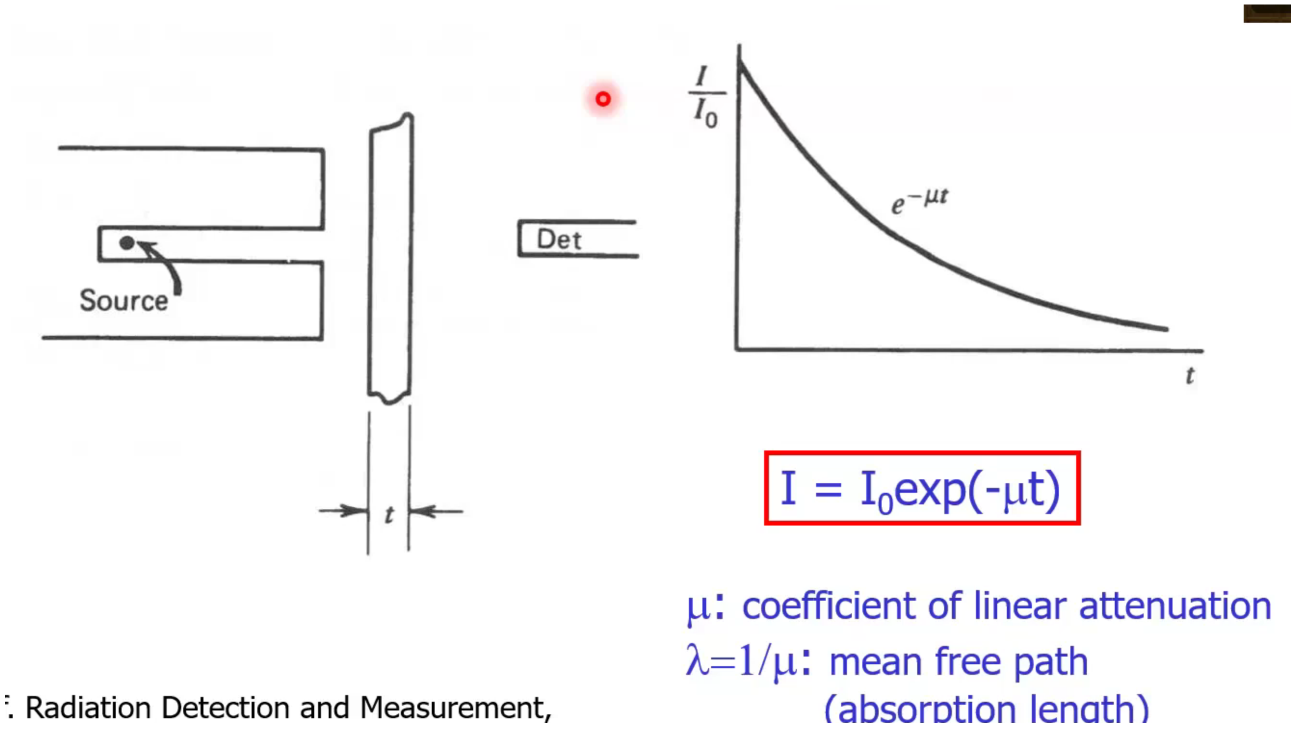
Radionuclide	Half-life (min)
^{11}C	20.4
^{15}O	2.07
^{13}N	9.96
^{18}F	109.7

- β+ radionuclides are produced in a cyclotron (directly in hospital) by means of irradiation with protons (~10MeV) or deuterium (~5MeV)
- they are bound to the molecule by means of chemical synthesis

$$^{12}C \xrightarrow{\text{proton}} {}^{11}C + \text{proton} + \text{neutron}$$
$$^{14}N \xrightarrow{\text{proton}} {}^{11}C + {}^{4}_{2}He$$
$$^{16}O \xrightarrow{\text{proton}} {}^{11}N + {}^{4}_{2}He$$
$$^{13}C \xrightarrow{\text{proton}} {}^{13}N + \text{neutron}$$
$$^{15}N \xrightarrow{\text{proton}} {}^{15}O + \text{neutron}$$
$$^{14}N \xrightarrow{\text{deuterium}} {}^{15}O + \text{neutron}$$
$$^{18}O \xrightarrow{\text{proton}\ \cancel{\text{deuterium}}} {}^{18}F + \text{neutron}$$
$$^{20}Ne \xrightarrow{\text{deuterium}} {}^{18}F + {}^{4}_{2}He$$

Interaction of X and γ rays with matter

We have that to detect X and γ rays we first need to be able to stop them so it is necessary to know how they interact with matter.



The interaction between high energy photons and matter is determined by a stochastic process.

The intensity of photons emerging from the material has an inverse exponential dependance with

- The thickness of the material t
- The inverse μ of the mean free path λ, $\mu = \frac{1}{\lambda}$

$$\boldsymbol{I = I_0 e^{-\mu t}}$$

Important

Note that the absorption coefficient is not only a function of the element but also of its density ρ for this reason we typically write utilizing the coefficient of mass absorption $\mu' = \frac{\mu}{\rho}$

$$\boldsymbol{I = I_0 e^{-\mu' \rho t}}$$

μ′ depends on 3 absorption mechanisms with which the photos can interact with matter

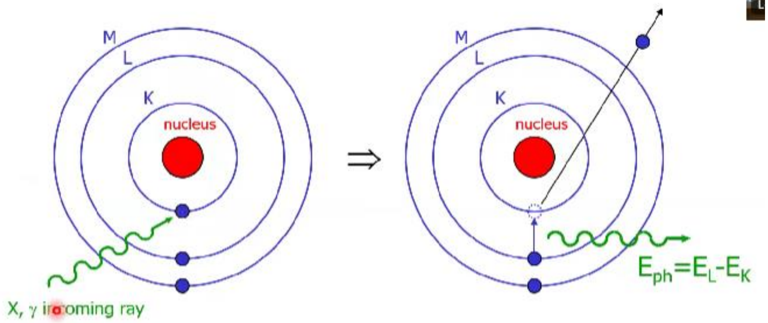
- Photoelectric absorption
- Compton absorption
- Production of e^-/e^+ pairs

$$\mu' = \mu'_{photoelectric} + \mu'_{compton} + \mu'_{pair}$$

Photoelectric absorption

In this case

1. The ray gives all of its energy to an electron which absorbs it
2. The now highly energetic electron leaves the atom
3. An x ray will be generated as an electron from an outer shell will decay to fill the spot left by the first electron



Measurement

This process allows us to measure the energy of the gamma ray by measuring the energy of the electron which will more easily interact with matter.

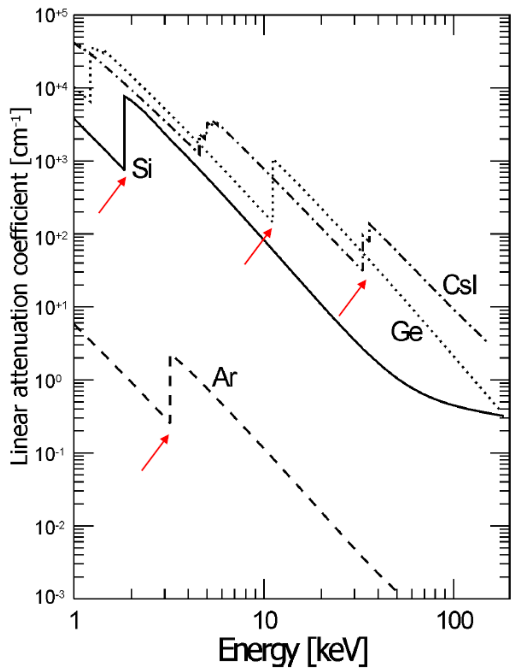
We make the electron travel through a medium, as it passes through it will release energy which will free charge, we collect the charge and form it determine the intensity of the gamma ray

Note

We need to consider that not all the energy is passed to the electron as some will be present in the X ray, this is not too significant since

1) The energy of the X ray will be much lower than the gamma ray total energy
2) The X ray could still interact with matter realizing charge, in this case all energy would be measured

Problem of photoelectric interaction



The interaction probability for photoelectric effect is given by

$$\mu'_{photoelectric} \propto \frac{Z^n}{E_\gamma^3}$$

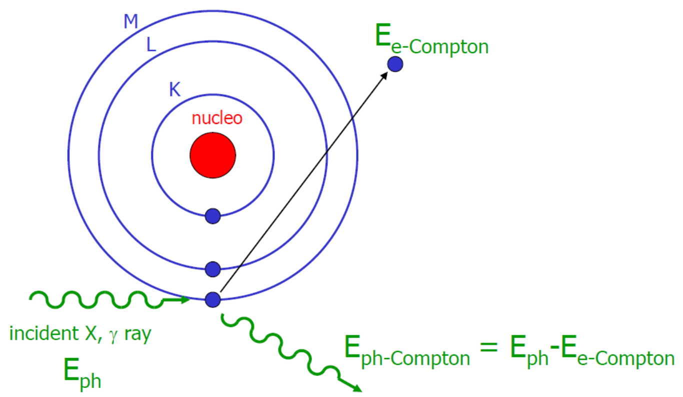
As we can see it is

- Directly proportional with the material density Z (note $n \simeq 4$)
- Inversely proportional to the energy of the gamma ray

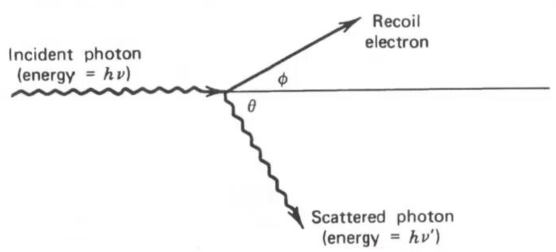
Additionally, the probability decreases sharply as the energy increases when the photon stops being able to interact with the election of a particular shell whose energy is now too low.

Compton scattering

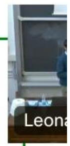
In this case the released electron is not able to absorb the energy of the gamma ray entirely, so after the interaction we still have a lower energy gamma ray also scattered from the impact.



Now less energy is transferred to the electron and so we need to improve the probability for the other gamma ray to also interact.
To do so we need to consider the conservation of momentum so that we can define the possible direction of the ray.



$$\boxed{h\nu' = \frac{h\nu}{1 + \frac{h\nu}{m_0 c^2}(1 - \cos\theta)}}$$



Energy of emitted photon

$$\boxed{\frac{d\sigma}{d\Omega} = Zr_0^2 \left(\frac{1}{1 + \alpha(1 - \cos\theta)}\right)^2 \left(\frac{1 + \cos^2\theta}{2}\right) \left(1 + \frac{\alpha^2(1 - \cos\theta)^2}{(1 + \cos^2\theta)[1 + \alpha(1 - \cos\theta)]}\right)}$$

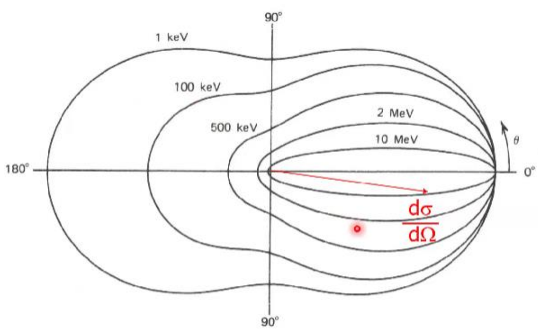
fraction of emitted photons at a given angle θ (Klein-Nishina)

r_0: e- radius
$\alpha = h\upsilon / m_0 c^2$

$\boxed{\mu'_{Compton} \div Z}$ (the probability depends on the number of electrons available for the hits and is therefore proportional to Z)

Basic concepts

- Only part of the energy is transferred
- The lower is the energy the smoller is the angle between the incoming and outgoing photon
- Depending on the energy it is more or less probable that the outgoing photon goes in the forward or backward direction: it is more probable for a lower energy photon to bounce back



Once we have a detector and we want to measure the energy of a gamma ray we must be able to obtain both photons so we design the collector so that we can capture also the scattered photon.

Elastic scattering (Rayleigh scattering)

Not very probable but still possible: the photon is bounced back without losing any energy, we do not lose energy but we lose information regarding direction.

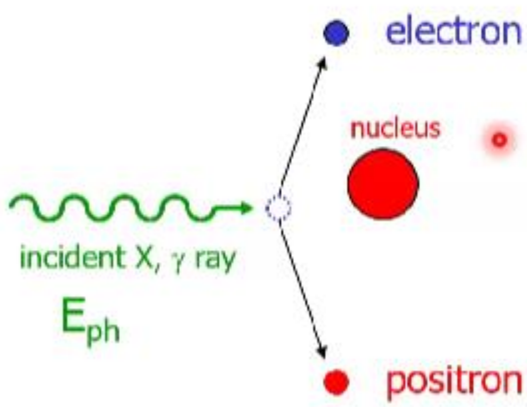
Problem with scattering

We wish to detect the source of the photons so if we have scattering when we detect the incoming photons, we have false data as the directions of the scattered photons are not that of the original photons.

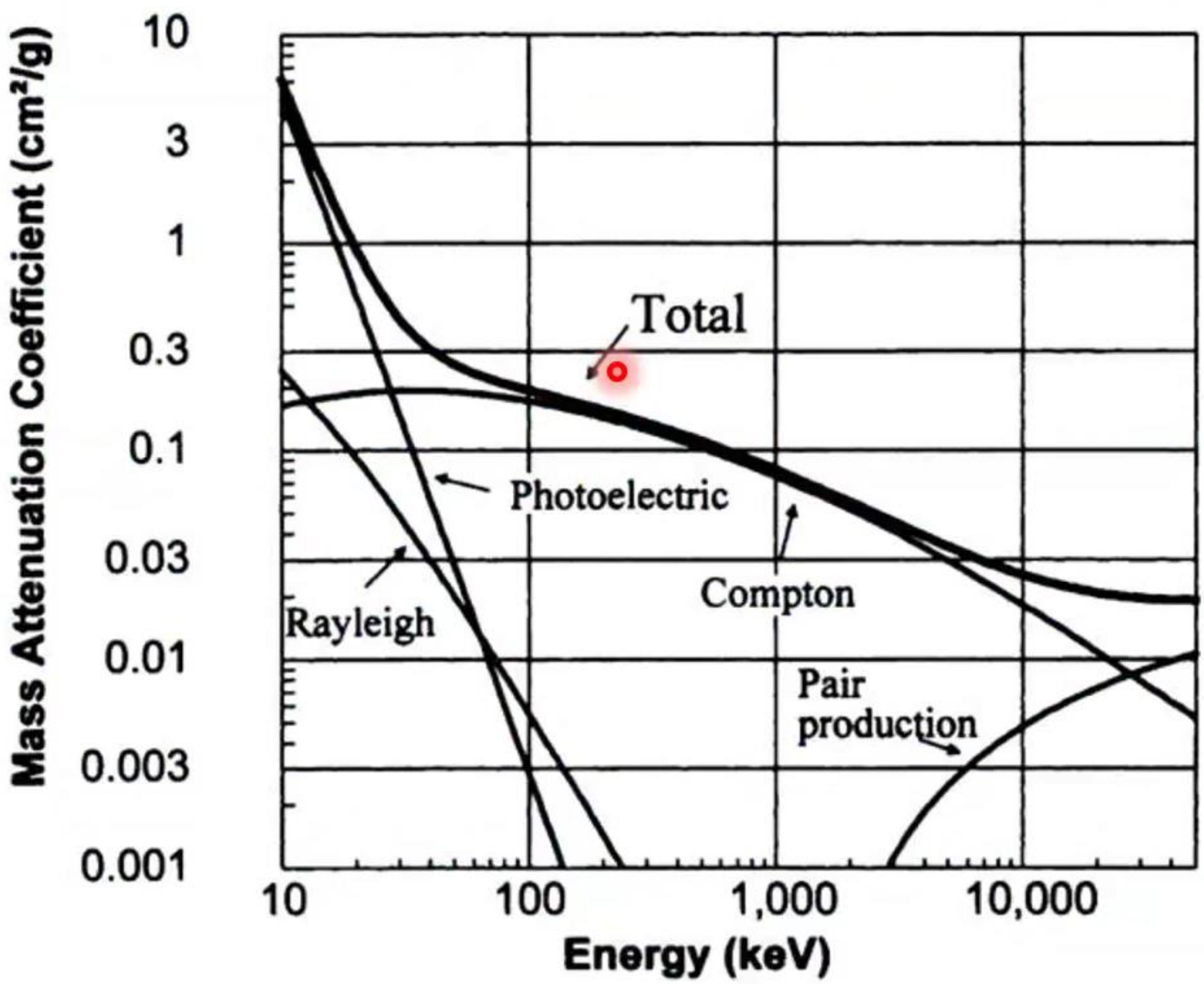
In the case of Compton scattering, we have that the scattered photons have less energy so we can filter out their results, this is not true for Rayleigh scattering (elastic scattering).

Generation of pairs

Not common in medical application as this process has a minimum threshold energy which corresponds to double the energy of an electron at rest in the Columbian field of a nucleus.
One such photon has a probability to split into an electron and a positron and this probability increases when the photon is affected by the Coulombian field of the nucleus.



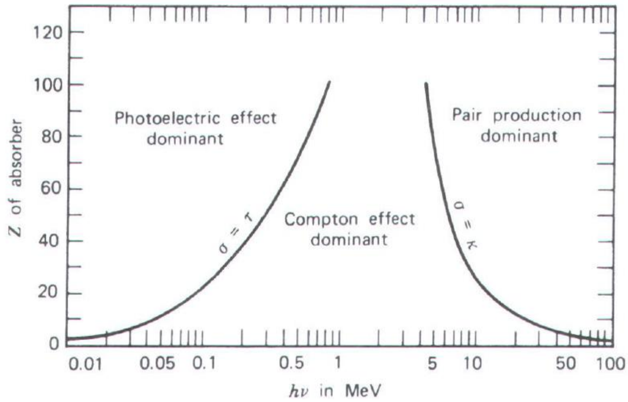
The reason why we have the energy limit is that for this to occur the photon should be able to provide enough energy to turn into the mass of the 2 particles



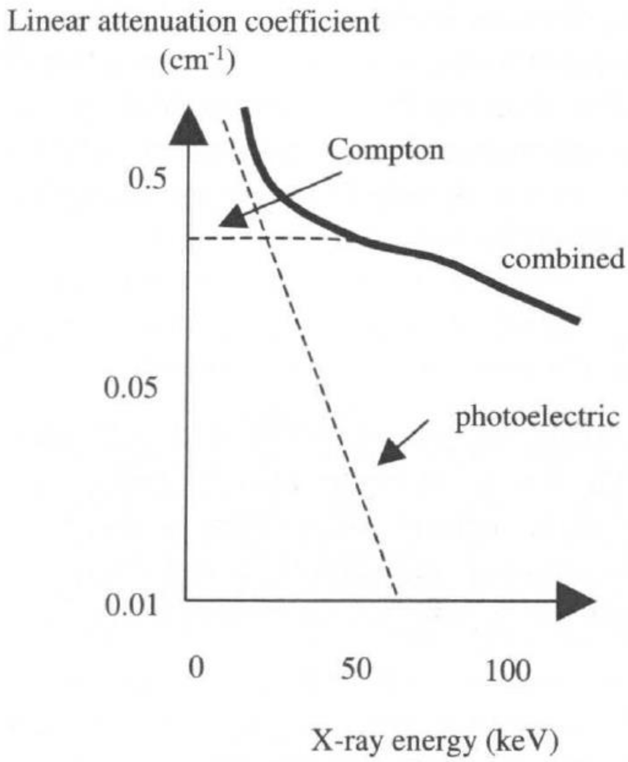
Dependence of the absorption probabilities for the 3 mechanisms as a function of energy and density

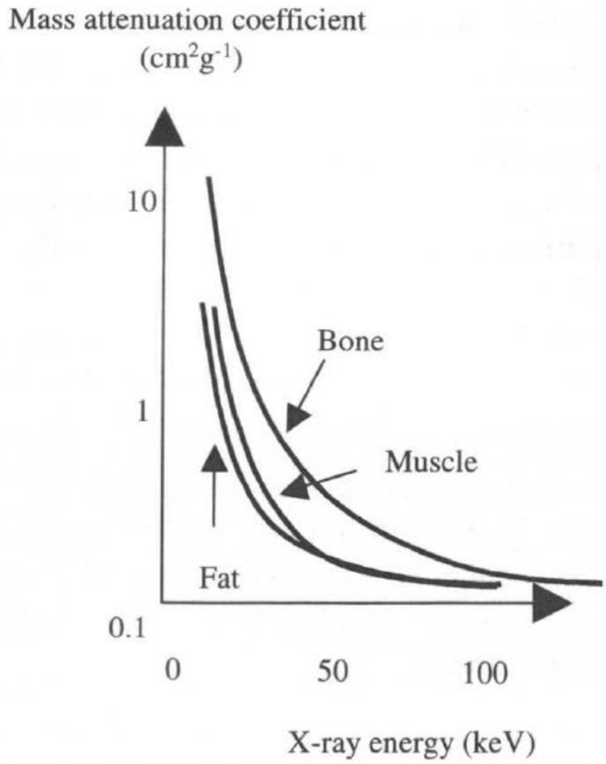
Note that the Z indicates the density of the material: as we increase the density of the material the probability of the photo electric effect increases for the same energy level.

We want to operate in ranges where the photoelectric effect is dominant rather than the Compton effect as the photoelectric effect makes it easier to detect the direction of the photons.



When we design a detector we need to know which photons we need to detect to know which material to select for the detector to have mostly photo electric interactions.





Detectors

A detector converts the energy released by a photon in the material into an electric signal.
The electric signal is then processed by a suitable electronic circuit to determine

- The interaction point
- The energy of the photon
- The time instant of the interaction

Types of detectors

Detectors can be divided according different characteristics like

Direct conversion detectors

In these detectors the photon energy is converted directly in a quantity of electric charge which is then collected at an output electrode

Indirect conversion detectors

In these detectors the energy of the photon is converted in another physical quantity (like for example visible photons) and a secondary detector is necessary to convert eh second physical quantity in an electric signal

Advantage

The material for the conversion can be different from the one which creates electric signal and thus selected more freely.

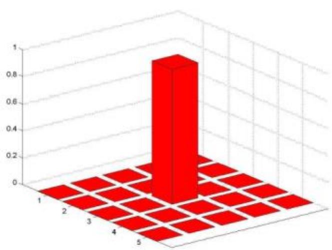
Disadvantage

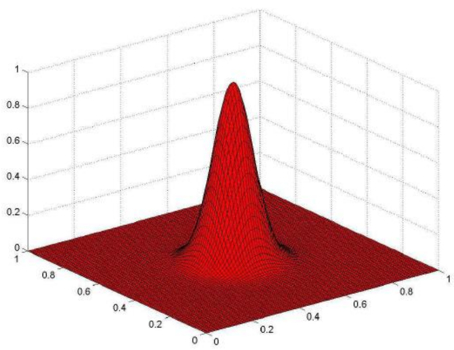
The cascade of 2 processes instead of one worsens the resolution on the overall conversion

Another subdivision is

Pixel detectors

This is a segmented detector, the resolution in the image is given by the dimension of a pixel





Continuous detectors

Values in the x, y axis change with continuity

Key parameters

Spatial resolution

Our ability to determine the position of interaction.
We obtain a PSF for the detector, note that this is not the same as the PSF of the scanner as it indicates only a single element of the total.

Energy resolution (values around 10%)

The precision with which we determine the photon energy
Note that the resolution is typically indicated for different energies.
Having a good energy resolution allows us to discard events caused by scattering which will present a lower energy and create noise in the measurement of the direction.

$$\boxed{G(E) = \frac{N_0}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(E-E_0)^2}{2\sigma^2}\right)}$$

$$\boxed{\mathrm{FWHM} = 2.35\,\sigma} \qquad \boxed{R = \Delta E_{FWHM}/E_0}$$

Detection efficiency

This indicates the **number of detected events over the number of events generated by the source**.
This can be found as the product of 3 elements

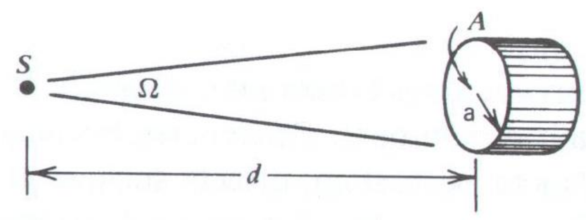
$$\boxed{\text{Detection efficiency} = \text{Geometrical efficiency} \times \text{Absorption efficiency} \times \text{'Photopeak' efficiency}}$$

Geometrical efficiency

This indicates the fraction of photons emitted which actually enters the detector.

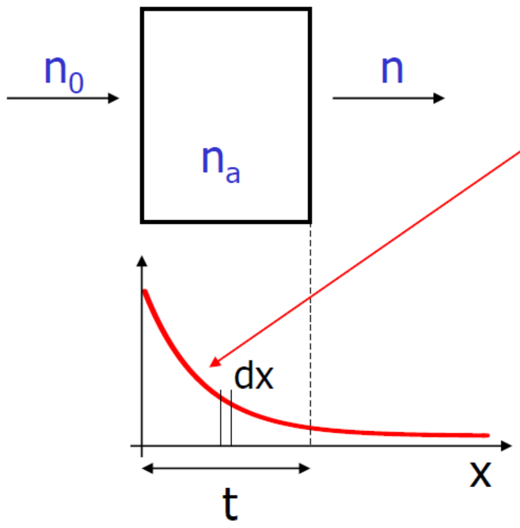
$$\eta_G = \frac{\Omega}{4\pi}$$

Where $\Omega = \frac{A}{d^2}$ is the solid angle under which the detectors intercepts the photons.



Absorption efficiency

Fraction of the photons entering the detector which is actually absorbed



$$\boxed{p(x)dx = \mu \exp(-\mu x)dx}$$

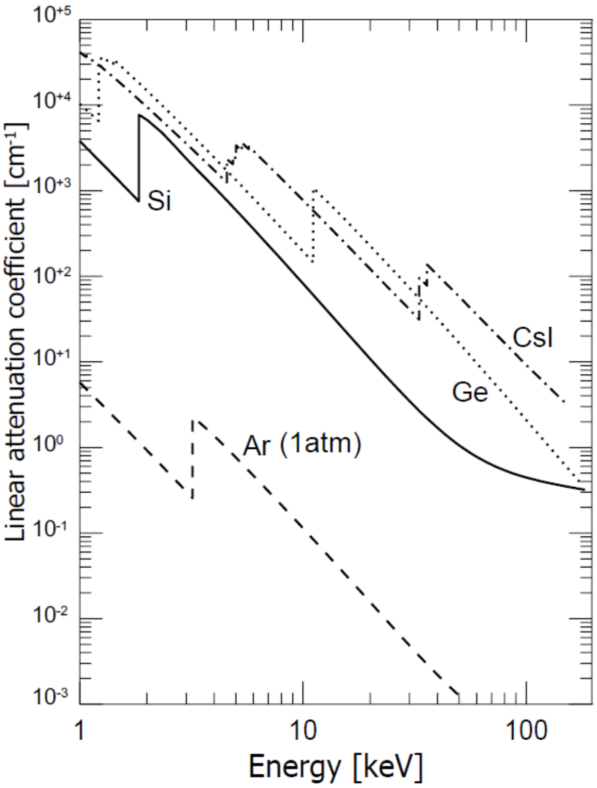
probability that a photon entering into the detector is absorbed in the thickness dx at x

$\Rightarrow$ probability that a photon entering into the detector is absorbed in the thickness t:

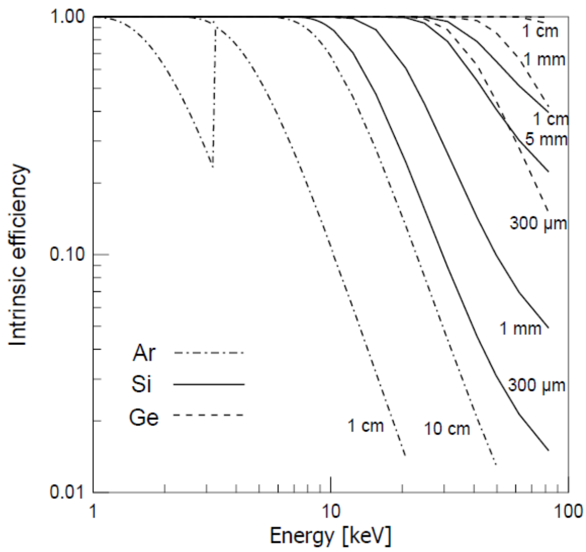
$$\int_0^t p(x)\,dx = \boxed{1 - \exp(-\mu t) = \eta_{abs.}}$$

Note that

- In SPECT the probability above is the probability of the absorption efficiency for the entire measurement
- In PET the probability is equal to the probability a single detector, and since we need to measure form both sides to find both the gamma rays generated the probability must be **squared to obtain the total probability**.



linear attenuation coefficient (μ) for different materials used for detectors



$\eta_{abs.}$ for different thickness

Photopeak efficiency

Fraction of photons that have interacted in the detector and that have released completely their energy. We need to remember that a percentage of the photons will interact with a Compton interaction and the secondary gamma ray may not be absorbed, so event if the photon has been absorbed the data is lost.

To increase efficiency it is necessary to size suitably the detector in order to maximize the probability that also the Compton photon is absorbed in the material.

Energy to electrical charge conversion

After the absorption of the photon in the material the energy absorbed causes the creation of an electron hole pair.
The charge generated is proportional to the photon energy

$$Q = \frac{qE}{\varepsilon}$$

Where ε is a conversion factor which indicates the efficiency of the material in transferring the energy of the photon into the electrons, it indicates how many electron volts are required to create a charge

Argon	$\varepsilon \sim 26$ eV for e-/ion pair
Silicon	$\varepsilon \sim 3.6$ eV for e-/h pair
Se-amorphous	$\varepsilon \sim 20$ eV for e-/h pair
CsI+PMT	$\varepsilon \sim 25$ eV for e-/h pair (ref. scintillator+photodiode: indirect conversion detector)

We can see that silicon has a much higher efficiency than other materials

Notes on materials

Silicon

Is an optimum material but it is efficient only up to $10 - 30keV$ because of limited thickness that can be depleted in practice

Germanium

Germanium has better ε than silicon however it has to be cooled down to reduce the dark current

Gas detectors

Have inherently low efficiency but can be still used for X rays because they can be fabricated in large dimensions even a few tens of centimeters.

Scintillator materials

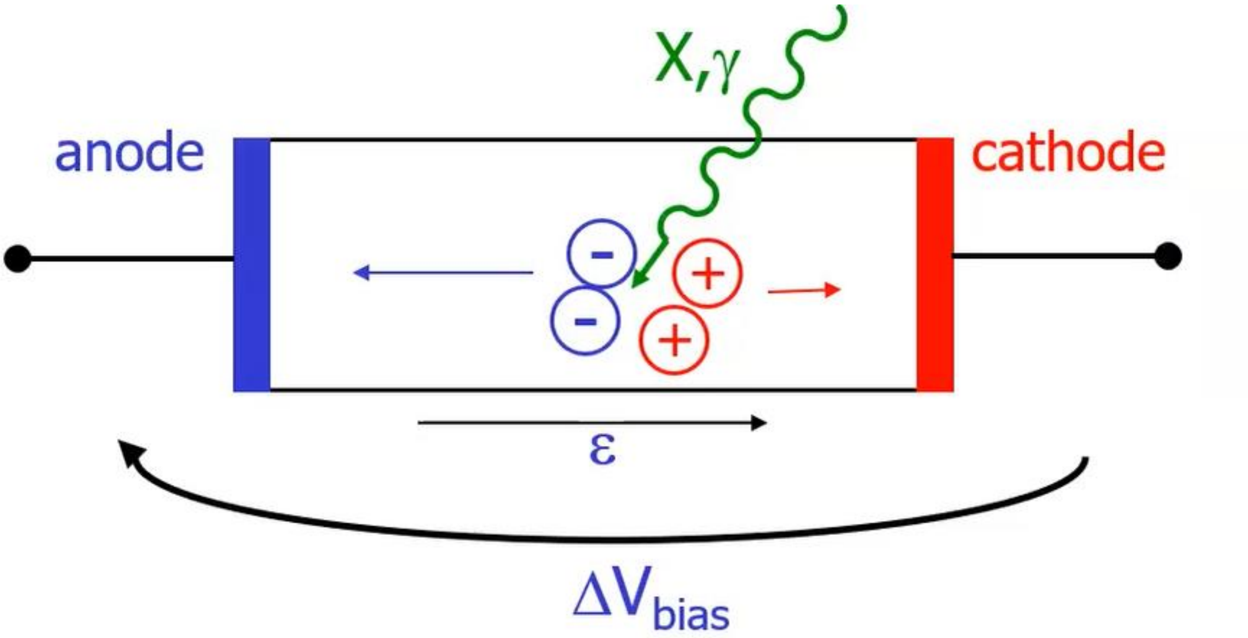
Are not able to create charge by ionization so they are used in indirect conversion thanks to their high efficiency

Other semiconductors

There are semiconductor materials with high Z which are more efficient than silicon they work fine at room temperature because of a large energy gap however they suffer form charge trapping effects.

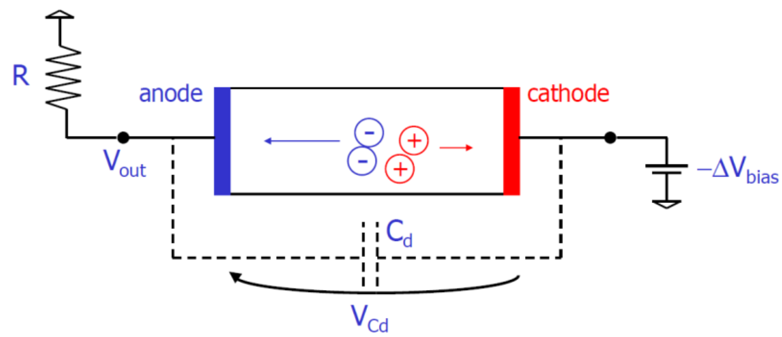
Ionizing detectors

Once we create the charge we need for an electric field to be present across the detector so that it can be collected otherwise it would simply remain there and then eventually decay again



After we need a read out circuit that can sense the amount of charge coming form the detector.

The basic circuit includes a bias generator at the cathode and a resistor at the anode so that we can read the output voltage since the detector will present a parasitic capacitance which will change its voltage as charge accumulates



$$V_{CD} = V_{CD,initial} + \Delta V_{CD}$$

Where $\Delta V_{CD} = -\frac{\Delta Q}{C_d}$

The output voltage will be

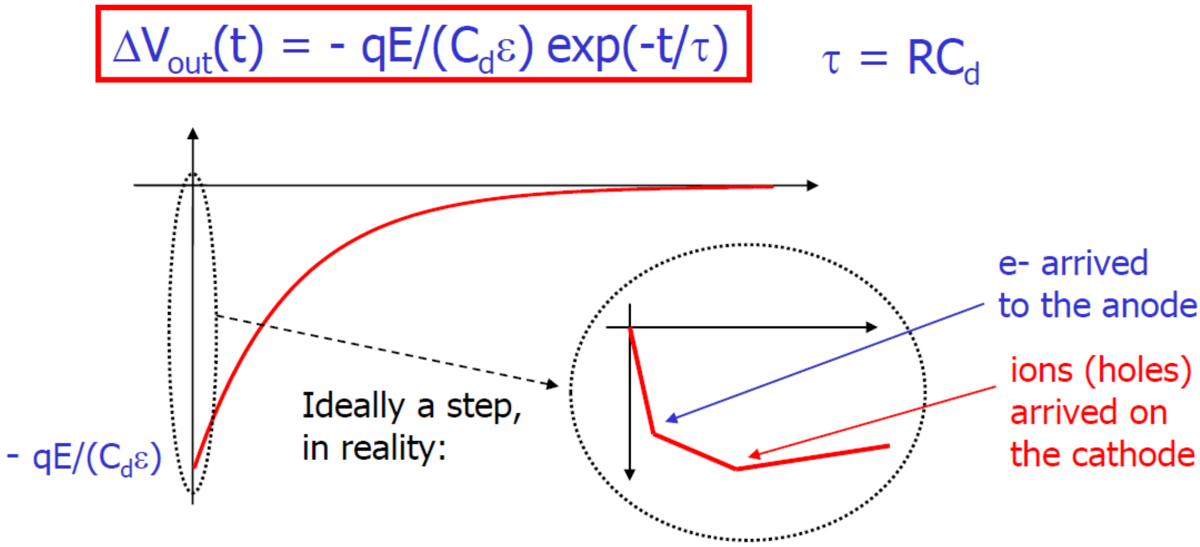
$$\Delta V_{out} = \Delta V_{Cd}$$

Then we can simply substitute with $\Delta Q = \frac{qE}{\varepsilon}$ so we get

$$\boldsymbol{\Delta V_{out} = -\frac{qE}{C_d \varepsilon}}$$

Discharging the charge

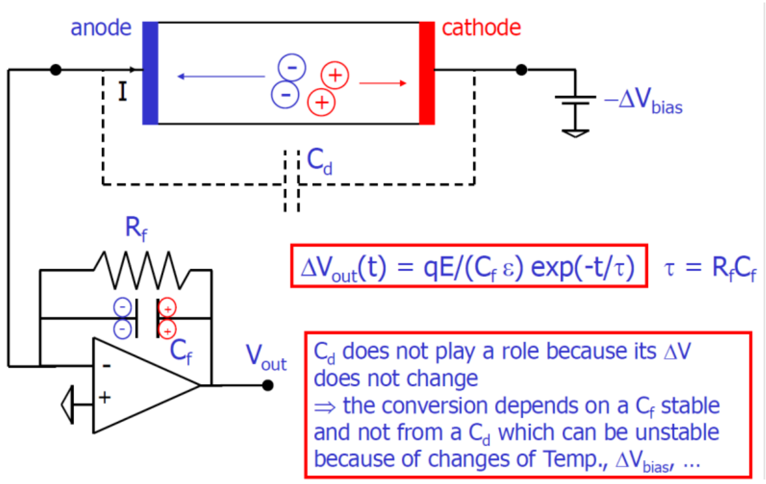
The accumulated charge is then discharged through the resistance R with exponential decau



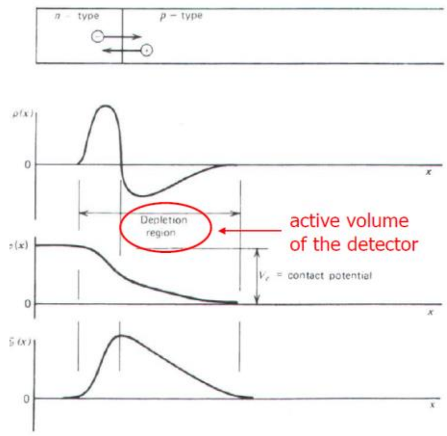
The discharge is needed to restore the detector to equilibrium otherwise the pulses would pile up. Note that the signal front is not an ideal step because of charges travelling time.

Charge preamplifier configuration

In this solution we connect to the output an integrator so that the anode voltage remains constant, the charge generated will not be stored in C_d but in the capacitance C_f. This is important because in some detector technologies the capacitance of the detector is not fixed, this solution allows us to have a more stable circuit



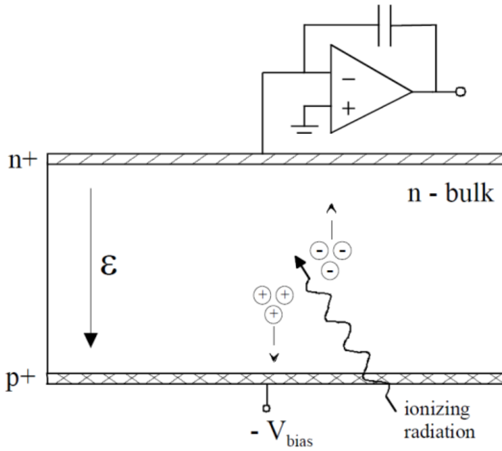
Semiconductor detectors



The basic semiconductor detector is the pn diode, we apply an inverse bias to a pn junction thus creating an area without free carriers.
When a photon frees a new carrier pair inside the depleted region the bias attracts the carriers to the contacts of the pn junction.

Thickness of the depleted region and efficiency

The active volume of the detector is the depletion region to increase detection efficiency we want to increase this areas as much as possible, however there are strong limitations to this introduced by the voltage value we would end up needing.



In the table below we can see how it is important to have high thickness of the active area to assure good efficiency

thickness	efficiency
1 mm	40% at 30keV
	10% at 50keV
1 cm	90% at 30keV
	40% at 80keV

However we can see that the voltage required to create ethe depleted region increases with the square of its thickness

$$\boldsymbol{V_{depl} = x_d^2 \frac{q \cdot N}{2\varepsilon_0 \varepsilon_r}}$$

Where

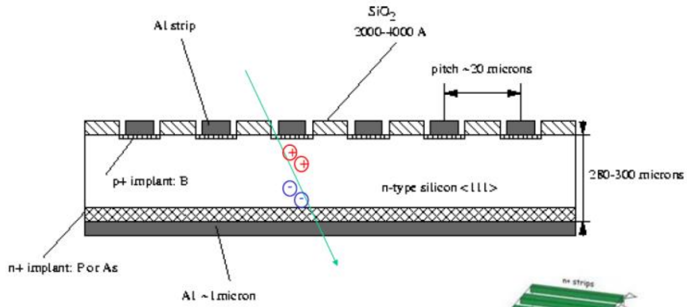
- $\varepsilon_0 = 8{,}8 \cdot 10^{-14} F/cm$
- $\varepsilon_r = 11{,}7$ for silicon
- $N \simeq 10^{12}$ is the dopant concentration

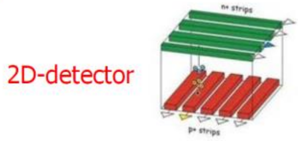
thickness	V_{depl}
100 μm	7.8 V
300 μm	70 V
1 mm	780 V
1 cm	78 kV (!)

We would need a colossal voltage to obtain a $1cm$ depleted region so **the use of pn detectors is limited to energy of a few tens of keV so we are limited to low energy X ray radiography**.

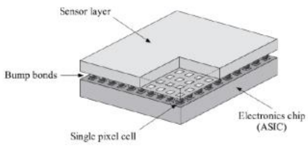
Microstrip detector

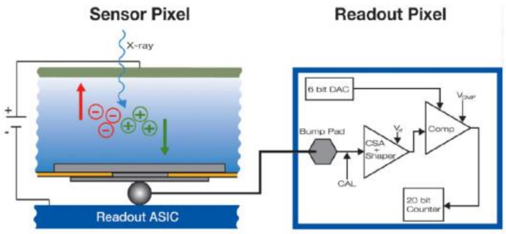
We segment the junction so that we can determine the position of interaction.
With this structure only one couple of the p and n stripes will collect the charge so we know that the interaction has happened at their intersection.





Pixel detector





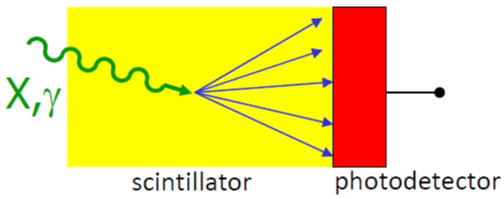
With this structure we segmented only one of the layers but in to directions thus obtaining many independent PN junction so we know directly the interaction position.

This is rather complex to create because we need a huge number of readout channels

Scintillators

These are indirect conversion detectors, here we have

1) A first element which absorbs the radiation and in this case converts it in a photon of visible light
2) A photosensor which will detect the light, from the light quantity we obtain the original signal



Scintillators can be divided in 2 main categories

Gamma camera a pixel

In this case the scintillator is segmented in many different and independent sections

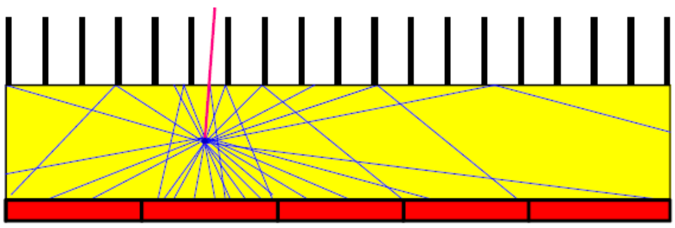
Pro

We have an individual pixel precision so we can easily measure multiple interactions if they occur on separate pixels

Con

We need a high number of photodetectors, this limits our spatial resolution, we have a resolution of a few millimeters.
Creating the small scintillator crystals are expensive to produce

Anger camera



In this case the scintillator is not segmented we have an smaller number of photodetectors and we extract the position of interaction by looking for the peak of detection of the signals analyzed as analog signals.

Pro

We can have a monolithic scintillator which is easier to produce and cheaper since we need less electronic channels.
We use less photo detectors
The resolution is not directly dependent on the number of photodetectors we can obtain a much higher resolution than the width of our photodetector.

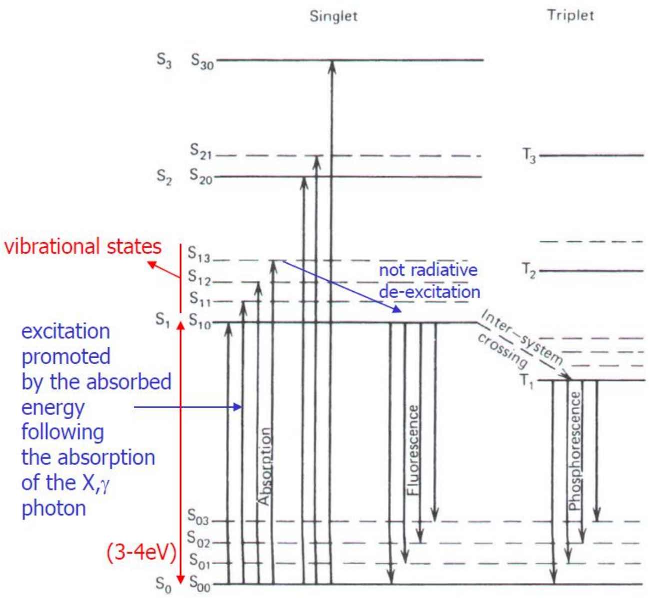
Con

Signals from multiple gamma rays overlap.

Scintillators categories

Organic scintillators

This is a class of scintillators where the scintillation property comes from an organic molecule, and since it is a property of the molecule itself we can obtain solid or liquid scintillators.



Above we have the state graph of an organic scintillation molecule, each state presents other substates which are vibration states.
When a high energy electron generated by the absorption of an X or gamma ray scatters in our medium if it hits a molecule it excites it to the S_1 state: this excitation is radiation less, afterwards we can deexcite by fluorescence and now the visible light photon is released.

The probability for a molecule to de excite is constant so if we start by a series of excited molecules we obtain an exponentially decreasing light emission

$$I = I_0 e^{-\frac{t}{\tau}} \qquad \tau \simeq ns$$

The time constant is the first parameter with which we can describe a scintillator material the other is the conversion efficiency.
It is better to have a **smaller time constant** because it means having a higher peak for the same amount of photons.

Conversion efficiency

Is the fraction of the absorbed energy actually converted in scintillation light, and it depends on the number of emitted photons from their energy.
The efficiency in organic scintillators is low between a few thousands and 70 thousands photons per MeV while in plastic scintillators we have a much higher efficiency.

Examples between 2 types of scintillators

NaI(Tl)

$G \sim 38 \times 10^3$ ph/MeV

$\lambda = 415$nm (well matched with PMT)

$\Rightarrow$ fraction of converted energy $= hc/\lambda \times G \sim 3\text{eV} \times G = 0.11$

CsI(Tl)

$G \sim 65 \times 10^3$ ph/MeV (+71%)

$\lambda = 540$nm (well matched with PD)

$\Rightarrow$ fraction of converted energy $= hc/\lambda \times G \sim 2.3\text{eV} \times G = 0.15$ (+36%)

Light yield

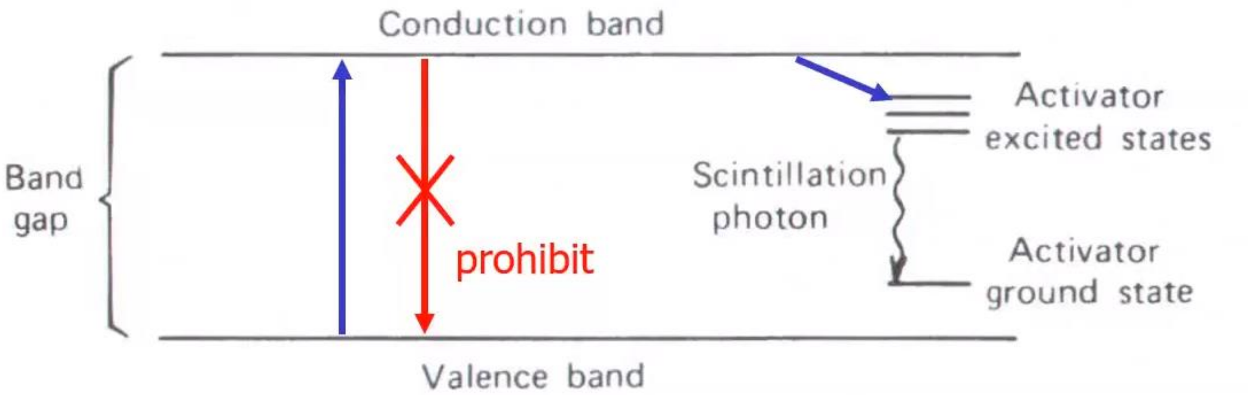
A more relevant measure than the energy conversion efficiency is the **light yield** which gives us the number of photons per MeV of energy provided.

The number of photons emitted is more important than their energy because the generated charge in the photodetector depends on the photon number and not their energy.

It is better to have more low energy photons than a few high energy ones because: as long as the photon has enough energy to generate an electron is enough.

The only factor is that the quantum efficiency of photodetectors changes with frequency so we should try to match the emitted photons wavelength to the optimum one of the photodetector.

Inorganic scintillators



In this kind of scintillators the scintillation is allowed thanks to the presence of **activators**: dopants that when introduced in the crystal structure create intermediate states between the conduction and the valence band of the intrinsic material, now when the electrons from the conduction band de excite rather than going directly from the conduction to the valence and they pass through the activator executed and ground state and in this intermediate passage radiate a photon.

Inorganic scintillators typically have a higher light yield than organic scintillators but they are also slower, conversely organic scintillators are faster but have a lower light yield

Organic scintillators present a decay time in the order of a few ns, while inorganic scintillators operate in the order of tens of nanoseconds to even a couple of microseconds.

Key parameters for scintillators

Table 8.3 Properties of Common Inorganic Scintillators

	Specific Gravity	Wavelength of Max. Emission	Refractive Index	Decay Time (μs)	Abs. Light Yield in Photons/MeV	Relative Pulse Height Using Bialk. PM tube	References
Alkali Halides							
NaI(Tl)	3.67	415	1.85	0.23	38 000	1.00	
CsI(Tl)	4.51	540	1.80	0.68 (64%), 3.34 (36%)	65 000	0.49	78, 90, 91
CsI(Na)	4.51	420	1.84	0.46, 4.18	39 000	1.10	92
Li(Eu)	4.08	470	1.96	1.4	11 000	0.23	
Other Slow Inorganics							
BGO	7.13	480	2.15	0.30	8200	0.13	
$CdWO_4$	7.90	470	2.3	1.1 (40%), 14.5 (60%)	15 000	0.4	98–100
ZnS(Ag) (polycrystalline)	4.09	450	2.36	0.2		1.3[a]	
CaF_2 (Eu)	3.19	435	1.47	0.9	24 000	0.5	
Unactivated Fast Inorganics							
BaF_2 (fast component)	4.89	220		0.0006	1400	na	107–109
BaF_2 (slow component)	4.89	310	1.56	0.63	9500	0.2	107–109
CsI (fast component)	4.51	305		0.002 (35%), 0.02 (65%)	2000	0.05	113–115
CsI (slow component)	4.51	450	1.80	multiple, up to several μs	varies	varies	114, 115
CeF_3	6.16	310, 340	1.68	0.005, 0.027	4400	0.04 to 0.05	76, 116, 117
Cerium-Activated Fast Inorganics							
GSO	6.71	440	1.85	0.056 (90%), 0.4 (10%)	9000	0.2	119–121
YAP	5.37	370	1.95	0.027	18 000	0.45	78, 125
YAG	4.56	550	1.82	0.088 (72%), 0.302 (28%)	17 000	0.5	78, 127
LSO	7.4	420	1.82	0.047	25 000	0.75	130, 131
LuAP	8.4	365	1.94	0.017	17 000	0.3	134, 136, 138
Glass Scintillators							
Ce activated Li glass[b]	2.64	400	1.59	0.05 to 0.1	3500	0.09	77, 145
Tb activated glass[b]	3.03	550	1.5	~3000 to 5000	~50 000	na	145
For comparison, a typical organic (plastic) scintillator:							
NE102A	1.03	423	1.58	0.002	10 000	0.25	

Specific gravity

Basically, the density expressed in grams over cubic centimeters, as we now the efficiency with which a scintillator captures a gamma ray is connected to the density

Wavelength of max emission

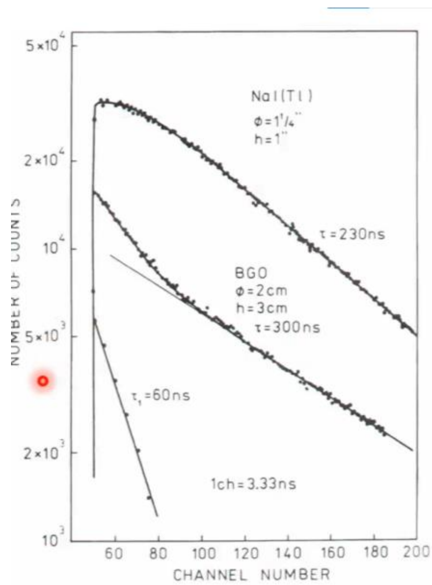
The wavelength for which we have the most emission of photons after fluorescence

Decay time: read slide below

Some scintillators have 2 or more components of scintillations with different time constants and relative weight in the scintillation process, this time constant as well as the weights $(a_1, a_2, \ldots)$ change with temperature.

$$I = I_0 \left(a_1 e^{-\frac{t}{\tau_1}} + a_2 e^{-\frac{t}{\tau_2}} \right)$$

Additionally, there are 2 processes fluorescence and phosphorescence the second is much slower and will generate an almost constant emission of light



Value of the parameters for different implementations

Specific gravity

For PET is more important to have a high density material than for SPECT because

1) The photons have a higher energy
2) We measure in paired so the detection efficiency of the detector is squared

Decay TIME

Decay time is not critical for SPECT we just need it to be fast enough to not limit our count rate, we do not measure the arrival time.
In PET we are interested in the exact arrival time of the gamma ray to reconstruct the constancies and then possibly use the time-of-flight technique

Additionally, we want to avoid slow decay time to avoid overlapping of the light emission of 2 different signals as this would create a pile up similarly as what we have seen with a direct detector (again since SPECT has a low event count we do not worry much for this in SPECT)

Light yield

Light yield is more important for SPECT because since the photons have a smaller energy we want more of them to be sure to obtain enough output signal for better operation.

Internal radioactivity

In nature elements present different isotopes, the crystal can contain isotopes which naturally decay releasing photons without a stimulation form outside. This adds noise to the measurement.
This is particularly significant for SPECT, in PET we can distinguish between paired events and background events both because of timing and energy levels which for background events will be a continuous spectrum.

Photodetectors

PMT

A PTM is a vacuum tube containing a photocathode and a series of dynodes.

- Photocathode: a thin material which when stimulated by a photon releases an electrode
- Dynodes: a series of mirrors

For this to work we need to excite the electron not to the conduction band but to the vacuum level, this requires energy and for the electron to be able to physically reach the vacuum for this reason the photocathode is very thin otherwise the electron might remain blocked inside the cathode.

To the drive the released atom towards the atom we need for a strong electric field to be present between cathode and anode.

As the electron proceeds it hits the different dynodes which each biased with a progressively higher voltage. As the electron moves it acquires kinetic energy so when it hits the dynode it transfer this energy which will be enough to cause the release of more than an electron.
By cascading this process we can obtain gains in the order of 10^7.

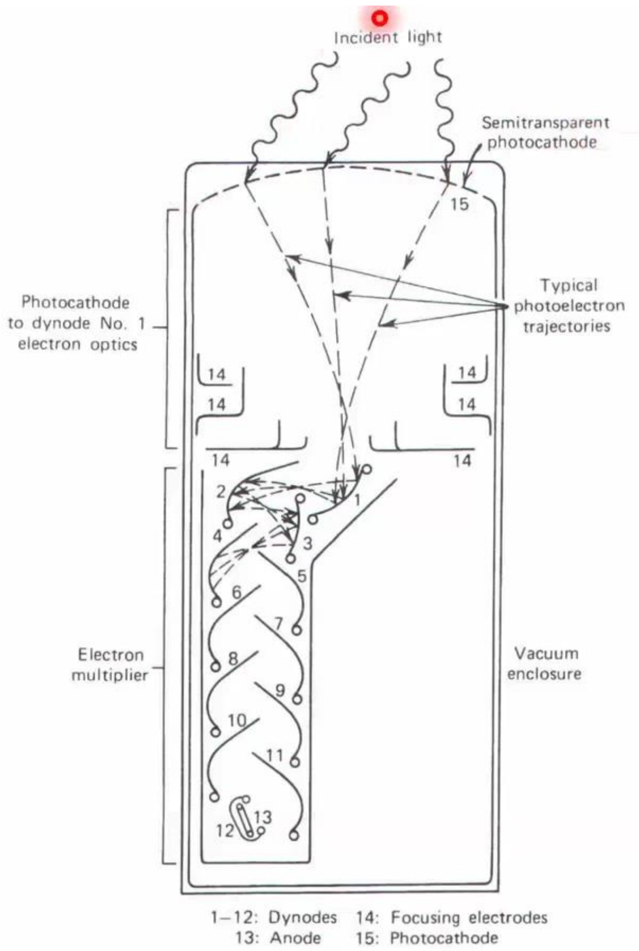
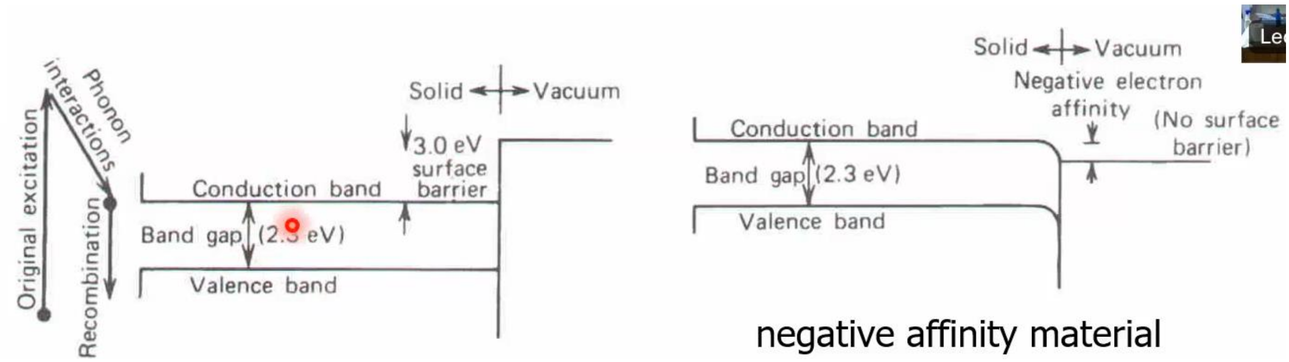


Figure 9.1 Basic elements of a PM tube. (From Ref. 1.)

The advantage of this high gain is that now the noise of the following component of the circuit is negligible as when we compare it to the signal we need to divide it by the gain of the PMT, instead we are just left with the primary statistical fluctuation introduced by the PMT which is independent form the PMT gain.

$$\sigma_{primary}{}^2 = n_{primary} \rightarrow n_{primary}/\sigma_{primary} = \sqrt{n_{primary}}$$

Photocathode



We want to make it as easy as possible for an electron to reach the vacuum level and thus the material.
To do so we need to create a pn junction to create a depleted region and thus band banding, this allows us to obtain a **negative affinity material** where the conduction band has an higher energy level than the vacuum level.

An example of negative affinity material is gallium phosphide $GaP(p)$ with high p doping around 10^{19}, then we connect the gallium phosphide with cesium Cs, the resulting band graph is shown above.
We have now that the conduction band of the GaP is higher than the vacuum level of the cesium so if an electron form the GaP conduction band reaches the Cesium it escapes into the vacuum.

Quantum efficiency

In a photocathode we can expect at best a quantum efficiency of 30%: the number of photoelectrons emitted by the photocathode over the number of incoming photons.

Additionally, this efficiency depends on the wavelength of incoming electrons, and it drops as the wavelength increases: for this reason we are interested in having high energy photons coming from the scintillator.

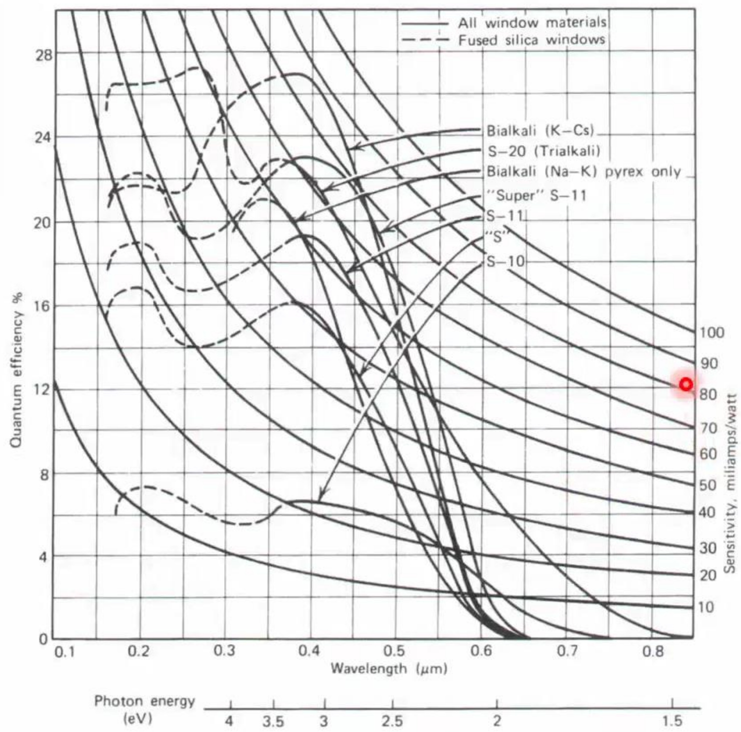


Figure 9.2 The spectral sensitivity of a number of photocathode materials of interest in PM tubes. The use of silica or quartz windows is necessary to extend the response into the ultraviolet region. (Courtesy of EMI GENCOM Inc., Plainview, NY.)

Multiplication factor at the dynodes

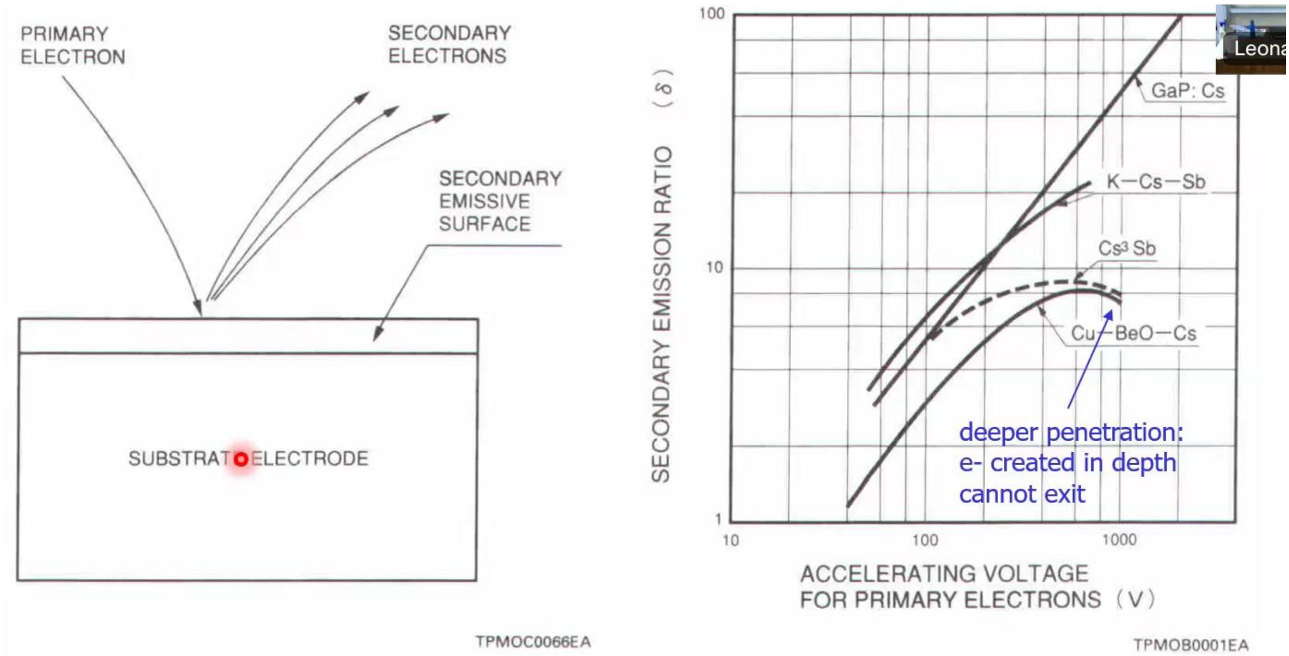


Figure 2-6: Secondary emission of dynode **Figure 2-7: Secondary emission ratio**

We have approximately a linear increase for low accelerating voltages then the value starts to plateau because while initially more energy corresponds to more electron freed, after a certain threshold the initial electrode starts to travel further and further inside the material until eventually the electrons generated to deep will have de exited when they reach the sourface.

Total multiplication

$\boxed{\text{total gain} = \alpha\, \delta^N}$

α= fraction of collected photoelectrons

δ= multiplication factor of single dynode

N= number of multiplication stages

$\delta = 5 \Rightarrow G = 5^{10} \sim 10^7$ with 10 stages

$\delta = 55$ (NEA) $\Rightarrow G = 55^4 \sim 10^7$ with 4 stages

multiplication statistics

δ fluctuates statistically event-by-event

variance related to 1 dynode (Poisson): $(\sigma/\delta)^2 = \delta/\delta^2 = 1/\delta$

variance of total G: $\boxed{(\sigma_G/\delta^N)^2 = 1/\delta + 1/\delta^2 + 1/\delta^3 + \ldots + 1/\delta^N \sim 1/(\delta - 1)}$

(note: variance dominated by the fluctuations of the first dynode for $\delta >> 1$)

Problem

The multiplication is also a statistical process, so cascading different multiplications adds noise to the final value. As said above the variance is dominated by the fluctuations of the first diodone if $\delta \gg 1$

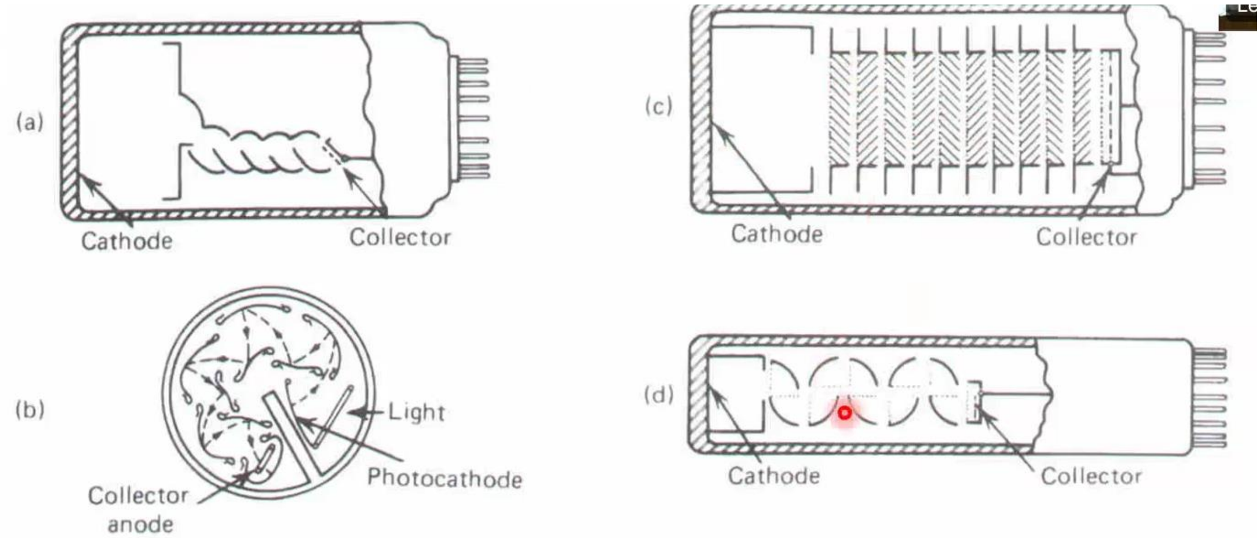
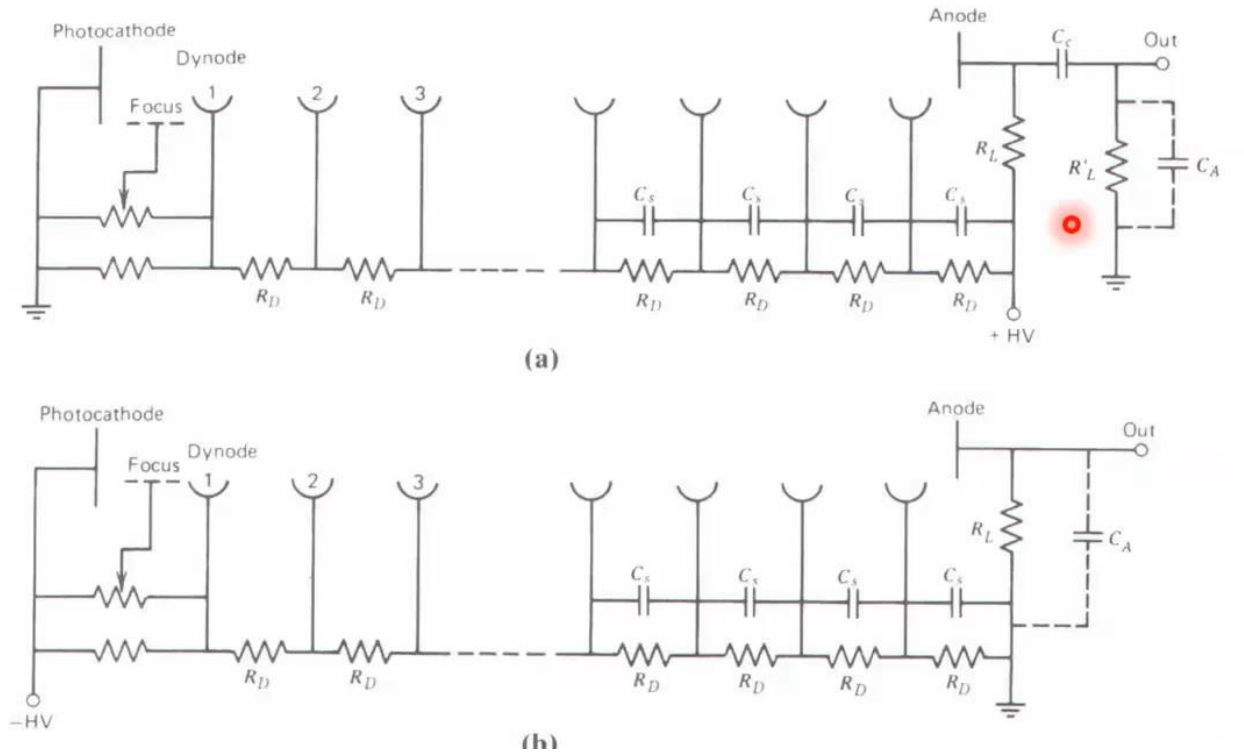


Figure 9.7 Configurations of some common types of PM tubes. (*a*) Focused linear structure. (*b*) Circular grid. (*c*) Venetian blind. (*d*) Box-and-grid. (Courtesy of EMI GENCOM Inc., Plainview, NY.)

Bias circuit for the photomultiplier



Same operation principle (voltage divider with resistors) the difference is that in one case we connect the cathode to ground so it is safe to touch however this means that the anode must be decoupled to read the signal.

In the second case we can read directly form the anode but the cathode is now at a high voltage.

Note that the capacitors are there to provide electrons for the signal generation while avoiding a flow of current in the bias resistors.

PMT signal readout

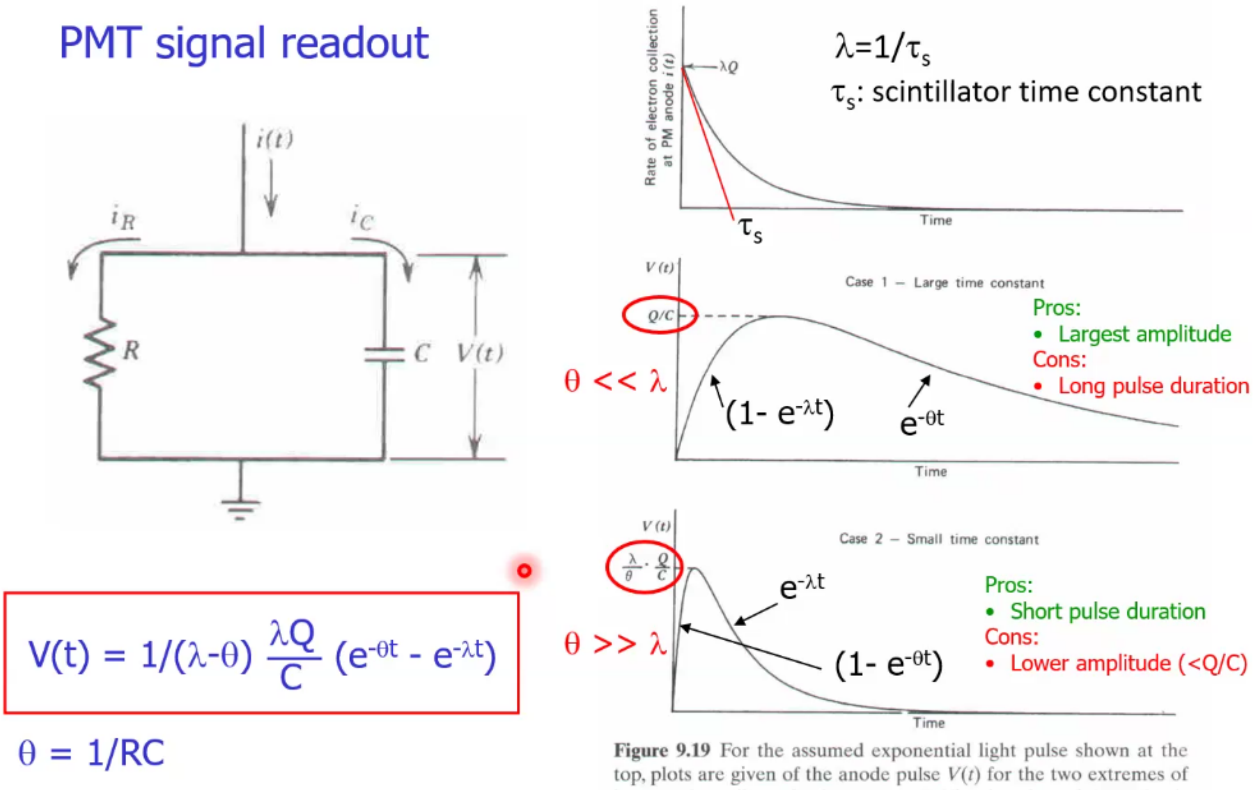


Figure 9.19 For the assumed exponential light pulse shown at the top, plots are given of the anode pulse $V(t)$ for the two extremes of large and small anode time constant. The duration of the pulse is shorter for Case 2, but the maximum amplitude is much smaller.

1) We start by assuming that the current signal has the same shape as the scintillator signal so we have a instant increase followed by an exponential decay.

The easiest signal to read out this is the parallel of a resistor and a capacitance, the charge accumulates first in the capacitance and then discharges over time through the resistance:

- With a high resistance we obtain a larger amplitude signal but also a longer pulse so read out will be slower, instead if we select a smaller R now the time constant is much shorter than the decay time of the scintillator now the peak is lower but the pulse is much shorter.

In the first case since the time constant is larger than the scintillator time constant we can assume that we integrated the entire signal so the output voltage value will be equal to

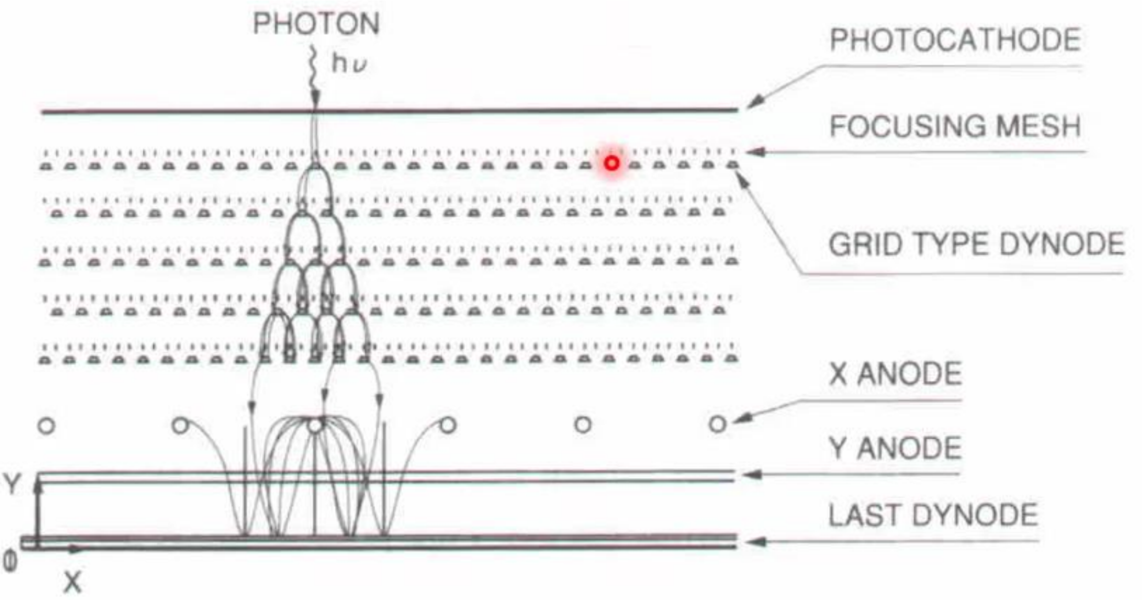
$$V_{peak} = \frac{Q}{C}$$

In the second case instead the capacitance starts to discharge before all of the charge pulse from the PMT has arrive so the output voltage will be approximately if we indicate with

- $\lambda = \frac{1}{\tau_S}$ the inverse of the scintillator time constant
- $\phi = \frac{1}{RC}$ the inverse of the RC circuit time constant

$$V_{peak} = \frac{\lambda}{\phi} \cdot \frac{Q}{C}$$

Position sensitive PMTs



We segment each dynode in many electrodes: the resulting dynodes is made by a series of electrodes at the same potential.
As the charge cascades downwards the resulting value will have a higher multiplication factor in the point where the initial photon hit.

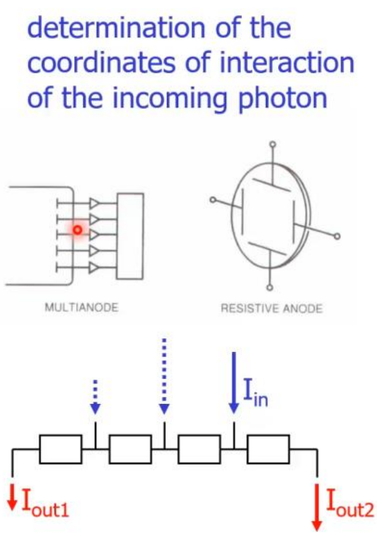
At the anode we read the current using 2 grids in perpendicular directions this allows us to detect the point where we have the most charge.

Timing
At the end we also have a last dynode used to sum together all the signals, the resulting pulse is used to determine the timing.

Read out and estimation of the position

with the easiest approach we simply read the output of each wire, this requires sensing form 2N wires, we maintain the maximum amount of information but the final circuit is very complex and expensive.

An alternative solution is that of utilizing resistances in series to delete the current signal in 2 directions depending on the positions, then we can sum the 2 currents with opposite signs and divide the result to obtain the value of the position from which the most current comes.



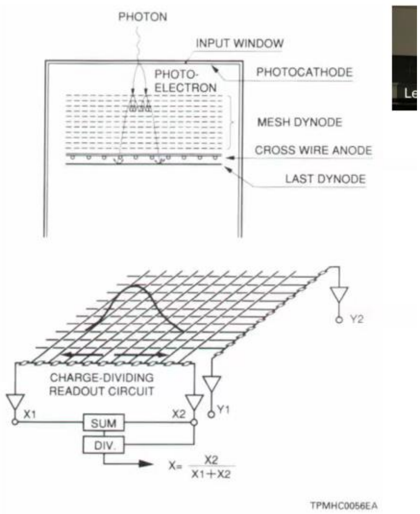


Figure 6-6: Center-of-gravity detection method for position-sensitive photomultiplier tubes using cross-wire anodes

Photodiode

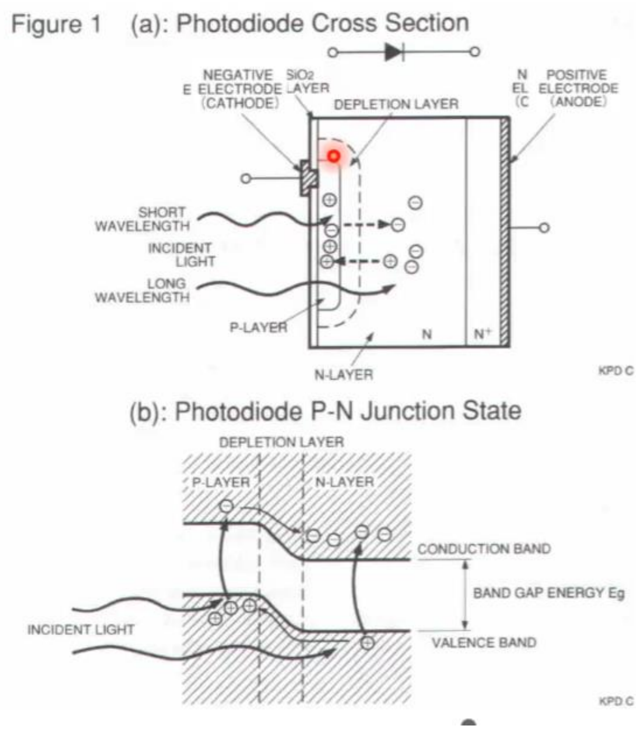
They are an alternative to PMT

Advantages

- Higher quantum efficiency above 80% rather than at barely 30%
- Higher compactness
- Lower bias voltage required
- Insensitive to magnetic fields
- Lower cost

Disadvantages

- Since we have no multiplication the amplifier noise will no longer negligible
- We have a significant dark current at room temperature



Operation

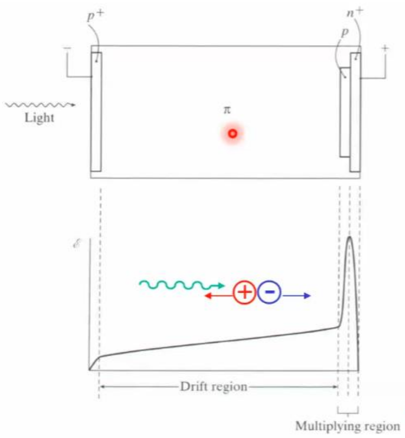
We have that higher energy photons will interact closer to the surface of the device.
As long as the interaction happens in the depleted region we will still have the generation of a hole electron pair which will then generate a current.
Again we do not have multiplication so one photon will generate at best 1 electron.

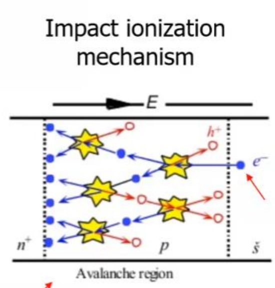
Important

We have transfer only if the photon interacts in the depletion region

Avalanche photodiodes APD

We still have a PN junction, we have at the input a p^+ contact with negative bias, then a drift region of almost intrinsic silicon with only slight p doping, then a junction with p doping and a n^+ doping.





Operation

By having the different doping concentrations we also modulate the voltage in each area obtaining first a drift region with small electric field in which we create our hole electron pair with the photon and then a smaller region with much higher electric field (required because of the higher doping) where we can expect **ionization by impact** as now the electrons will reach a high enough kinetic energy between each collision to be able to free another electron and a hole after they collide with another atom.

With this solution we can obtain a gain of about 1000.

It is possible also for holes to be multiply however their multiplication factor will be lower than that of electrons since, holes have a smaller mobility than electrons.

Problems

We do not multiply only the charge of the signal we also increase ethe dark current.
Let's now consider all the multiplication factors for the different current sources

What is the variance of the output signal

The output signal is equal to

output signal $$N_{Sout} = N_S M_{(e-)}$$ N_S: primary e- generated
$M_{(e-)}$: multiplication coeff. for e-

If the multiplication process would be deterministic than the variance of the output signal would be multiplied by the same factor of the signal, this means that the overall signal to noise ratio would remain exactly the same.
Unfortunately in reality we need to consider an excess noise factor due to the fact that the multiplication is itself a statistical process

variance of the output signal $$\sigma^2_{Sout} = N_S M_{(e-)}^2 F_{(e-)}$$ $F_{(e-)}$: excess noise factor (multiplication due to e-)

This means hat the final signal to noise ration is worsened with respect to the original and it is equal to

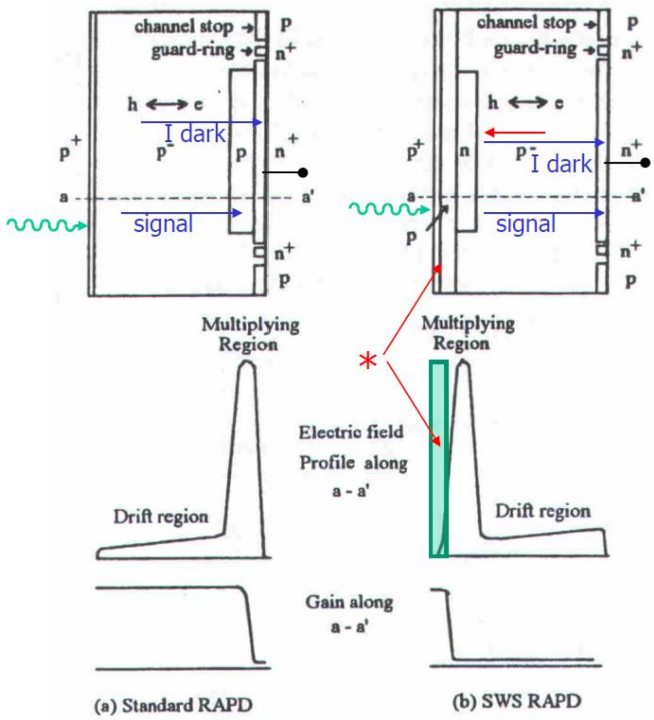
$$\boxed{S/N = N_{Sout}/\sigma_{Sout} = \sqrt{(N_s/F_{(e-)})}}$$

$$F_{(e-)} \sim 2\text{-}3$$

Additionally the noise spectral of the dark current will also increase

$$2qI_D M_{(e-)}^2 F_{(e-)}$$ this noise component also worsens by a factor $F_{(e-)}$ (with respect to $2qI_D M_{(e-)}^2$)

Reach through ADP



We want to reduce the contribution of the dark current shot noise, to do this we want to make it so that the dark current and the signal are generated in different parts of the junction so that they are multiplied by a different factor.

We want to make it so that we multiply
- The electrons of the signals
- The holes of the dark current

Because the multiplication factor of the elections is higher than that of the holes.

What we do is multiply the structure so that the high electric field region is not multiplying all of the charge generated in the drift region.
What we do is move the implant of n^+ doping near the entrance window rather than on the opposite side.
We have that in reverse bias
- The holes flow towards the n doped region
- The electrons flow towards the n doped region

With the configuration described above we still have a deletion region at the entrance so the photon still have time to interact and free electrons which will then pass through the high electric field region be multiplied an collected at the n electrode.
Instead the electron of the dark current whose generation is uniform in the volume will for the vast majority just travel towards the electrode without passing for the high electric field region so they will not be multiplied.

Only the holes of the dark current which travel in the opposite direction will be multiplied.

We obtain a circuit that makes it so that
- The signal is multiplied by the multiplication factor of the electrons
- The dark current is multiplied by the multiplication factor of the holes

$$N_{Sout} = N_S M_{(e-)}$$

$$\sigma^2_{I\ dark\ out} = \sigma^2_{I\ dark(h)} M_{(h)}^2 F_{(h)}$$

$$\boxed{S/N_{dark} = N_s \times M_{(e-)}/M_{(h)} \,/ (\sigma_{I\ dark(h)} \sqrt{F_{(h)}})}$$

$$\boxed{\begin{array}{c} M_{(e-)} > M_{(h)} \\ F_{(h)} > F_{(e-)} \end{array}}$$

it is higher than the previous thanks to M(e-)/M(h)

Important

This works only if the photons interact in the first layer

Summary

PMT
- gain
- reliable/mature fabrication and several available topologies
- low quantum efficiency
- costs, size
- high bias voltage, sensitivity to magnetic fields

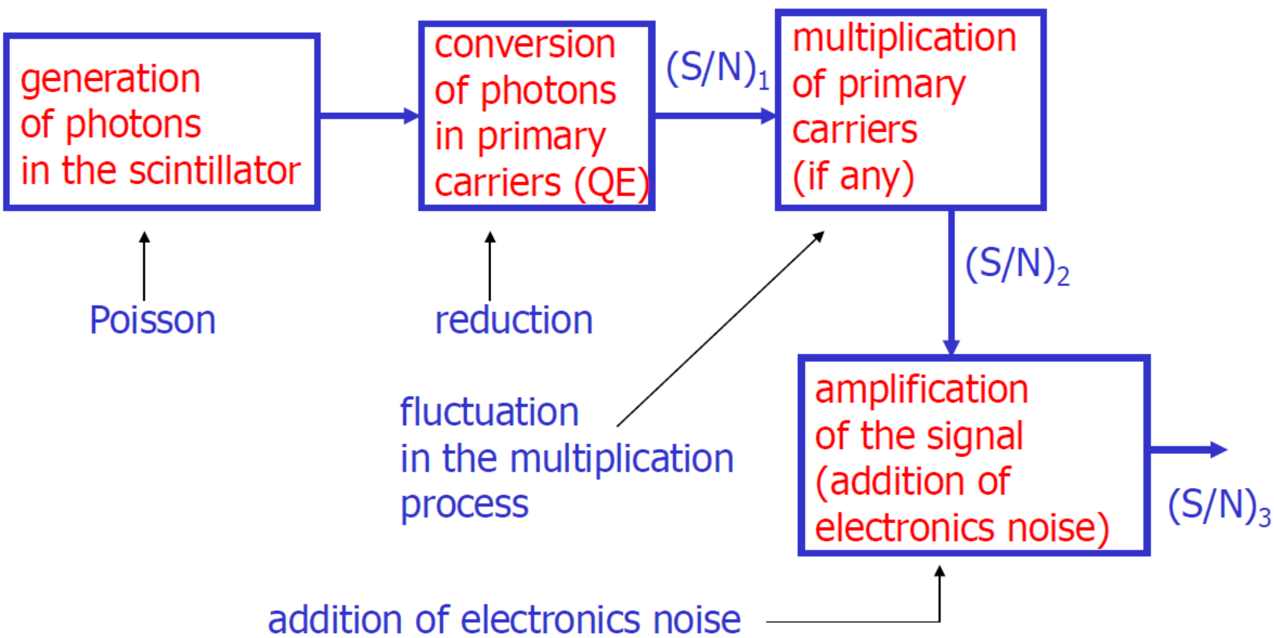
PD
- high quantum efficiency
- small size, low bias voltage, low cost
- fabrication of arrays of uniform units
- lack of gain (amplifier noise plays a role)

APD
- gain
- high quantum efficiency
- small size
- gain fluctuation, high bias voltage
- difficulty to fabricate arrays of uniform units

Parte 2 medical imaging

Comparison between photodetectors

Below we have a scheme of all the different noise sources in the detection chain



- When the photons are generated in the scintillator we have a Poisson distribution of the amount of photons an thus a noise is naturally present
- When the photons are converted into charge carriers only a fraction will be converted depending on the detector quantum efficiency, this means that the noise distribution remains the same but at the output will be equal to the square root of the number of photoelectrons and not that of the photons themselves

- We might then have a multiplication of the carriers, when present this multiplication adds an excess noise factor F
- Finally we have the addition of noise coming from the electronics, so we can define the ENC as the amount of input charge which gives us a signal to noise ratio of 1 at the output of the circuit
$$Q = ENC \Rightarrow S/N = 1$$

Noise summary

- First term due to Poisson statistics
$$SNR_1 = \sqrt{N_S}$$
where N_S is the number of carriers generated after photodetection
- Second term
$$SNR_1 = \sqrt{\frac{N_S}{F}}$$
Where F is the worsening factor of the statistics of carriers due to multiplication

Finally the total signal to noise ratio will be
$$SNR_3 = \frac{N_S M}{\sqrt{N_S M^2 F + ENC^2}} = \frac{N_S}{\sqrt{N_S F + \frac{ENC^2}{M^2}}}$$
Finally we can write
$$\boldsymbol{SNR_3 = \frac{1}{\sqrt{\frac{F}{N_S} + \frac{ENC^2}{M^2}\frac{1}{N_S}}}}$$

Analysis

Multiplication of the photodetectors

- It improves the component related to electronic noise (as we can see M is at the denominator)
- It worsen the statistic component as we introduce the excess noise factor F

Electronic vs statistic noise

Electronic component noise dominates in case of low signals since it has a proportionality to $\frac{1}{N_S^2}$ which increases faster than $\frac{1}{N_S}$ when N_s decreases.

This affects which detector we select since it means that when we have low signal we prefer a detector with multiplications since the electronic noise will be dominant.

APD vs PMT

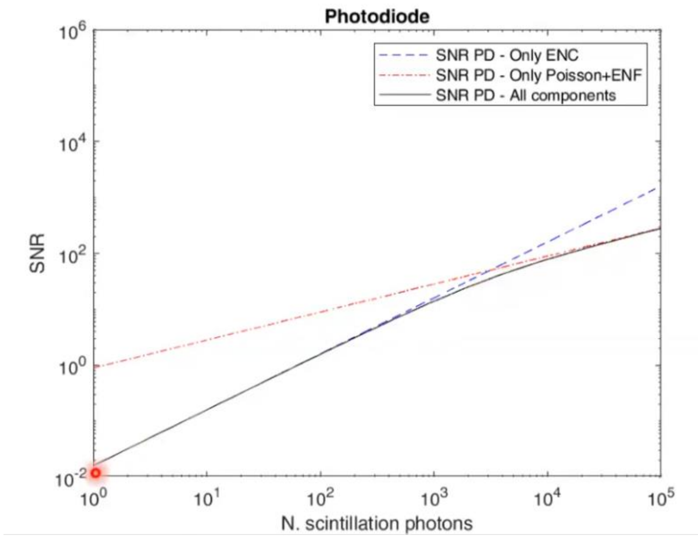
We have that APD typically provides a better statistic component because N_s is higher because of the higher QE.

PMT instead provide a much larger M and a better ENC since their dark current is lower, so these will present better electronic noise components.

Photodiode example

Let's consider

- $QE = 80\%$
- $F = 1$
- $ENC = 50$
- $M = 1$ (no multiplication)



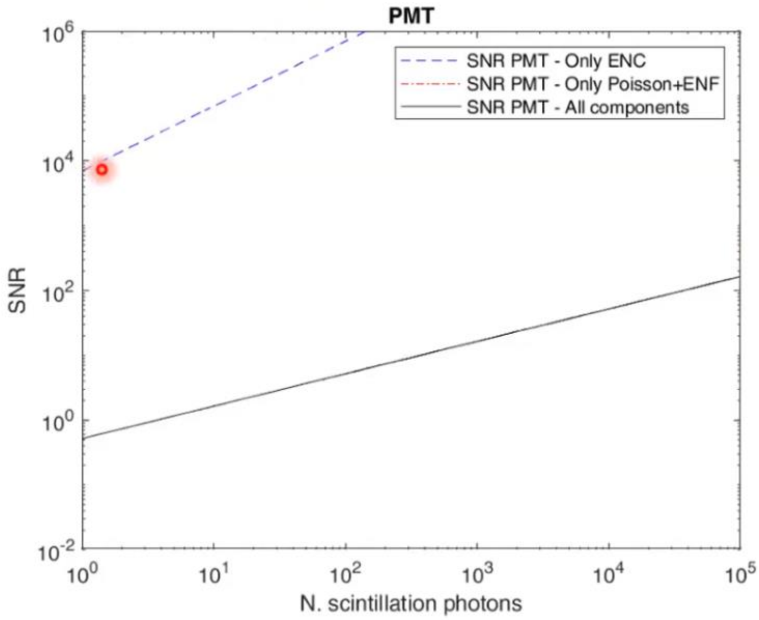
We can see that the Poisson noise is higher for lower signal and then is surpassed by the electronic noise as the number of photons increases.

The total SNR is determined by the lowest component.

PMT example

Let's consider

- $QE = 35\%$
- $F = 1{,}3$
- $ENC = 50$
- $M = 10^6$



As we can see the electronic noise is always negligible with respect to the statistical noise

Comparison

For a small number of photons the PMT has better performances because the electronic noise is dominant while statistical noise becomes dominant once the photons increase.

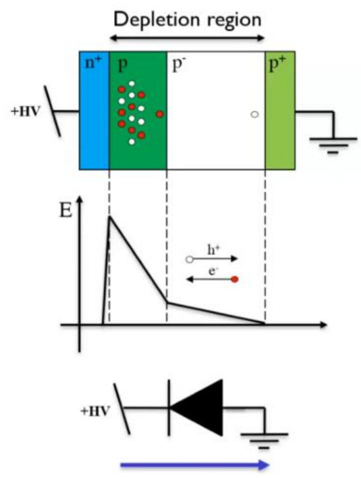
Note

The SiPM has a similar curve to the PMT but slightly better (higher SNR)

SPAD and SiPM

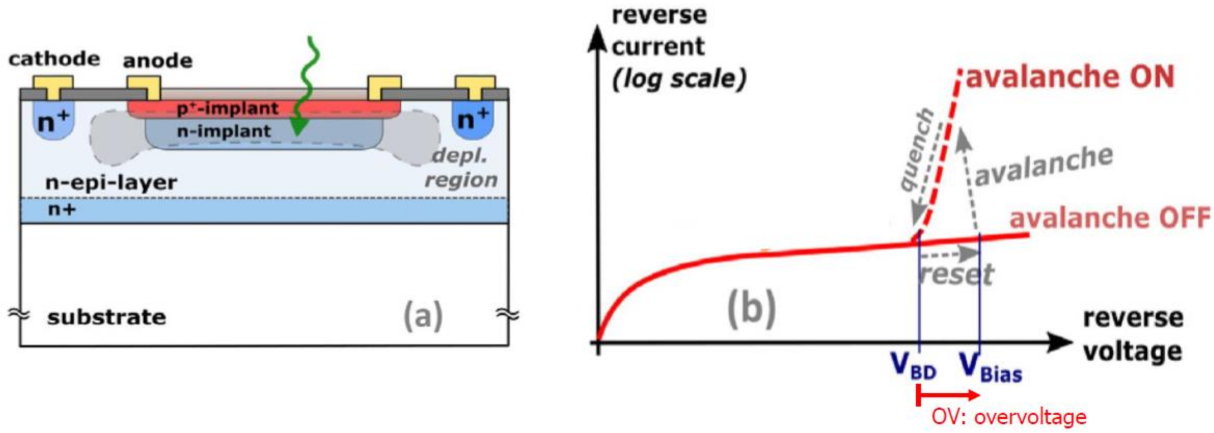
We mentioned that APD which are pn junctions with a high electric field region which multiplies our charge thus reducing the effect of electronic noise, they however have some limitaitons

- The gain is low when compared to PMT
- The variance of the gain is high



A SPAD (single photon APD) in an APD in which the high electric field region reaches the breakdown voltage.
At this point the multiplication becomes divergent an thus a single photon releasing an electron will cause a huge amount of charge to flow to the output

Operation



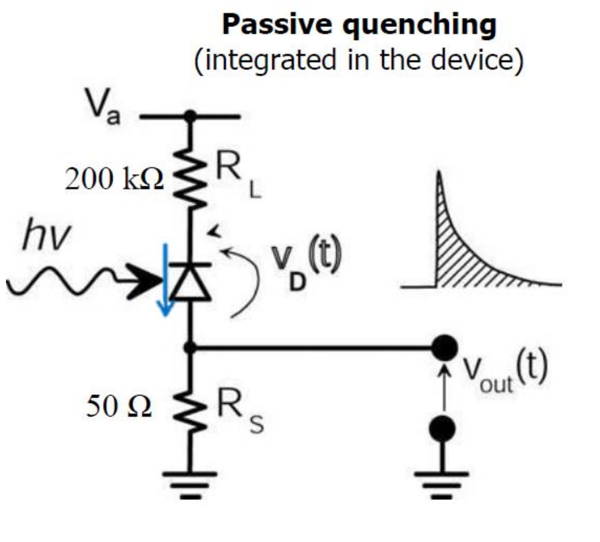
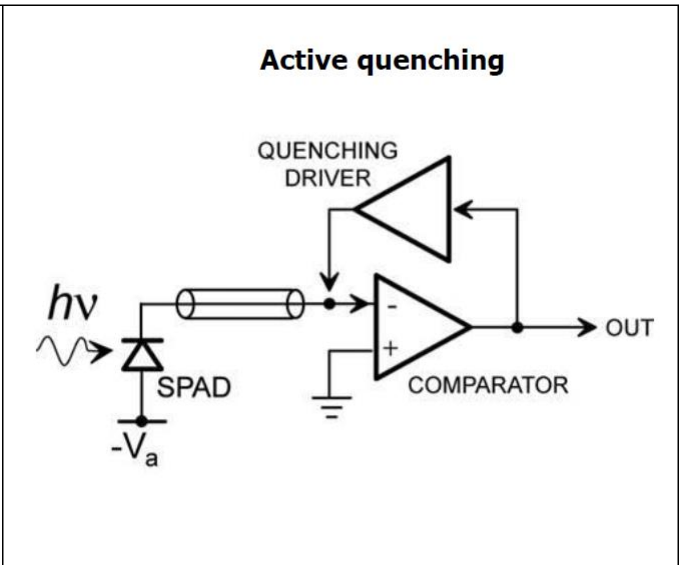
The structure is the same of an APD, what we do is increase the reverse bias until we reach the breakdown voltage leaving the linear region in which typical APD operate.

Metastable region

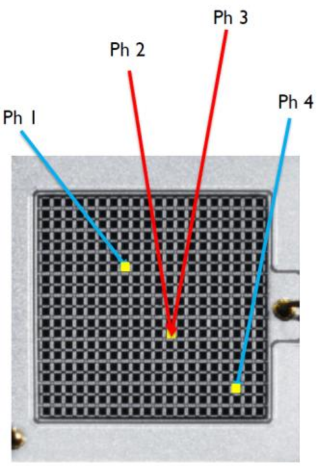
The reason why we are able to cross the breakdown voltage without having instantly a charge avalanche is because we are in a metastable region until a charge is able to move and start the avalanche, after that point we have a current that would theoretically continue to flow until the **avalanche is quenched by lowering the voltage below the breakdown voltage**. The circuit that does this is called quenching circuit.
To detect a new photon we need to restore the bias voltage to above the breakdown.

Quenching the avalanche

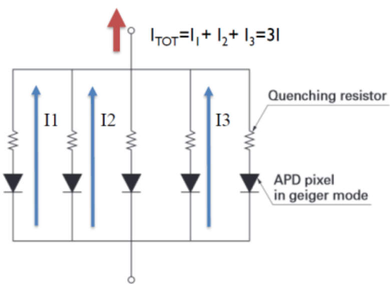
Quenching can be done actively or passively

	
In this solution the current signal itself leads to a drop in the voltage of the bias resistor so the bias voltage lowers automatically. This allows us to have • Cheaper solution • Less active area However we have • Low counting rate $100kcps$ • Worse photon timing $> 100ps$	In this solution once we detect the pulse a circuit actively lowers the voltage. This allows us to have • High count rate $1MCps$ • Good photon timing $< 100ps$ However we have • Higher cost • More area required

SiPM principle

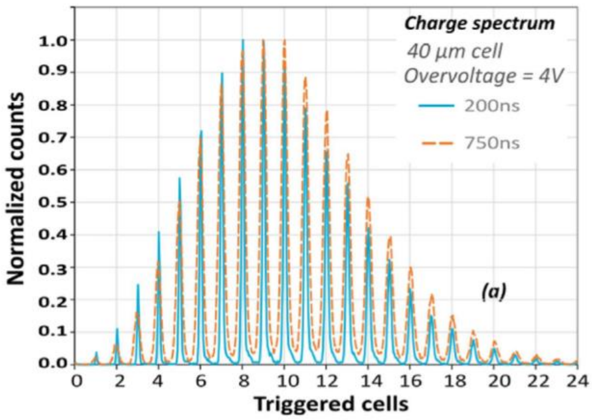


We create a 2d array of SPAD cells connected in parallel with quench resistors (active quenching is too expensive).
The total signal is proportional to the number of fired cells, we can measure the output current to retrieve the number of photon absorbed.



Note

Photons interacting on the same cell are counted as a single hit.



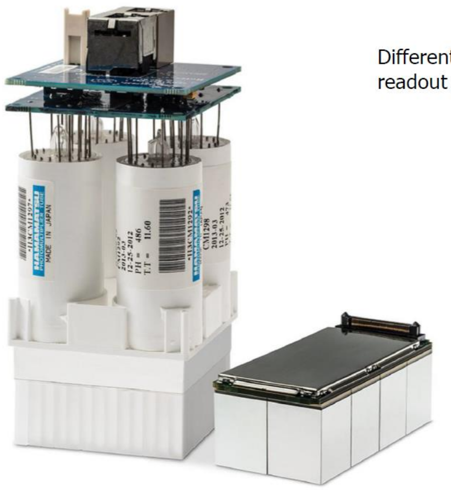
On the right we have the plot of the amount of charge created for a given amount of triggered SPADs.
Note that there will be some variance of the amount of charge released by the SPAD for each trigger event, for this reason the lines are not perfect however they are still spaced enough that we can distinguish between different number of SPADs being triggered.

Trade off between SiPM size and cell dimension
If we increase the SPID size but maintain the same cell dimension (size of the SPAD) we can measure more photons.
If we maintain the same SPID size but decrease the cell dimension, then we have more cells for the same area so we are now able to measure a more intense light
SPAD cell size varies form $15\mu m \times 15\mu m$ to $100\mu m \times 100\mu m$.

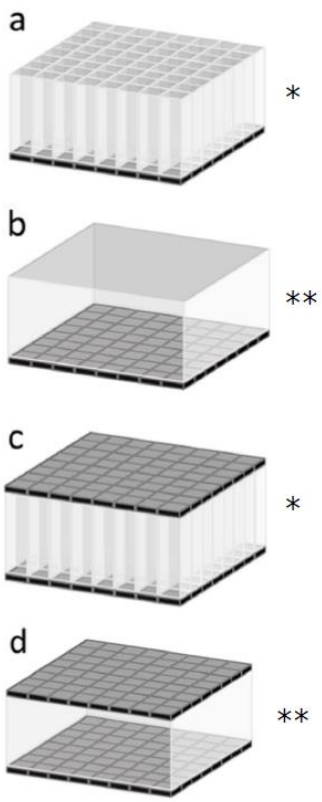
SiPM matrices for imaging applications
We create a matrix of SiPM sensors so that we can also distinguish the interaction point of the photon since otherwise in the SPID all position information are lost.

Advantages of SiPM
- They are very compact each SiPM array is just a few centimeters
- It is possible to connect them in a compact manner this reducing the dead area

SiPMs in medical imaging
In the image we can see the size different between a scanner based on PMT and one based on SiPM, as we can see the size reduction is significant.



Different readout



We can have different configurations:
In case a we have a pixelated configuration where the scintillator is divided and a single SiPM is used to read each pixel.

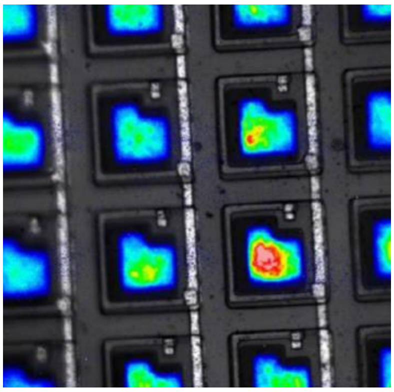
In case b we have a monolithic scintillator and we measure with the same logic as the angler camera, utilizing an analog analysis of the output to recover the interaction position with higher resolution than the pixels.

Configurations c and d are more compels and not currently used and allow to read the signal from 2 different sides, this allows us to read the light more quickly thus increasing the time resolution and also to measure the proportion of light going upward and downward, this allows us to deduce the interaction point inside the crystal.
In these configurations one of the detectors will be in between the crystal and the source of gamma rays.

SiPM and magnetic fields
An important element that contributes to the popularity of SiPM is that their operation is not affected by magnetic fields unlike PMTs, so it is possible to create a single machine capable of executing both MRI and SPECT.

Photodetection efficiency

In a SiPM it is more difficult to define the quantum efficiency with respect to the case of a PMT or a APD. We instead utilize the photodetection efficiency PDEW which is determined by 3 factors



The geometrical efficiency or fill factor FF

The ratio between the active area and total device area, we have that each SPAD cell is surrounded by a dead region determined by the guard ring and the structure to prevent optical cross talk, the FF depends on the cell size, a normal value could be 45%

The quantum efficiency QE

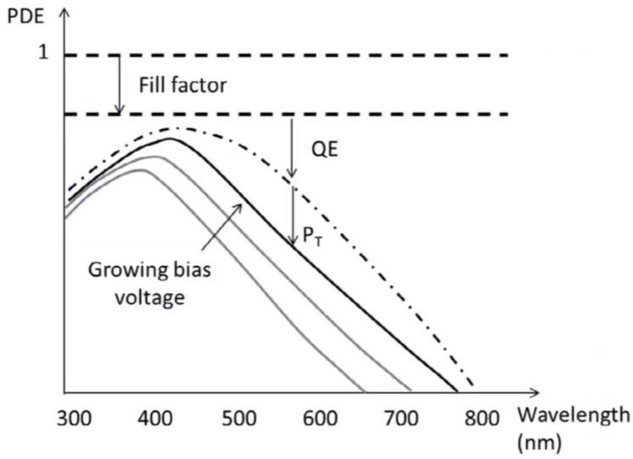
This is the probability that an impinging photon creates a primary electron hole pair in the active volume, i

The turn on probability P_T

The probability for an electron to start an avalanche breakdown, not it depends n wavelength because different wavelengths will interact ad different depths inside the SPAD

$$PDE = FF \cdot QE \cdot P_T$$

Example of PDE

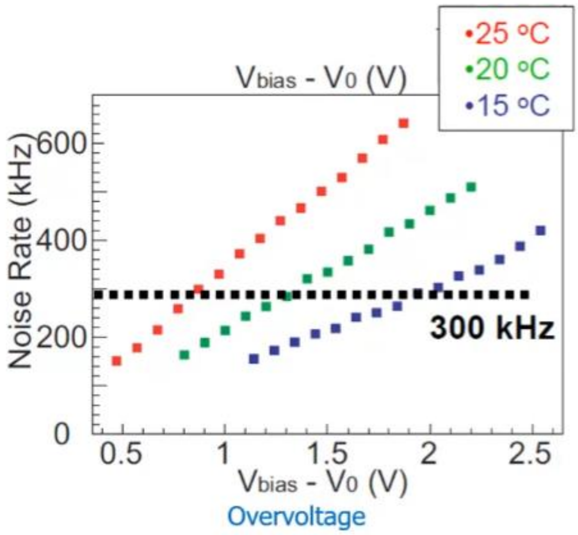


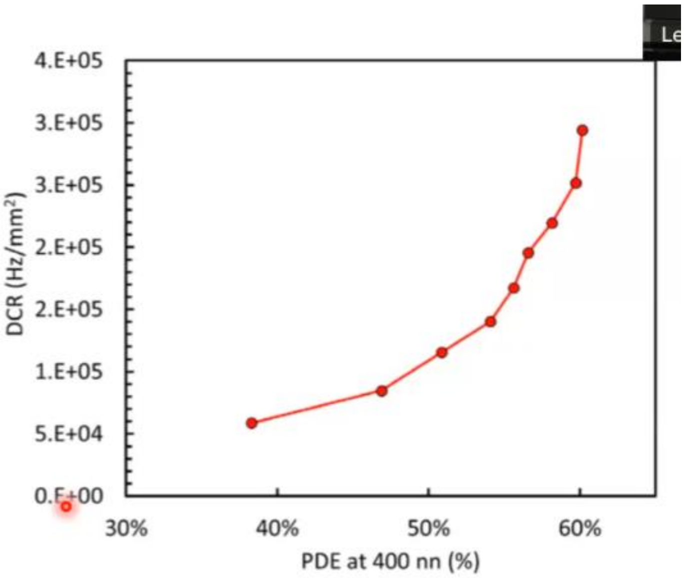
In the plot we can see the different contributions:

1) **The fil factor removes a fixed percentage at all frequencies**
2) **The QE depends on the wavelength**.
   We see that longer wavelength tend to interact deeper in the silicon thus they may go past the depletion region and be lost.
3) **Trigger probability is dependent on the reverse bias** that we apply, a larger bias implies a larger probability.
   It does depend on the wavelength as different wavelengths interact at different depths thus causing electrons/holes pair to travel for different distances.
   Holes have a smaller trigger probability than electrons so the interaction points makes it so that the holes travel for longer than the electrons the probability will be smaller than the opposite.

Dark count rate DCR

When we have a PN junction we have that one electron may promote form valance to conduction band just for thermal excitation, another effects which may cause dark current in SPADs are tunneling (favoured by the high reverse voltage).
The dark current is the current we would have in the sensor even without light.

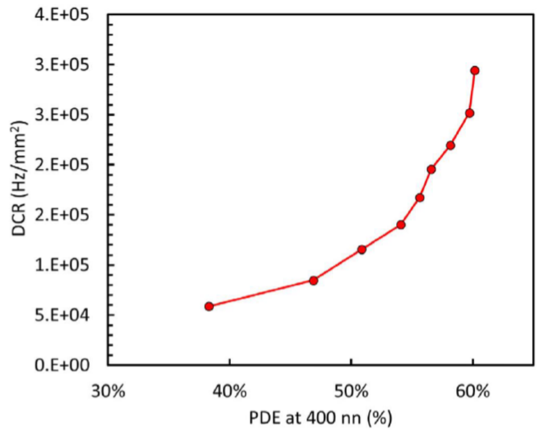




In SiPMs we indicate dark current noise not as ampere but as dark count rate, we measure the number of SPADs triggering every second for every unit of area.
The typical order of magnitude of a dark count rate is rather large hundreds of thousands.

Dependances on the different parameters

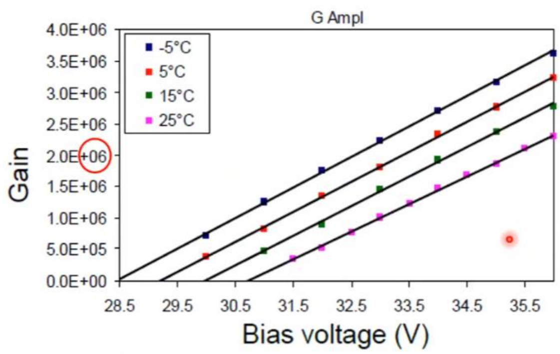
The DCR increases with the overvoltage and decreases if the temperature decreases.
We can see from the graph on the right that we can balance the PDE and the DCR so that we give up a small amount of PDE while reducing greatly the DCR.
We find the optimum point experimentally.



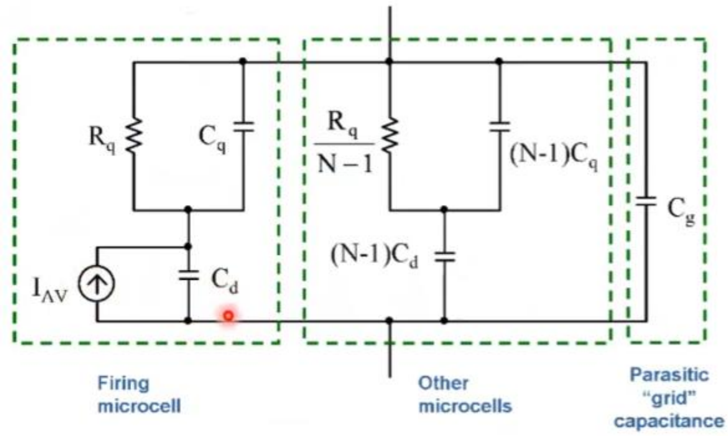
SPAD gain

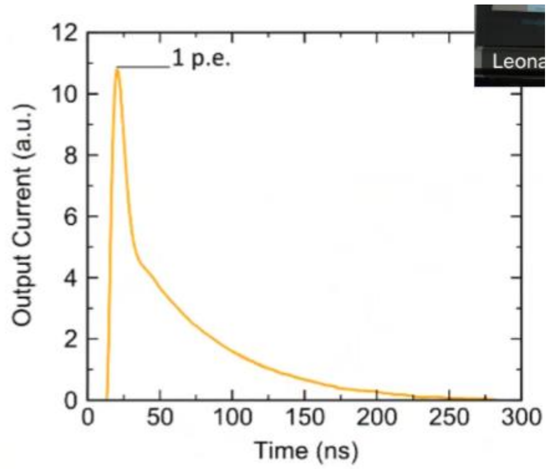
The gain of a SPAD corresponds to the amount of charge provided by a single avalanche.

The multiplication factor M is proportional to the applied voltage and increases at lower temperature because now electrons and holes have an higher mobility since they can move for a longer distance before hitting an atom and thus impact ionization I smore effective.
A higher M allows us to reduce noise since we increase the signal.



SiPM signal shape





The real output current is complex because many SPADs are connected in parallel the model includes

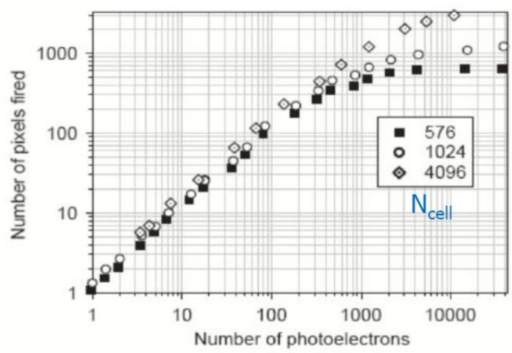
- The equivalent model of the fired cell
- The equivalent circuit of the other microcells
- The total parasitic capacitance associated to the large routing interconnections among all microcells
- The front-end electronics input impedance

The output shows a fast rise time and a slow exponential decay which may present multiple time constants.

If we have multiple SPAD firing at the same time the curve will have the same shape but a proportionally greater amplitude.

Dynamic range

As we know 2 photons interacting on the same SPAD will be read as a single hit, the probability of this happening increases as the light intensity increases eventually, we stop having a linear relationship between the signal and the light intensity because we loose a progressively more significant part of the light signal.



The number of fired cells can be expressed as a function of the number of cells and the number of photons

$$N_{fired} = N_{cell}\left(1 - e^{-\frac{N_{ph} \cdot PDE}{N_{Cell}}}\right)$$

If the signal intensity increases then we need to either

- Increase the size of the SiPM: this solution is not always possible as sometimes the dimension of the scintillator crystal is fixed
- Decrease the size of the microcell (SPAD), this however decreases efficiency as the FF worsens

Noise in SiPM

There a re 3 main sources in SiPM

Dark count

Cells triggered by electrons released because of thermal energy or tunneling effects cause a noise of single cell pulses like the one indicated with X.

Single cell dark count can be recognized because they have the same amplitude since they are caused by a single electron, this can also be used to calibrate the detector: if we measure that for the dark count we have a pulse of for example $47mV$ in the amplifier we know that this is the value corresponding to a single photoelectron and thus use this as base for extracting the number of photoelectrons from a generic signal.

After pulse

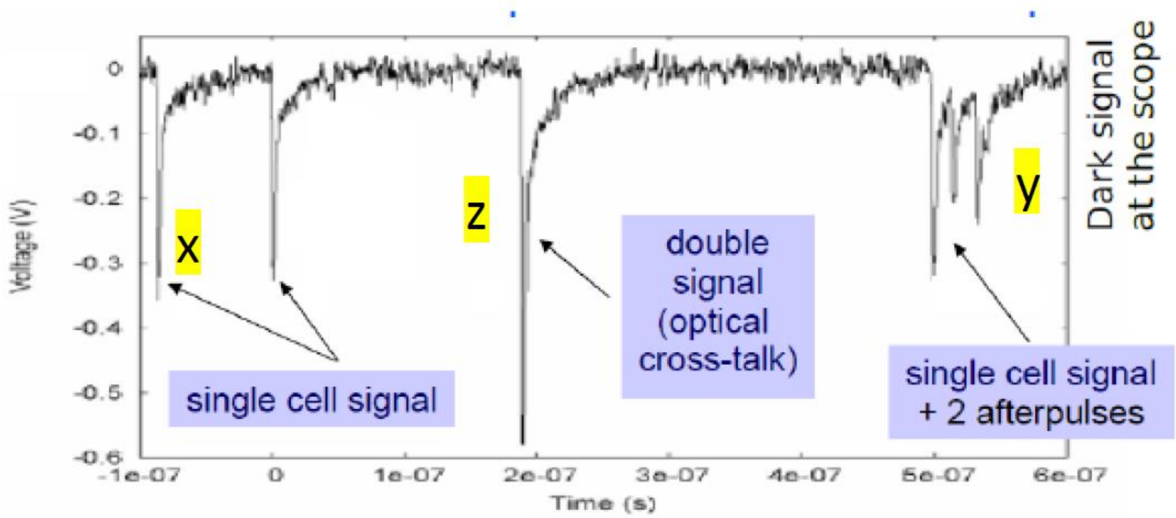
These are carriers created during the avalanche and trapped, if they are not released until the avalanche has already been quenched then they may create a new trigger event and thus a second fake pulse.

In the graph they are indicated with y.

Cross talk

This occurs during the ionization, some of the impacts may cause the promotion of electrons rather than freeing them, so when they relax a photon could be released.

These photons become problematic in the case when they are able to reach the next pixel over and start another avalanche, this causes multiple cells to be fired and it is catastrophic. Case Z.



Energy resolution in scintillation detection

The energy resolution is defined as the full width at half maximum of the energy distribution divided by the peak of the energy distribution

To compute it we use this formula

$$\frac{\Delta E_\gamma}{E_\gamma} = 2.355\frac{\sigma_{N_{pe,out}}}{N_{pe,out}} = 2.355\sqrt{\underbrace{\upsilon(E_{int})}_{\text{Intrinsic resolution}} + \underbrace{\frac{ENF}{N_{pe}}}_{\text{Poisson resolution}} + \underbrace{\left(\frac{ENC_{TOT}}{N_{pe}M}\right)^2}_{\text{Electronic noise resolution}}}$$

$$N_{pe} = N_{ph}\eta = E_\gamma Y\eta$$

We can see that we have identified 3 components determining the energy resolution and a conversion term between sigma and noise

Intrinsic resolution

This is the intrinsic resolution of the scintillator crystal, each scintillator has an average number of photons emitted for a certain incoming energy, this is typically quoted in photons over electron volt, typical value 25000ph/eV.

This number will also be subjected to the Poisson statistic so we have a variance equal to the square root of the number of photons.

Poison resolution

This is the contribution of the Posson statistic connected to the generated by the SiPM

Electronic noise resolution

Contribution of the electronic noise

Singal to noise ratio

The signal to noise ratio formula

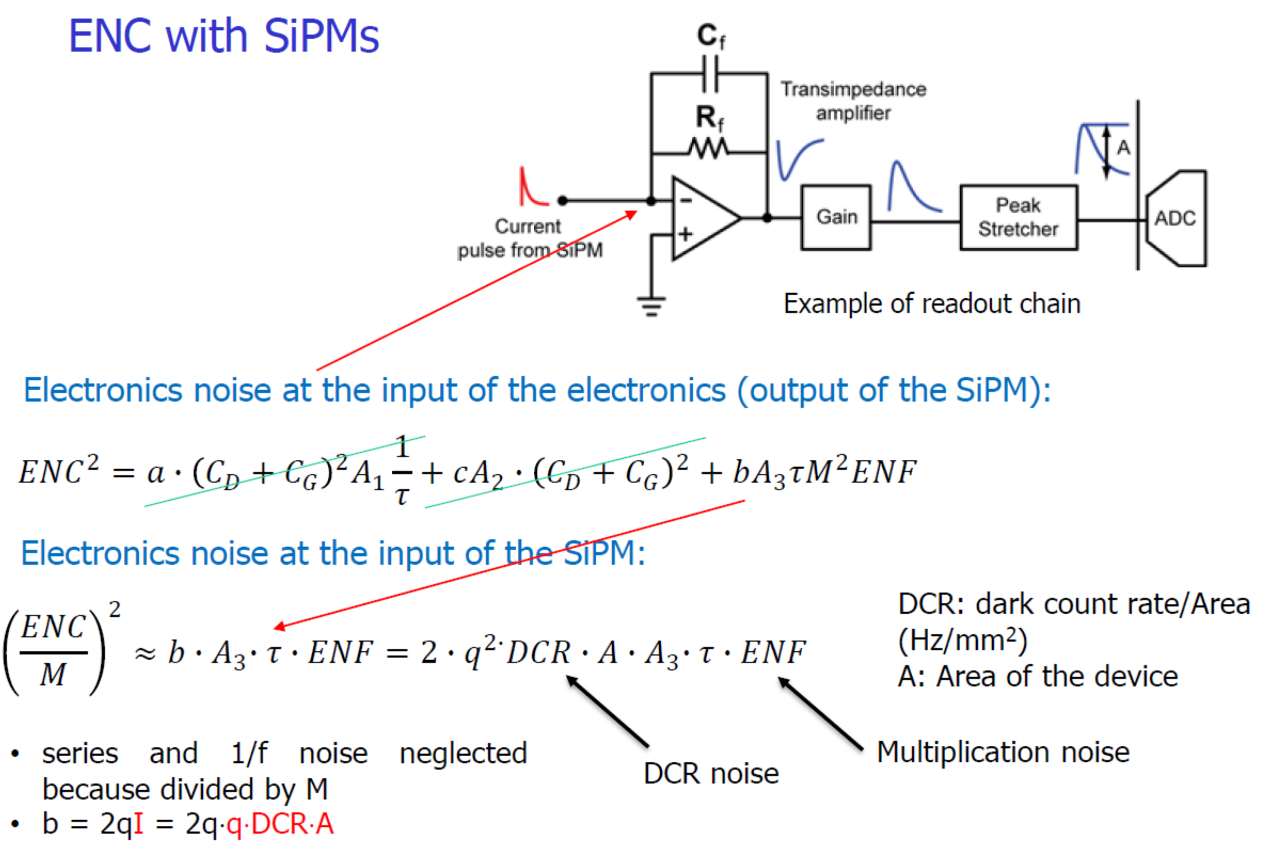
$$SNR = \frac{N_S \cdot M}{\sqrt{N_S M^2 F + ENC^2}} = \frac{N_S}{\sqrt{N_S F + \frac{ENC^2}{M^2}}}$$

At the numerator we have the signal expressed as number of photoelectrons

At the denominator we have 2 terms

- The signal multiplied by the excess noise factor to indicate the Poisson spread
- The electronic noise (ENC^2) divided by the gain squared

ENC with SiPM



We have shown in the previous formula that the electronic noise of the SiPM is divided by the gain, which is very large in the order of 10^6, this however does not mean that we can just neglect the electronic noise because while the series noise and the $\frac{1}{f}$ noise will become negligible the parallel noise (given by the dark current) will remain significant since it is multiplied by the gain so the term simplifies and we are left with the remaining noise

Electronics noise at the input of the ~~SiPM~~:

$$\left(\frac{ENC}{M}\right)^2 \approx b \cdot A_3 \cdot \tau \cdot ENF = 2 \cdot q^2 \cdot DCR \cdot A \cdot A_3 \cdot \tau \cdot ENF$$

DCR: dark count rate/Area (Hz/mm²)
A: Area of the device

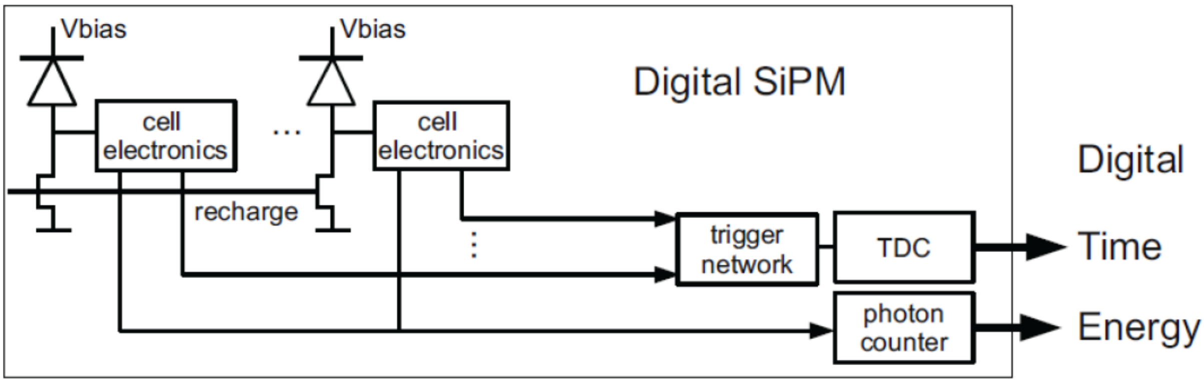
DCR noise

Multiplication noise

- series and 1/f noise neglected because divided by M
- b = 2qI = 2q·q·DCR·A

The dark count rate is the only noise relevant when we make a measurement of energy resolution.

Digital SiPMs



In this kind of SiPM each SPAD has its own readout circuit, basically, a discriminator which detects when the SPAD has been triggered.
This solution requires more electronic and is much more expensive, additionally the SPADs tend to be noisier because we need to create not only them but also the readout circuit so the production process is not fully optimized (higher dark count), however it has a series of advantages

Easier photon counting

Counting the photons is easier since now each interaction will have a distinct signal so we just need to sum the total

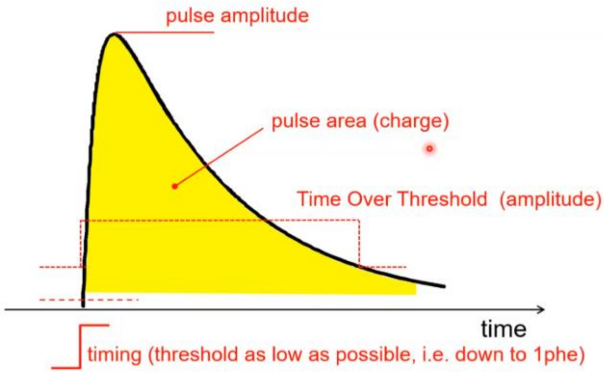
Easier timing

We have a time stamp for each individual event without needing pass the cumulative signal in a common discriminator.

Possibility to switch off Hot SPADs

Not all SPADs have the same DCR, typically in a SiPM the dark current is provided mostly by a few (order of tens) SPADs with defects in their structure which end up firing all the time
The digital SiPM solution allows us to both identify these hot SPADs and also disable them by turning off their readout circuits.

Measurement on a SiPM signal



There are 2 main terms that we are interested in measuring the amplitude of the pulse and its timing.

Measure of the amplitude

This allows us to determine the number of events occurring.

Measure of the peak of the pulse

To measure it we need a very quick peak stretcher as it needs to match the bandwidth of the signal to track the amplitude.

Measure of the pulse area

This is allows us to measure the total charge generated utilizing a simple integrator, with this solution we do not have bandwidth requirements

Measure of the timing

We use a comparator to determine when the pulse surpasses a certain threshold and use the output of the comparator to create the timing signal.

Threshold

To have a more precise measurement we want to set the threshold as low as possible, in particular we want it to match ethe amplitude corresponding to a single photoelectron, as this is the smallest trigger event.

Note

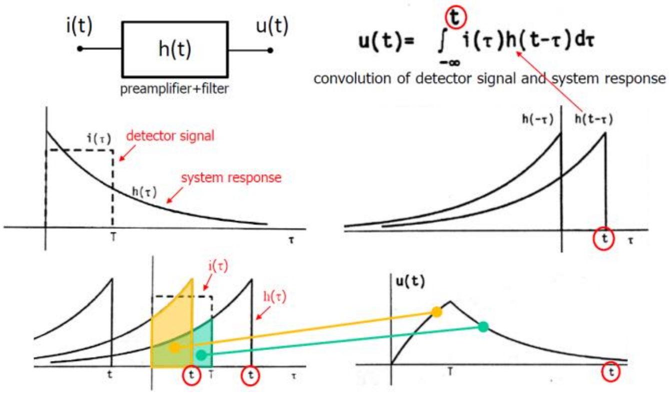
This selection is possible only if the electronic noise is lower than this amplitude, which is often the case.

Time over threshold

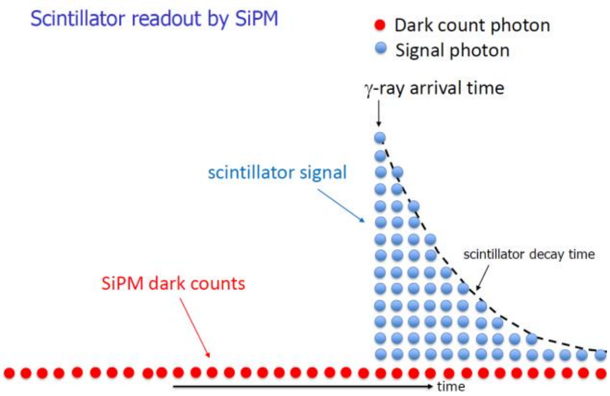
Another value we are interested in measuring when dealing with timing is the time over threshold: the time interval during which the pulse is above the threshold as this gives us information's about the signal energy. Having this measurement allows us to extract timing and energy information with a single measurement chain, the disadvantage is that the relationship between time over threshold and energy is not linear but logarithmic so it is difficult to obtain the energy value.

Selecting the shape of the filter

the output of the filter is the convolution of the signal with the filter transfer function.

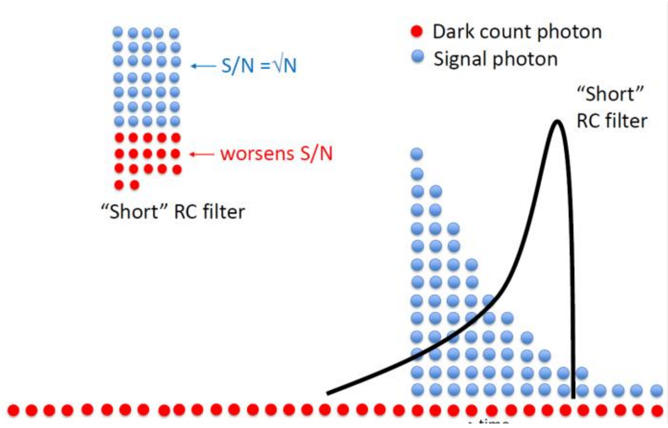


We want to select the filter shape so that the contribution to the output of the SiPM dark counts is minimized.



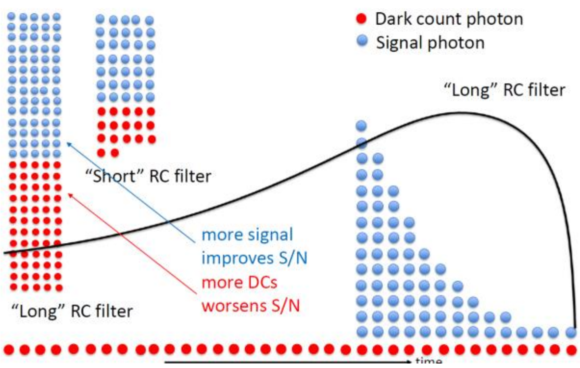
We can see that while this solution collects small amounts of noise it also cuts off a lot of the signal thus deteriorating the signal.

Short RC filter



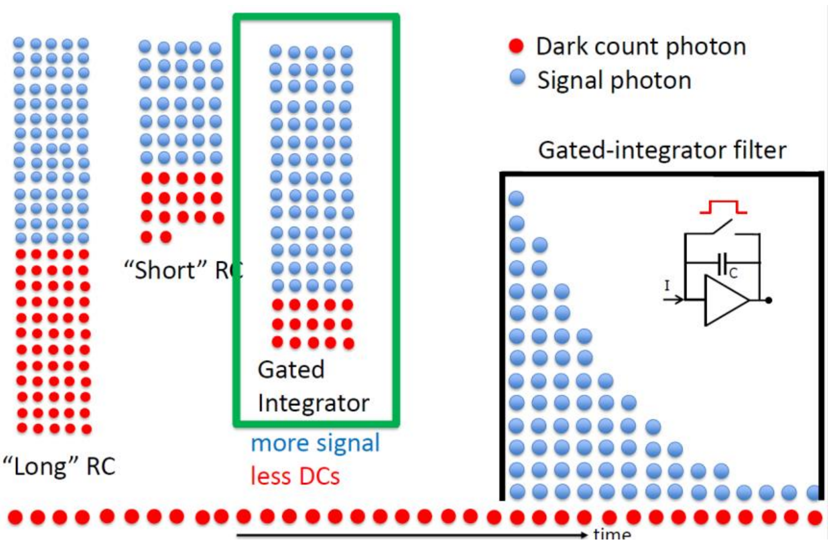
Long RC filter

This solution increases the amount of both signal and noise components.



Gated integrator filter

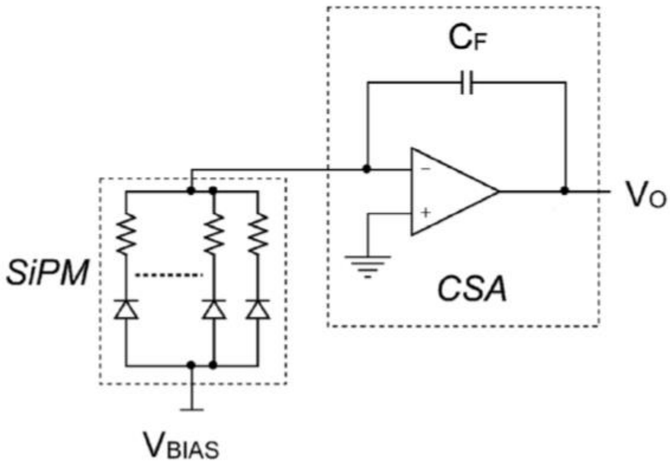
This is the best solution as its shape is ideally a rectangle and allows us to integrate al much as the signal as possible while not including periods with only noise unlike what happens with the slow RC filter.
The integration time is usually around 4/5 times the exponential pulse decay time constant.



Electronics readout options for SiPMs

Charge preamplifier

For SiPM the charge preamplifier is seldomly used because the SiPM delivers too much charge, so the feedback capacitor C_F would need to be very large, larger than it is reasonable to make in an integrated circuit.

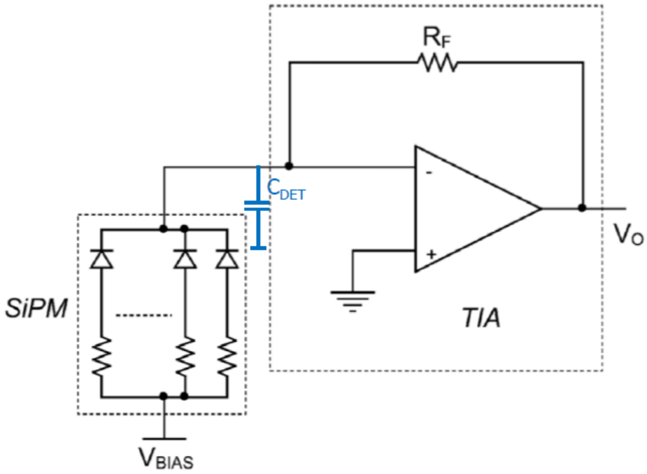


Example

If we consider a PET gamma ray. $511keV$ then a good scintillator will release about 2000 electrons, adding the gain of the SiPM we obtain a charge of about $320pC$, to obtain an output voltage swing of the amplifier of $1V$ the capacitor would need to be $320pF$ which is too large for an integrate4d cMOS technology.

Transimpedance amplifier

With this solution the current pulse is directly converted into a voltage



Timing

If we have a large enough bandwidth we can preserve the fast rise time of the input signal and thus allow for good timing performances as now a discriminator can be placed at the output of the TIA to extract the time information

Energy/ amplitude

The amplitude measurement remains an open problem since to measure it we would need to convert back the voltage signal into a current signal and then feed it into an integrator.

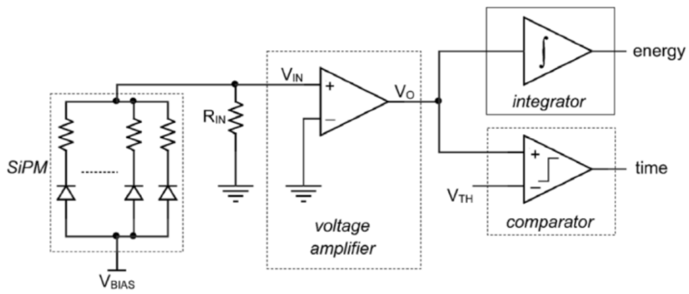
Drawback: instability

We have a slow pole introduced by the resistance R_F and the capacitance of the detector C_{DET}, this capacitance is typically very large even in the order of nF, this may cause instability and force us to place a compensation capacitance C_F in parallel to R_F

Voltage readout

this is the simplest solution: we place a resistor at the output of the SiPM, since the charge emitted is so large even a small resistance in the order of $50m\Omega$ will create a measurable voltage.

We then use an open loop voltage amplifier to amplify the voltage and sent the output to

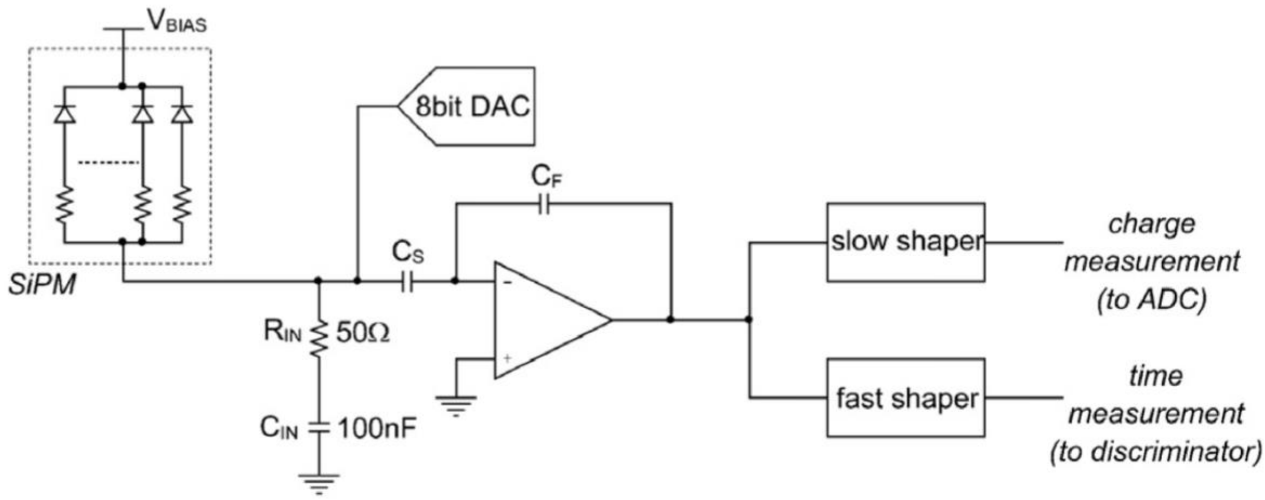


- An integrator for energy measurement
- A comparator for timing measurement

Drawback

Since we are changing the voltage at the end of the SiPM, when a voltage is generated than the overvoltage of the SiPM and thus its multiplication factor thus introducing a non linearity since the signal itself affects the gain. This means that there is a limit to the signal we can obtain because if the resistance is too high and thus the voltage increases too much we are going to have significant non linearities.

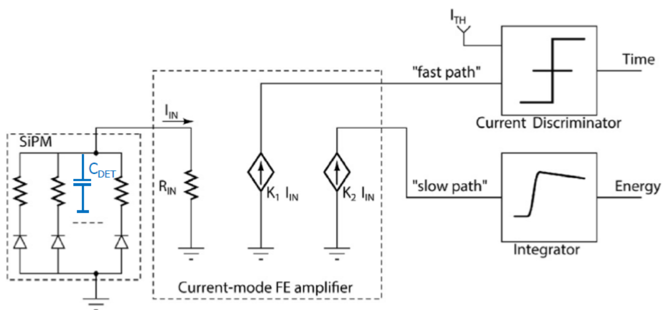
Real circuit



We can see that in this implementation the bias of the SiPM is done with an 8 bit DAC this allows us to fine tune the overvoltage and thus the gain.
A capacitor C_{in} is placed in series with R_{in} to allow for a DC bias without having current flowing in R_{IN}.
Then we have an inverting amplifier utilizing capacitors rather than resistor.

Current readout

in this configuration we utilize a current buffer to transfer the current signal coming from the SiPM.



The current buffer input resistance R_{IN}(small 30Ω) and the detector capacitance C_{DET} (big $1nF$) introduce a time constant meaning that even if the incoming signal form the SiPM is fast the signal reaching the following circuits will be slow down.

Outputs of the buffer

The buffer will have 2 outputs with different gain factors k, one fast used for time measurement and one slow for energy measurement.

Energy output

To obtain the energy we utilize an integrator, to be able to do so the signal must be attenuated (we said that otherwise the required capacitance would need to be to large).
Attenuating a signal makes the electronic noise introduced by the circuit more relevant so, the attenuation level should be carefully selected, in general in this kind of circuits the noise introduced by the integrator will be dominant over other noise sources.

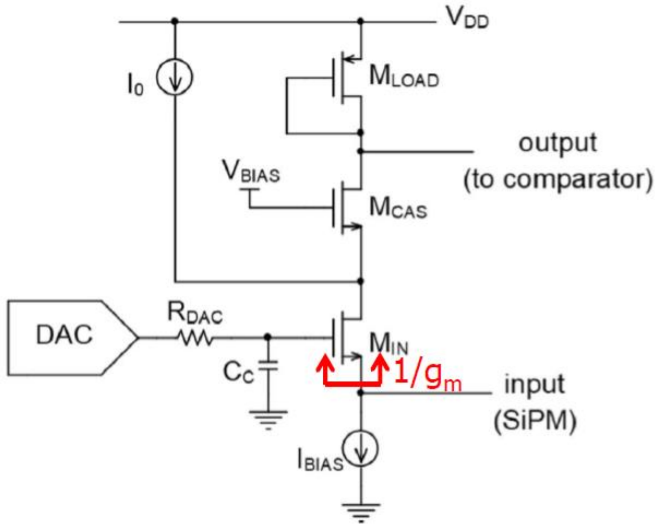
Timing output

For the timing path the attenuation can be smaller because we do not care if the signal saturates. We only care when the signal crosses the threshold.

Configuration 1 cascode

This is the simplest configuration, we use a cascode transistor to read the input current which is then converted to a voltage thanks to the transdiode M_{load}.

We utilize the input cascode transistor as a source follower to provide the bias voltage for the SiPM.



Operation in detail

The input resistance is the $\frac{1}{g_m}$ of the input transistor, we want it to be as small as possible so we want for the bias current to be large, however using a large current introduces 2 problems

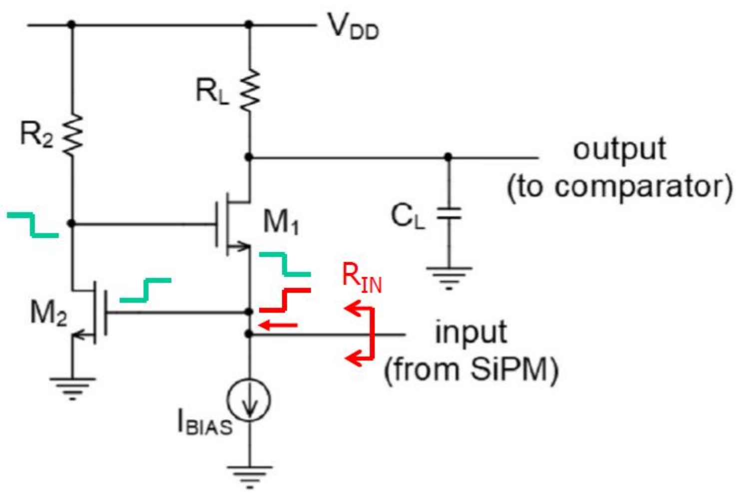
1) Power consumption
2) If the current is the same bias current as M_{load} also its transimpedance will decrease and we attenuate the signal

For this reason we have the current generator I_o as well as the second cascode transistor M_{cas}, this configuration allows us to have 2 separate bias currents for M_{LOAD} and M_{IN} as now M_{load} will need to provide only $I_{bias} - I_0$.

Configuration 2 regulated cascode

this is a variation of the cascode configuration created with the objective of reducing the power consumption.

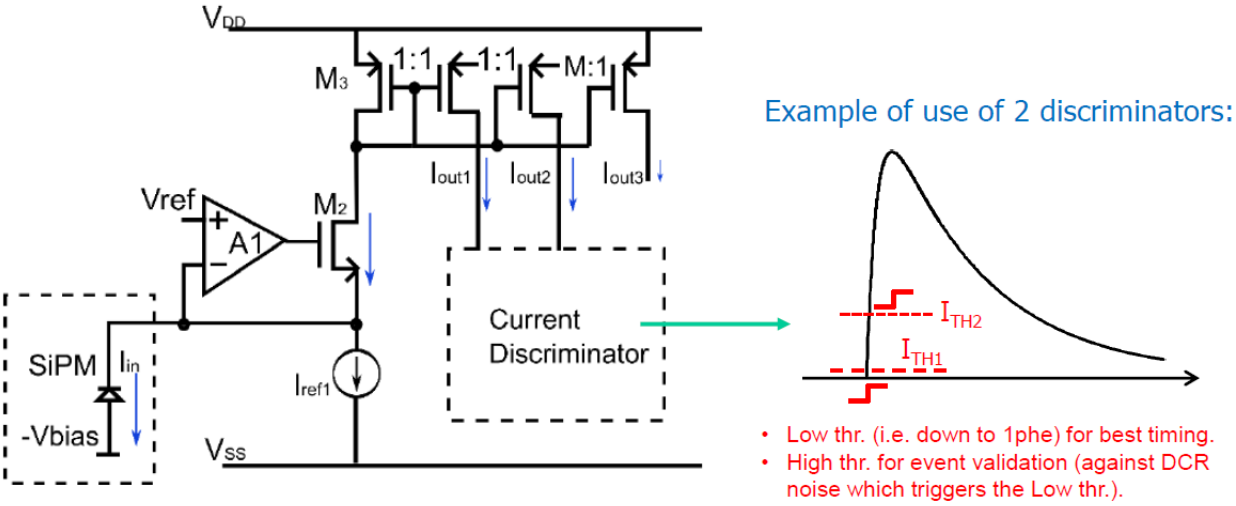
We introduce a feedback so that the input impedance can be reduced by the loop gain which is equal to $g_{m2} \cdot R_2$, this allows us to reduce the requirements of the current also because if we reach weak inversion the g_m to I relationship becomes linear and not connected to the square root.



Since we introduce a loop we need to consider stability since now we have multiple poles.

Configuration 3 regulated cascode with improved loop

We create the loop utilizing a real amplifier and not only a resistor this allows us to have a much higher gain and thus a much lower input impedance.



We can see that we have multiple outputs

- Output 3 is demultiplied so that then we are able to integrate it
- Outputs 1 and 2 are transferred as is because we do not care if they saturate

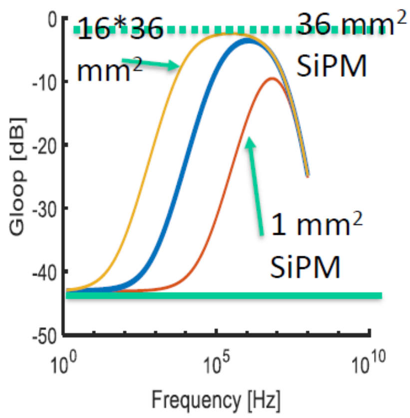
The reason why the discriminator utilizes 2 equal inputs is because we test with 2 separate thresholds

- A lower threshold at the level of a single photon released which start the count
- A higher threshold which is used to verify that the pulse is generated by more than 1 photon as this would most likely be simply a dark current electron.

Doint this allows us to have a high precision in timing measurements while greatly reducing the amount of dals triggers because of dark current.

Advantage is stability

We said before that the capacitance of the SiPM creates a low frequency pole which united with the loop of the regulated cascode can cause instability, instead in this configuration even if C_{DET} is very large we are stable because increasing the capacitance never makes the loop go past 1.

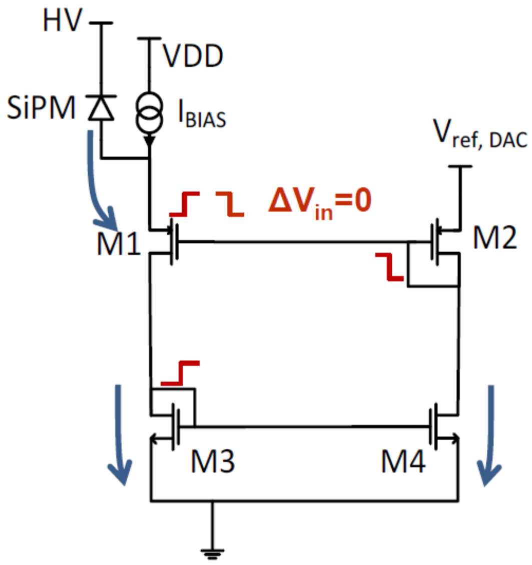


Disadvantage

The transistor sizing is critical as the loop gain must never cross 1.

Alternative configuration for current readout with positive feedback

- $\left(\frac{W}{L}\right)\Big|_{M1} = \left(\frac{W}{L}\right)\Big|_{M2}$
- $\left(\frac{W}{L}\right)\Big|_{M3} = \left(\frac{W}{L}\right)\Big|_{M4}$



"Fast CMOS current amplifier and buffer stage"
G.C. Temes, W.H. Ki

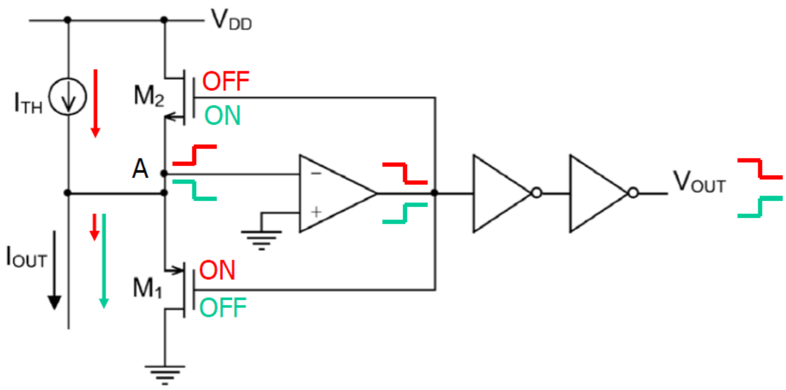
In this configuration we have a positive loop, we however remain stable because the loop gain is between 0 and 1.

Discriminator

This is a bistable circuit.

If $I_{OUT} < I_{TH}$ then the voltage at A is positive the output of the amplifier is negative, the pMOS M1 is on and allows the excess current to flow.

If $I_{OUT} > I_{TH}$ the voltage at A is negative the output is positive so M1 is off and M2 is on providing the current required.



Gamma cameras for SPECT and PET

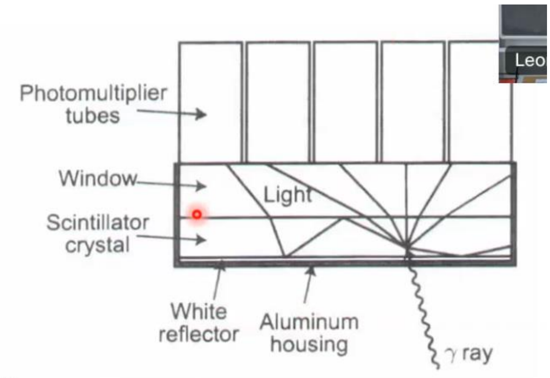
Anger camera

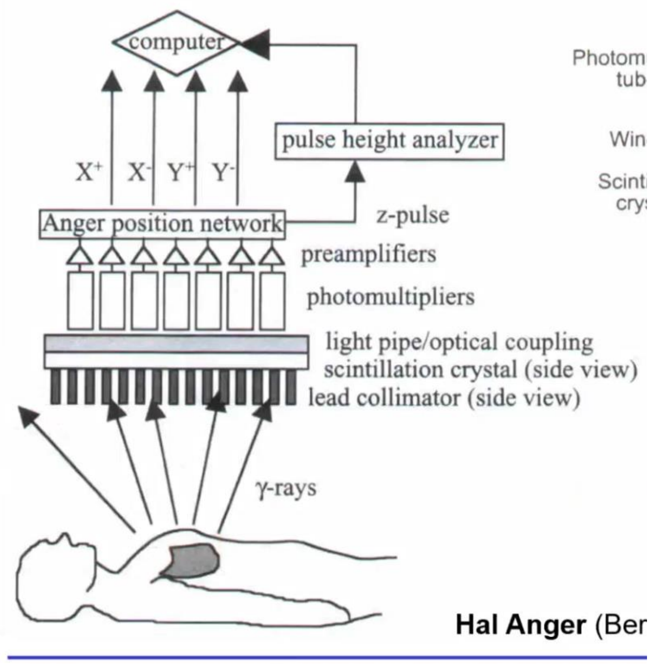
An Anger camera is a detector based on

- A monolithic crystal
- A segmented photosensor

We have that each sensor will acquire a different amount of light depending on the solid angle with respect to the interaction between ray and scintillator.

From the distribution we can obtain the interaction position.

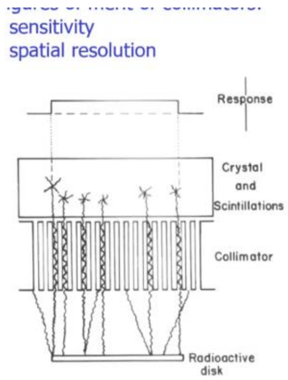


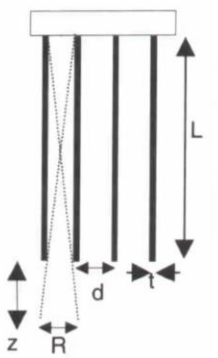


Window

Note that there is a window between the scintillator and the photosensor this is done because the additional distance allows to spread the light on a larger surface providing a response from more channels so the distribution includes more elements and it becomes possible to better determine the position of interaction, otherwise if the light is collected by a single photosensor we can not do any analysis on he sampled data.

The collimator





The collimator is utilized to mechanically select the gamma rays which are able to interact with the scintillator.

Only the gamma rays perpendicular to the scintillator will arrive so that it is possible to detect their direction.

Figures of merit

Sensitivity

Is the ratio between the gamma rays arriving on the collimator and the number of photon which can pass through it and can thus be detected.

the larger the dimension of the hole and the shorter its length the larger its sensitivity.

$$S = k\frac{d^2}{L(d+t)}$$

Spatial resolution

It indicates our ability to distinguish the volume of response form which the gamma ray is originated.

$$R_G = geometric\ resolution = \frac{d(L+z)}{L}$$

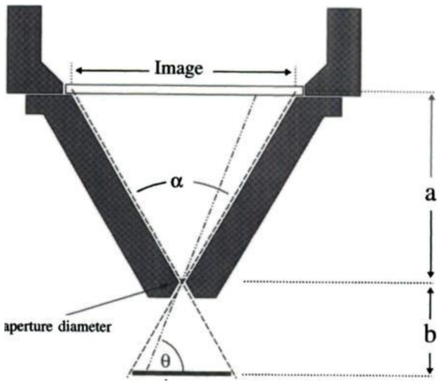
The total resolution will be given by the quadratic sum of the geometric resolution and the intrinsic resolution of the detector

$$R_{tot}^2 = R_G^2 + R_i^2$$

We can see that there is a tradeoff between resolution and sensitivity when we determine the dimension of the collimator holes.

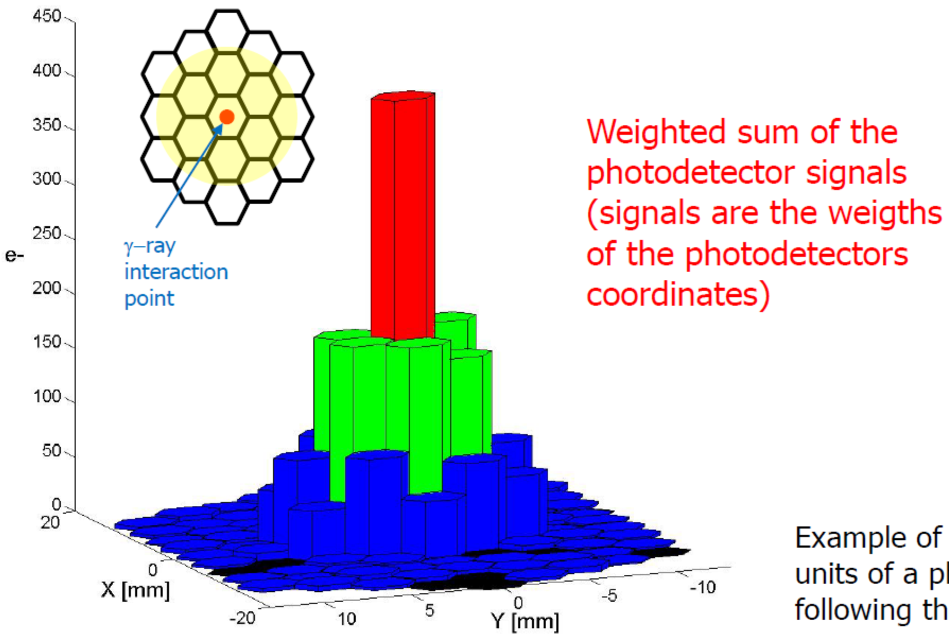
Pin hole collimator

We leave only a small conical input for the light so that we can still detect the direction



Techniques for the reconstruction of the points of interaction

Center of gravity



$$x_0 = \frac{\sum_i N_i \cdot x_i}{\sum_i N_i}$$

$$y_0 = \frac{\sum_i N_i \cdot y_i}{\sum_i N_i}$$

Example of signals collected by the units of a photodetector array following the absorption of a γ–ray.

Note the detector may be hexagonal, this makes them more expensive but also increases the final resolution.

Determining the position with the center of gravity

We simply create an histogram indicating the among of charge collected by each detector.
The position can be found by making a weighted sum of the photodetector's coordinates weighted by the signal that they received.

Example

We sum all the coordinates of the signals multiplying each for the signal it has received then we normalize for the sum of all signals

$$x_0 = \frac{\sum N_i \cdot x_i}{\sum N_i}$$

Limitations of this method

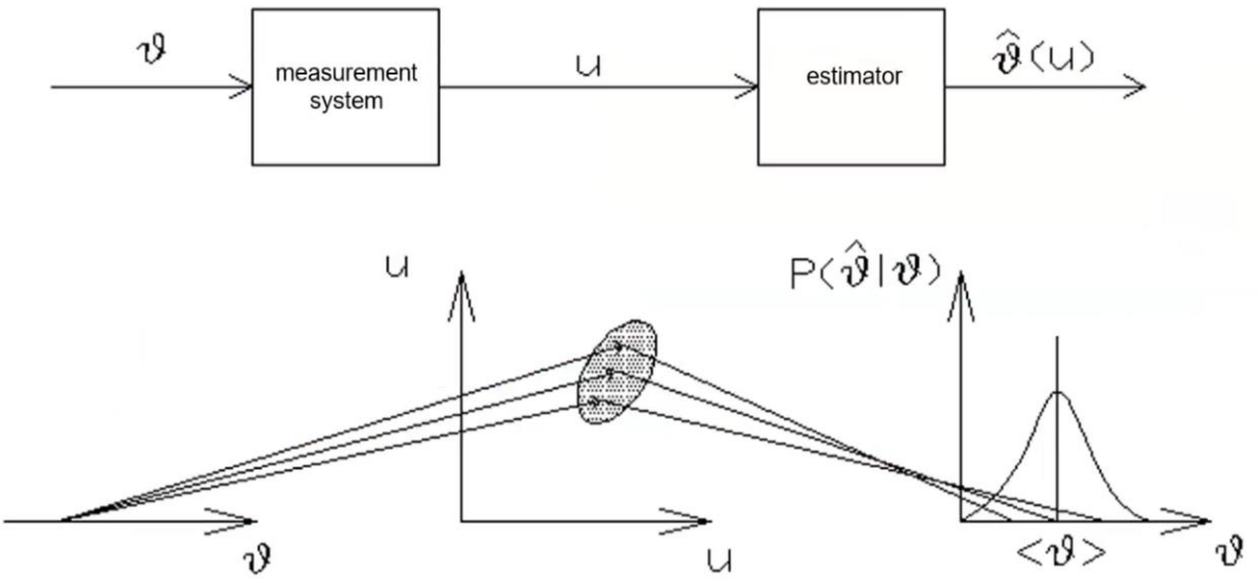
Interactions at the edges

When we have an interaction close to the edge of a crystal only part of the light signal will be lost on the border of the wall becoming homogeneous background so there is a bias towards the center of the cell.

Broken photosensor channel

If we have a broken photosensor channel, then we lose a significant part of the information's and thus the final result will be distorted

Maximum likelihood ML

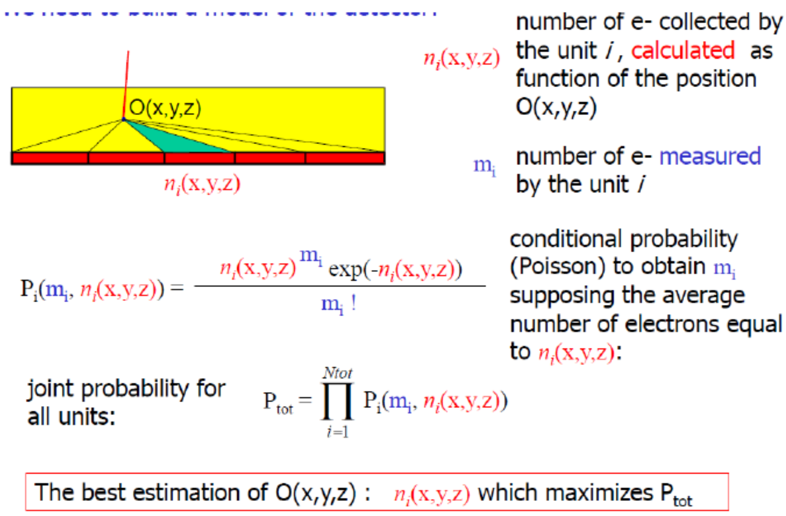


This is a more sophisticated technique based on the idea of building an estimator.

Building an estimator means building a function which takes as input the measurements of the sensors and provides as an output an estimation of the coordinates of interaction and thus of origin of the gamma ray. It is a function build out of probability.

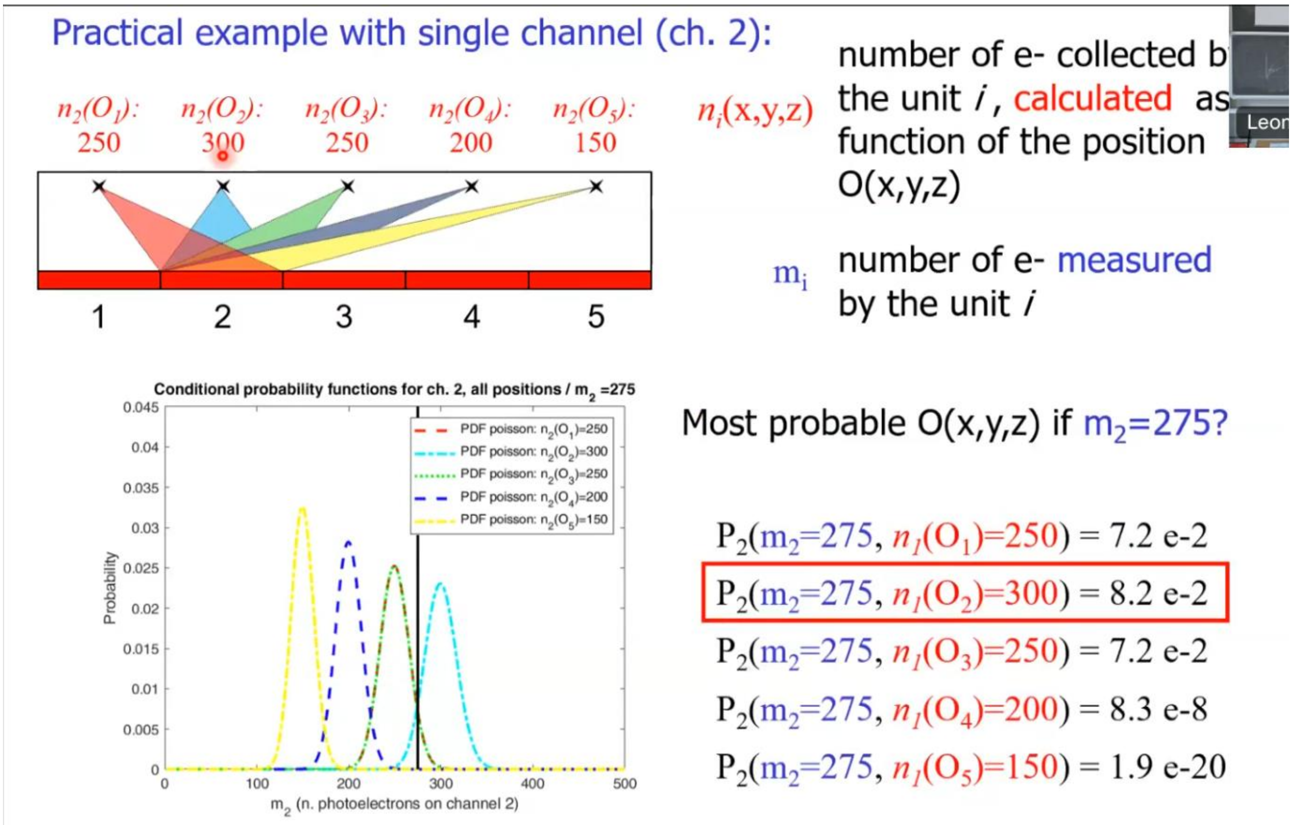
Modeling the detector

First of all we need to create the model of the photodetector which should be something like the image where we indicate ethe monolithic scintillator crystal in yellow and the photodetectors in red.



We want to create a look up table where the output is the **number of photoelectrons** $\boldsymbol{n_i}$ collected by each **photodetector** $\boldsymbol{i}$ as a function of the **position of interaction** $\boldsymbol{O(x,y,z)}$.

The model can be built considering that each photodetectors covers a certain solid angle which connects is surface to the point of interaction, meaning that depending on the position of O the angle will be larger or smaller and thus the amount of photons detected n_i will change accordingly.



Applying the model

Once we have the model the next step is to reconstruct the position of interaction given a set of measurements m_i indicating the response of each photodetector i.
What we need to do is build a conditional probability: **the probability to have measured $\boldsymbol{m_i}$ photons with that photodetector with an incoming signal $\boldsymbol{n_i}$**.

Example

Let's focus on a specific detector n_3, which has measured $m_3 = 140$ electrons, I need to define a function which gives me the probability for a given average signal n_i to have measured m_3.
This probability is given by the Poisson probability.

Now since I'm operating on a single pixel right now the average signal n_i which gives me the highest probability to have measured m_3 electrons is also a signal with $n_i = m_3 = 140$ electrons.

Now I can go back using the look up table to find which coordinates for O can give $n_i = 140$ electrons (note that it will not be a single result).

This result is not good enough with a single pixel so we repeat for each pixel and determine the joint probability.

Joint probability

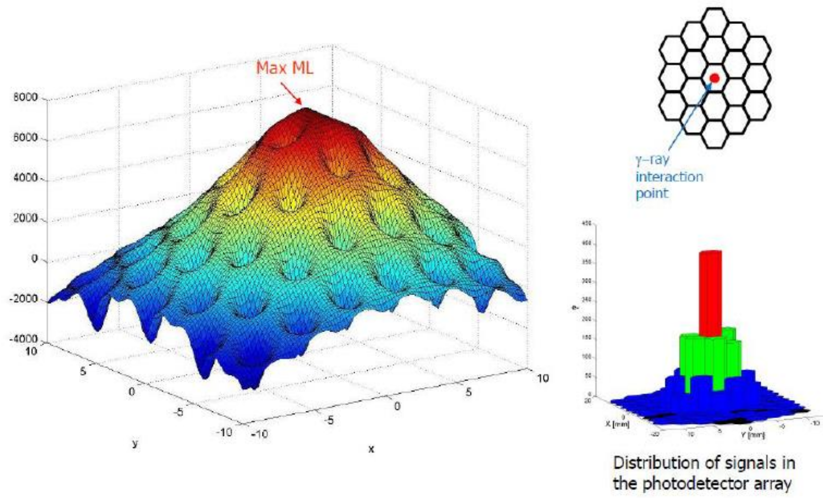
We multiply the probability for each $n_i(x, y, z)$ result for each pixel and then determine the correct $n_i(x, y, z)$ as the maximum of this function.
This solutions much more robust since we are taking into account all pixels.

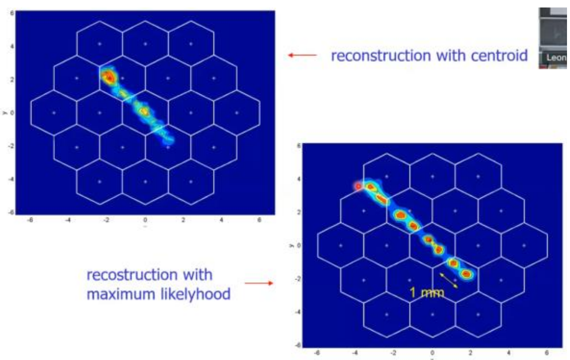
Example

Below we see a gamma ray interaction at the center of a camera made out of hexagonal pixels.

We applied the ML method and obtain the results.

We can see that across the map there are local minimums corresponding to the centers of neighbor pixels.
This balances what in the gravity center was the bias towards the center of the pixel allowing us to reconstruct points of interactions at the boundaries.



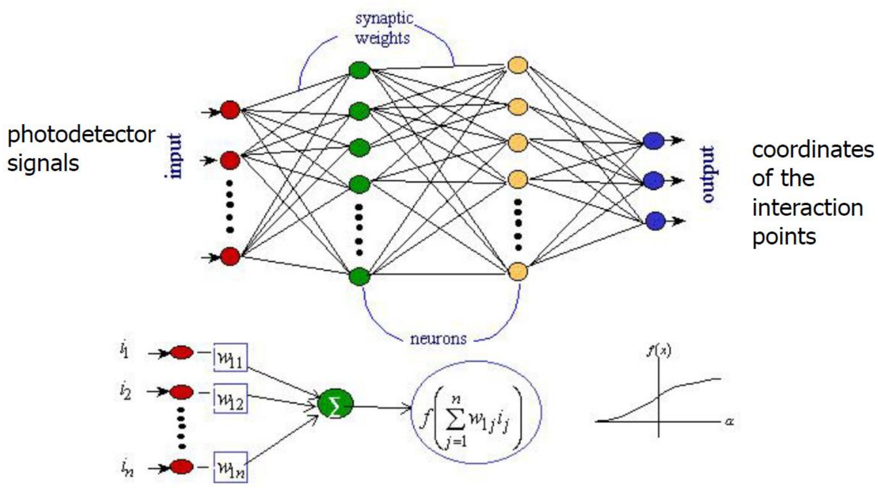


Conclusion

In this solution the detectors plays a role in determining the interaction point and we do not simply utilize an algebraic formula, this means that we can circumvent limitations of boundaries and broken units, each detector provides information so having a broke detector just deprives us of some information rather than giving us wrong ones since we are able to exclude them.

Since the formula becomes more and more complex as he number of detector increases we can consider only those which have a relevant signals, this is typically done by selecting the detector with the highest measured signal and selecting only the one close to it in a certain range.

The neural network



A neural model tries to emulate the operation of the brain.

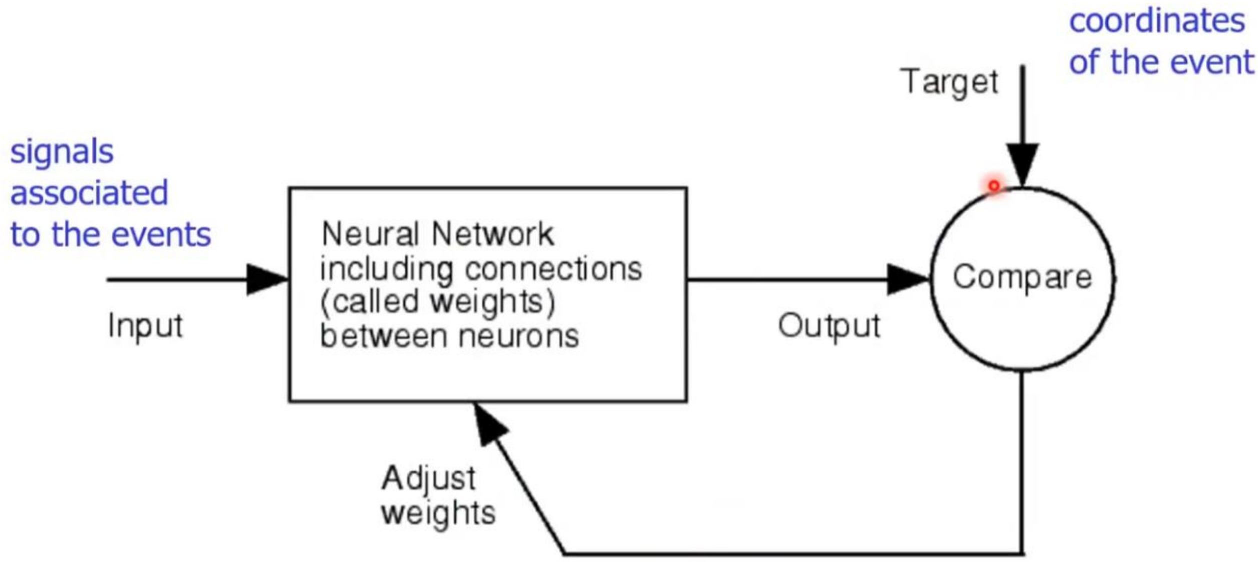
It is an algorithm with a certain number of inputs and outputs, between we have a series of weights and nodes which create the transfer function.

Operation

We can see in the plot above we have 4 layers a red a greed a yellow and finally a blue one, the value of each layer is obtained as the weighted sum of the values of the previous layer.

Training phase

We need to train our neural network to determine the best weight to obtain a correct prediction.
We have a set of input data together with the correspond correct inputs, we initialize the weights of the



neural network to random value and then we execute the data and modify the weights with different algorithms to improve the results.

Pro

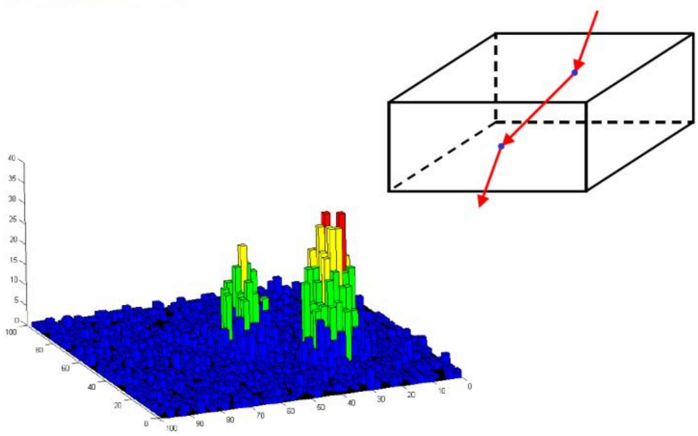
- Very suitable for FPGA implementation (multiplication and sums)

Cons

- Very long training phase

Application of NN for reconstruction of Compton events

As we know it is difficult to reconstruct Compton events because we obtain 2 simultaneous light signals which can be either separated (like in the image) or superimposed.

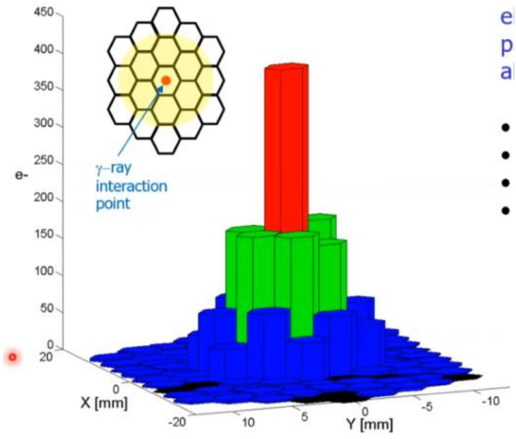


These are **impossible to reconstruct with maximal likelihood** because our model would need to include all possible combinations of interactions.

Instead if we train a neural network with data containing interactions of the Compton effect we can obtain a way to reconstruct them reliably.

Impact of the photodetector/amplifier electronic noise on the anger camera spatial resolution

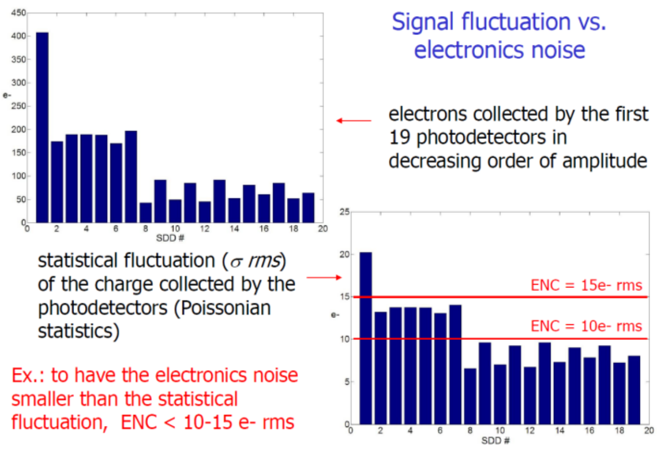
We can assume to have a fluctuation of the number of photoelectrons emitted by the photodetector given by the Poisson statistic.



Photodetector variance

We cab see in the image the effect of these fluctuations: the green columns are equally distant from the interaction point and thus should all have the same height, however because of the fluctuations the results are different.

2D graph



The top graph is the same as the 3D graph instead the bottom one indicates the expected variance for each signal which is set equal to the square root of the number of electrons.

Electronic noise

We need to consider also the presence of electronic noise of the amplifier which we express as the ENC, the overall variance will be given by the quadratic sum of the ENC and the Poisson variance.
Since it is a quadratic sum the electronic noise becomes relevant if the signal become smaller, while its effect is almost negligible when we deal with a high signal.
The total variations are what prevents the anger camera to perfectly reconstruct the interaction position.

Energy resolution

The total energy of the gamma ray measured is obtained as the sum of the readouts of all channels, so we get

$$E_g = G \times \sum_i (N_i + noise_i)$$

Where

- G is the gain factor which tells us how to convert the number of photoelectrons into energy of the gamma ray
- N_i is the number of events

The sum of the electronic noise of all the units of the array contributes to the statistical fluctuation of the computed energy.

Why energy resolution is improtant

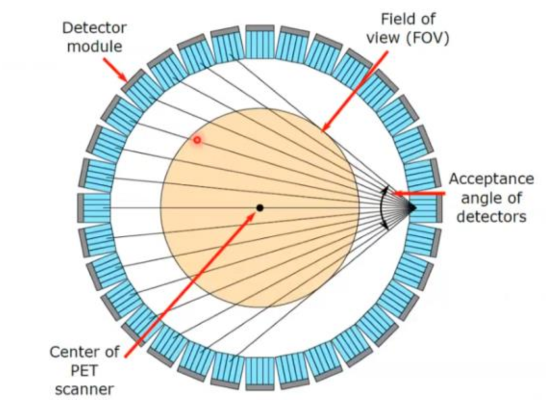
We require a good energy resolution to separate the Compton interactions which will otherwise provide false data.

Positron emission computed tomography PET

Operation

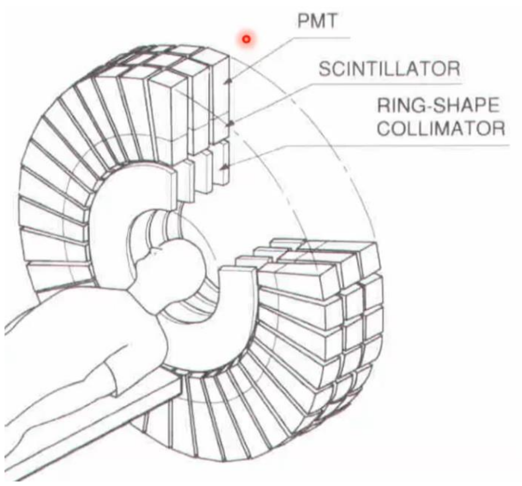
The radioemitter emits positron which travel a short distance before annihilating with electrons and producing a couple of gamma rays travelling in opposite directions (we assume that the overall system is at rest so the total momentum is zero).
Once we detect the 2 rays which will arrive almost instantaneously (a few ns between them) we can draw a line connecting them, if we have multiple lines we can obtain an image of the internal of the patient. Note that if we can extract also the timing of the 2 rays starting dorm the idea that they are generated at the same time it becomes possible to increase the resolution of our image.

Structure of the scanner



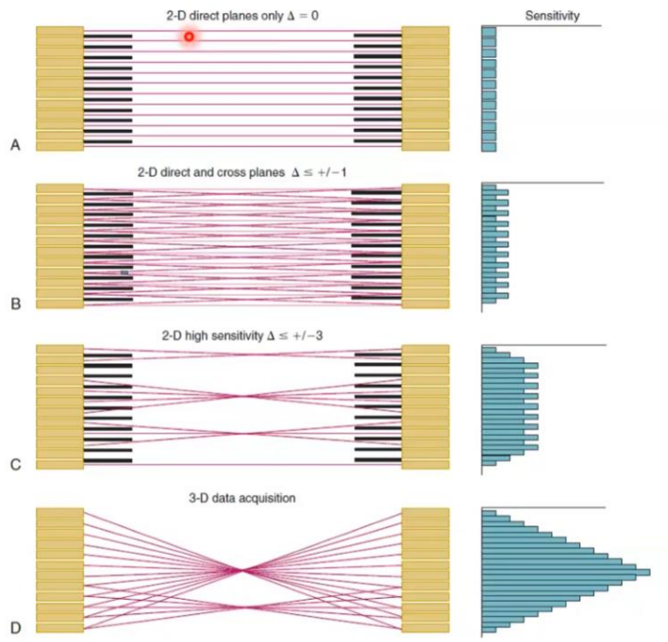
We check for coincidences and accept only signal which have a coupled ray arriving within a very small delay on a detector in the opposite side of the first one activated.

Typical dimensions for a pet scanner is few tens of centimeters.



Collimators

Note that especially in older scanners we also need to have collimators, these collimators are used to separate the different detectors modules.



This reduces the count rate allowing coincidences only along certain directions, this eases up the requirements for the computing power.
Now days with the improvement in performances we are able to utilize PET without collimators.

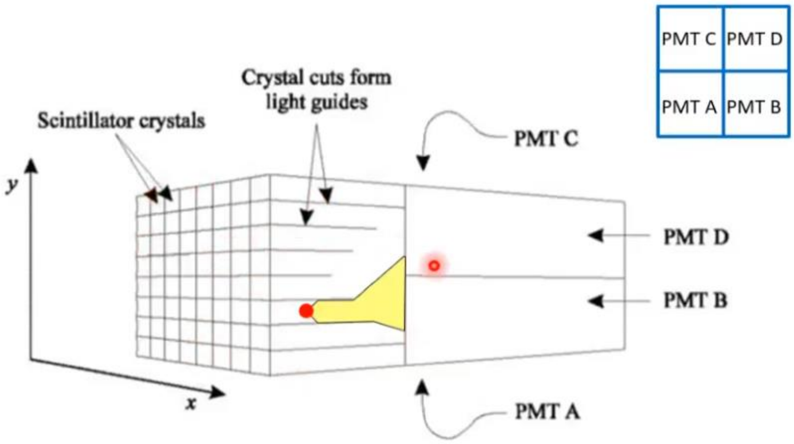
The block detector for PET

This is the first type of detector that was created specifically for PET scanners.
The idea was to build a much thicker scintillator crystal (order of a few centimeters) to compensate for the higher energy of the gamma rays, because of this however the detector needed to be pixelated otherwise the excessive thickness would have allowed the photons to spread too much to obtain a readable result.

$$X = \frac{(PMT_A + PMT_B) - (PMT_C + PMT_D)}{PMT_A + PMT_B + PMT_C + PMT_D}$$

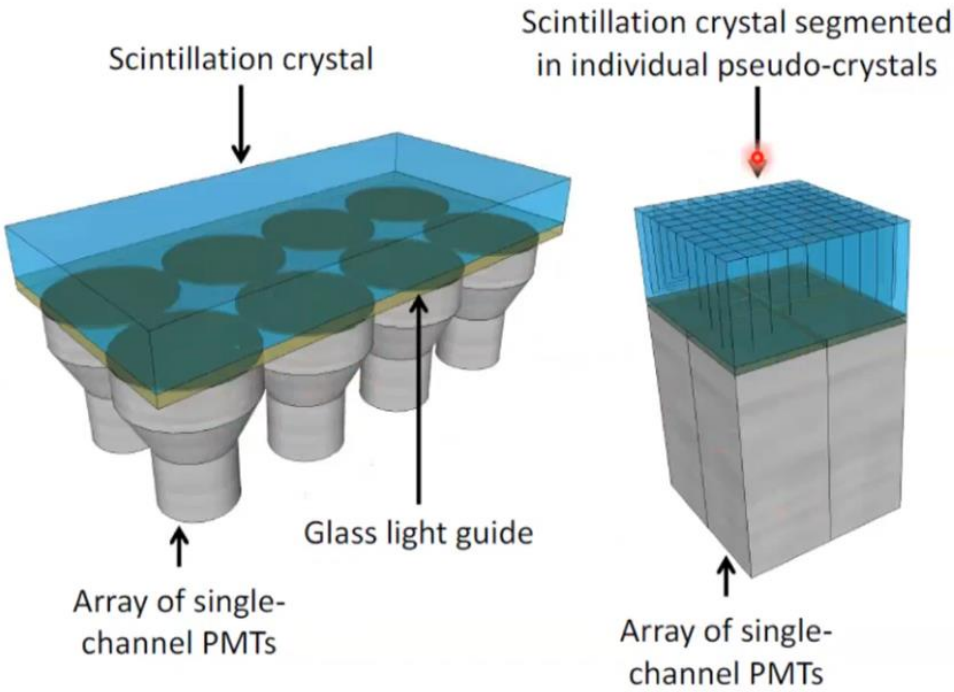
$$Y = \frac{(PMT_A + PMT_C) - (PMT_B + PMT_D)}{PMT_A + PMT_B + PMT_C + PMT_D}$$

where PMT_A, PMT_B, PMT_C and PMT_D are the fractional amounts of light detected by each PMT



To obtain the pixelated crystal we started form a monolithic crystal and then cut it in sections, the cuts were made longer at the edges and shorter as we moved towards the center of the crystal.
This way the light was initially channeled but then could spread, this was necessary so that we would not need a single detector for each pixel and instead we could apply the same reconstruction methods used in anger cameras.

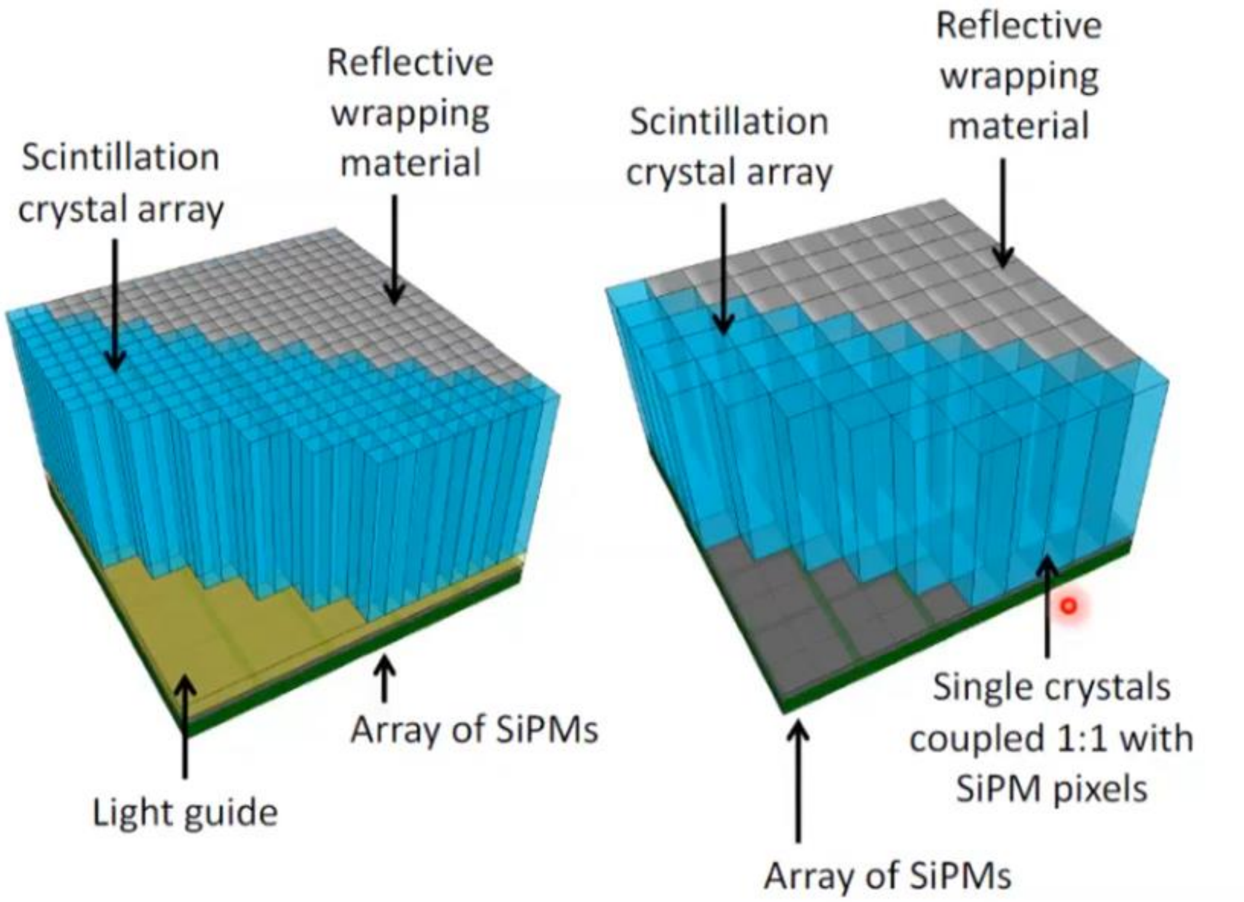
Comparison between anger camera for SPECT and PET



Comparison of different detector architectures

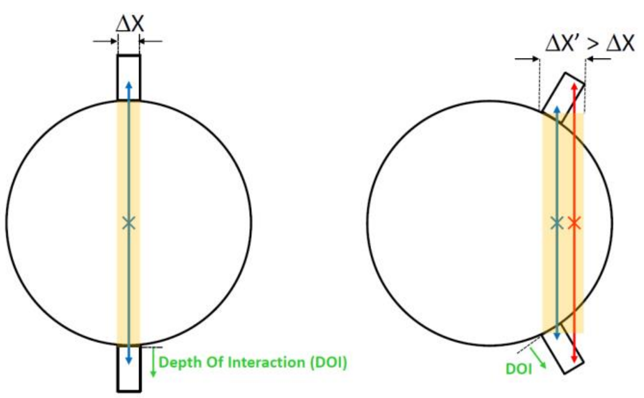
In recent years in block detectors PMT have been substituted by SiPM so we are now able to have fully pixelated crystals.
There are 2 main types of pixelated detectors

- One in which each crustal is coupled with a single SiPM, we have a limit because there is a limit on how small we can make the sensor
- One where the SiPM are larger than a single crystal pixel and we calculate the position of interaction by measuring the light from multiple sensors



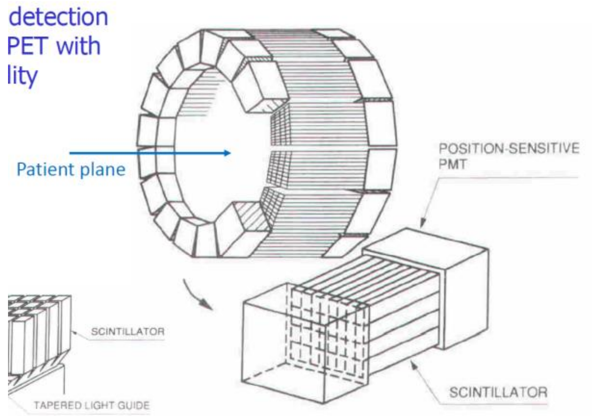
The parallax error

We have a ring of detectors, fi the event comes from the center of the detector we have no problem, however events coming from off center are affected by the depth of interaction, the events can come from a wider region than the simple width of the crystal thus the spatial resolution is worsened.



How to estimate the effect of the Depth of interaction in the measurement

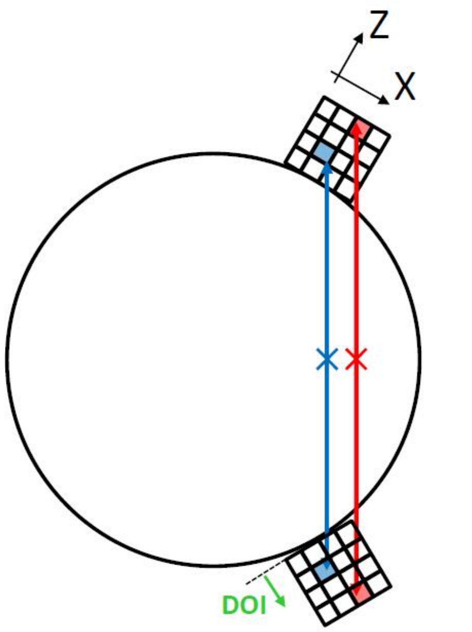
We rotate the crystal array and make it longer.
Now the pixelation is not radial but longitudinal: the scintillators crystal sections are parallel with the patient.



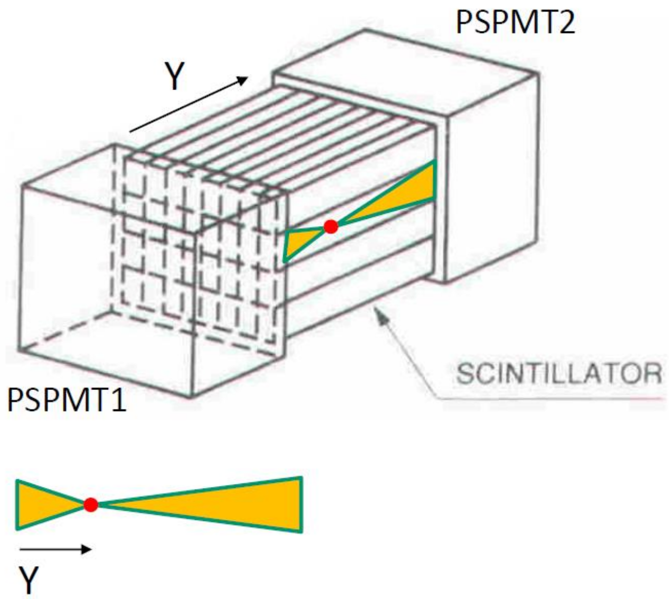
It is possible to utilize position sensitive PMT to distinguish the position of interaction in the X and Y axis.

Y axis

It is important to note that with this solution we lose information about the position along the y axis (parallel with the patient).
To get it back we add a second detector on the opposite side of the scintillator crystal, we will be able to extract the y coordinates by comparing the amounts of light detected: part of the light will be dissipated by the crystal so the sensor further away form the interaction position will receive less light.



Pixelization in Z direction (in addition to X) allows to implement DOI reconstruction



Reconstruction on Y coordinate is obtained by the ratio of the signals collected by the PMTs which is dependent on the position of interaction

Conclusion

We obtain coordinates x, z thanks to the pixelization while we use the light sharing to obtain the y data, this allows us to have a **fully 3d reconstruction of the gamma ray interaction**.

Alternative detection system for PET with DOI capability

In this case we utilize a monolithic scintillator.

We estimate the position on the base that we expect the light peak to change depending on the interaction point position

- It moves for changes in the horizontal position
- It increases or decreases depending on the depth of interaction

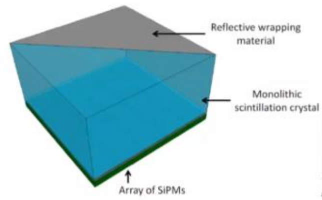


Figure 1.11: Schematic drawing of a monolithic scintillator detector.

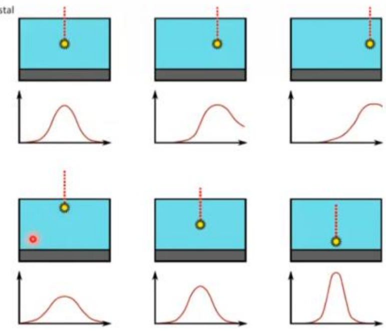
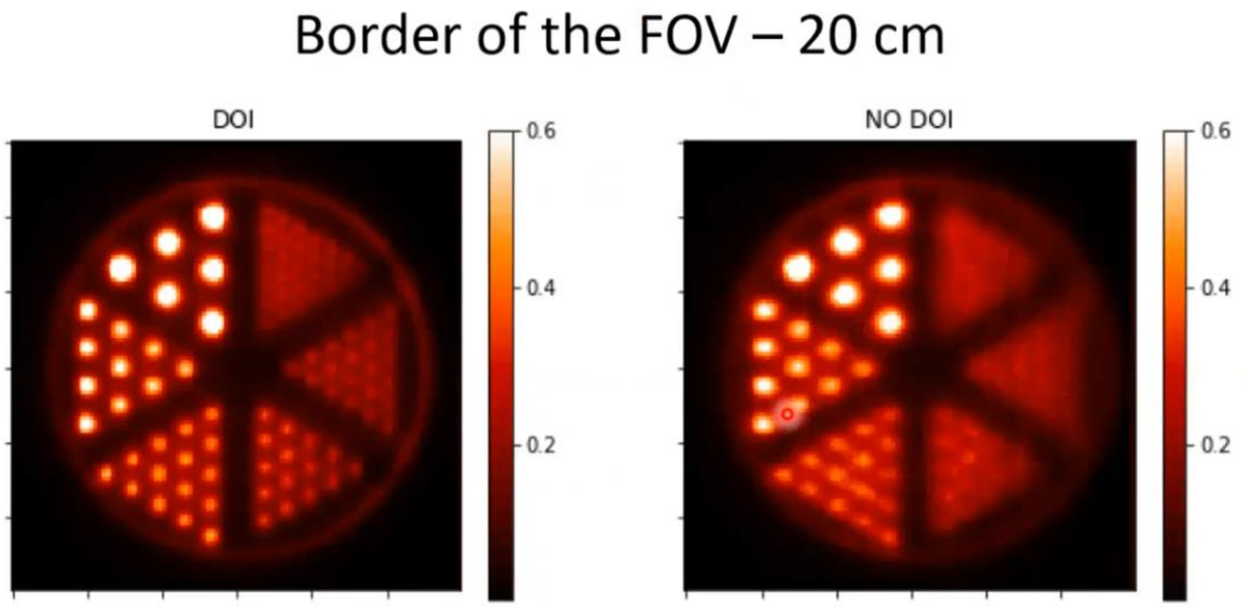


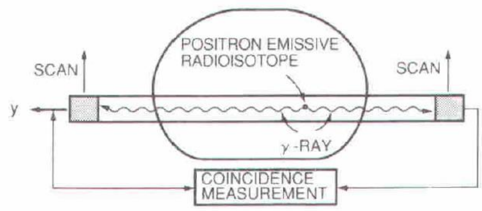
Figure 1.12: Simplified illustration of the working principle of monolithic scintillator detectors: if the position of interaction shifts laterally, the peak of the light distribution is expected to shift accordingly (top row); if the interaction point changes depth, the light distribution is expected to change width, resulting in a narrower light distribution when the interaction is closer to the photosensor.

Example of improvement with DOI

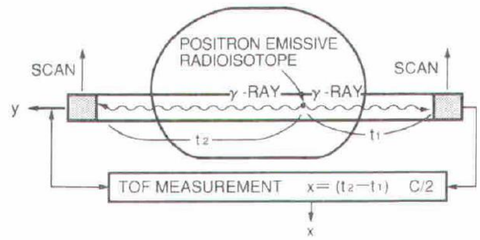


PET Time of flight

As we know in PET we measure coincidences between gamma rays, if we are able to accurately measure di time difference between the 2 rays it becomes possible to detect not only the direction but the exact position in which the gamma rays were generated



b) TOF: Time of Flight



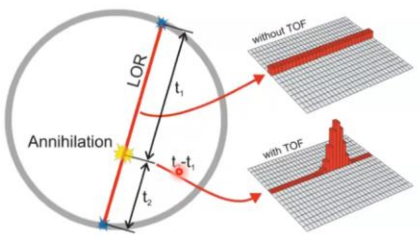
We can find the x position as

$$\Delta x = \frac{c \cdot \Delta t}{2}$$

The variance of this measurement is proportional to the variance in our timing measurement

$$\sigma_x = \frac{c \cdot \sigma_t}{2}$$

Note that we have the term 2 is present because we measure form the 2 edges towards the center and thus make basically 2 measurements.



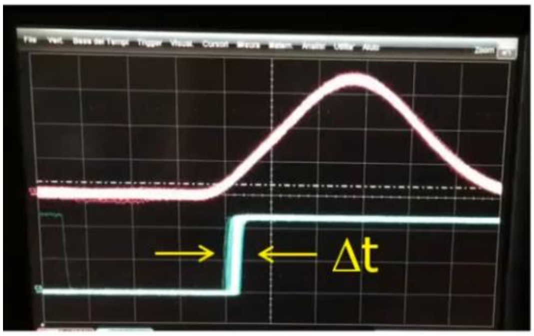
$$\begin{cases} \Delta t = t_2 - t_1 \\ \Delta x = distance\ from\ the\ center\ of\ the\ FOV \\ \sigma_x = uncertainty\ in\ estimating\ \Delta x \\ \sigma_t = uncertainty\ in\ estimating\ \Delta t \end{cases}$$

Resolution of the measurement

We are going to have an uncertainty in the timing because we need to verify when the signal crosses a threshold, so we are going to be influenced by jitter caused by elements both in the scintillator and the photodetector but also to the electronic chain

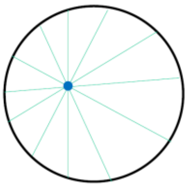
$$\sigma_x = \sigma_t \times \frac{c}{2}$$

Since $c = 30\frac{cm}{ns}$ we get that with a timing uncertainty of just $3ns$ we get a space variance of $\sigma = 45cm$.

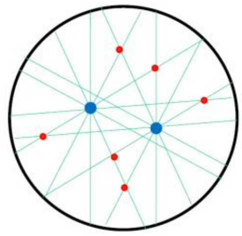


The most recent scanners have a timing resolution of around $\sigma_t = 400ps$ and thus a spatial resolution of $\sigma_x = 4{,}5cm$

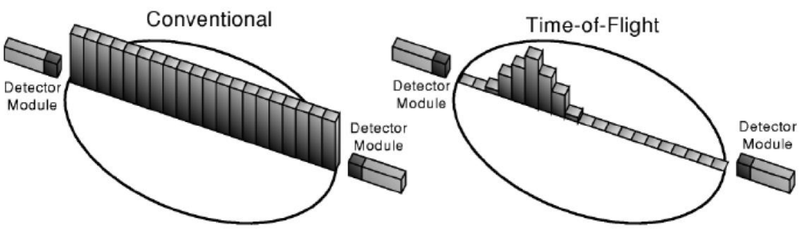
Image reconstruction in PET

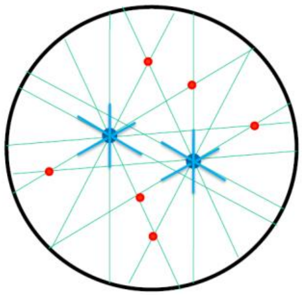


Here we can see the image reconstruction in bet without time of flight, if we have only a single source then determining the intersection between the different events is easy, however if we have multiple source we end up determining a huge number of spurious intersections which reduce the SNR of the reconstructed image and thus its resolution.



This happens because in classic PET we assign the same probability to each pixel between the 2 activated detectors, instead if we have time of flight we can reduce our interest in a smaller area thus allowing us to discard spurious intersections

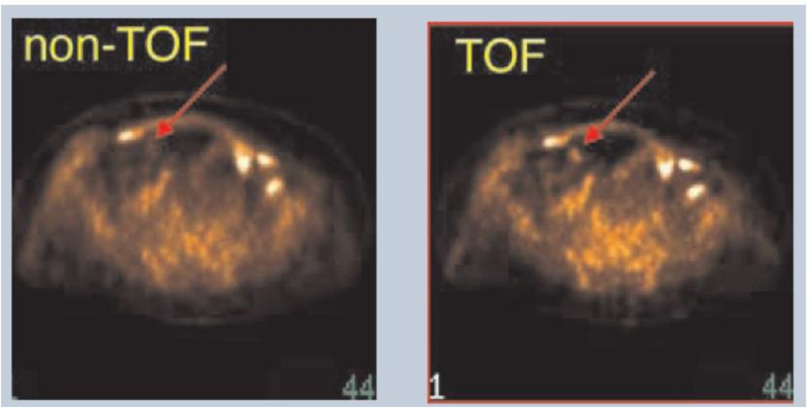




Signal to noise ratio gain in the image

$$SNR_{gain} = \frac{SNR_{TOF}}{SNR_{nonTOF}} \propto \sqrt{\frac{D}{\sigma_x}}$$

Where D is the diameter of the object and as we know $\sigma_x = \frac{c}{2}\sigma_t$, this means that with a $500ps$ timing resolution we obtain an $8cm$ spatial resolution which on an object of $40cm$ gives an improvement on the variance by a factor of 5



Future goal

If we reach a time resolution of $10ps$ we are going to obtain a resolution of $1{,}5mm$, this means that we would be able to reconstruct the origin of the gamma rays in the line with just a couple of gamma rays and we would not need to reconstruct utilizing the crossing of the lines.

Ultimo seminario su PET (quello che manca Lezione 42)

Artificial neural network in detail

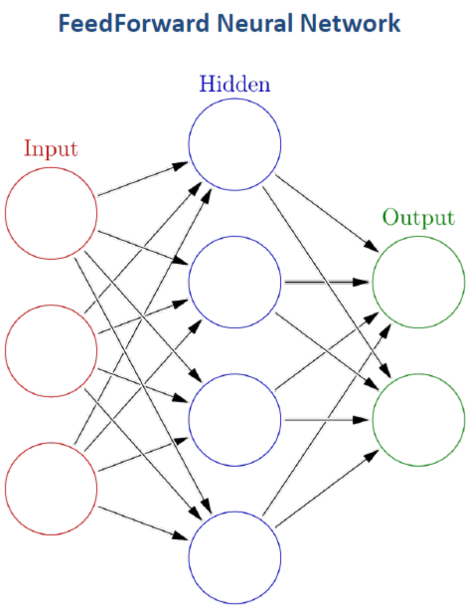
An artificial neural network can be used to solve classification problems: in these problems we have a series of observations that we place in an n dimensional space and we wish to divide them in classes.

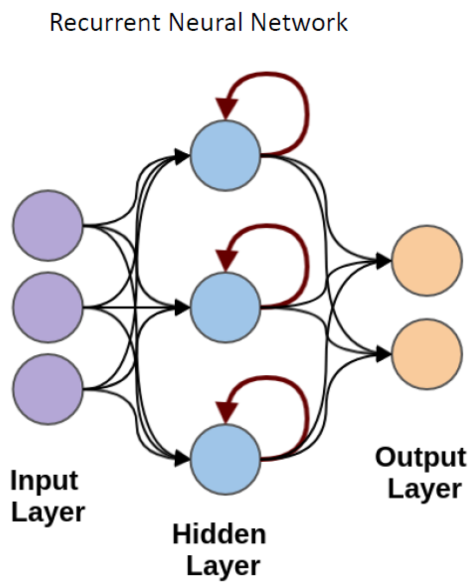
In our case observations are the light signal coming form all the detector channels after the interaction of a gamma ray. So for every observation, sample, event we will acquire a light distribution on all channels and place in a n dimensional space where e n is the number of detectors. Then we want to define an algorithm which assign to each point a class which in our case corresponds to the position of interaction (we could also use it for timing).

What are artificial neural networks

We have already said that neural networks are networks with a certain amount of inputs which are then connected to the outputs through a series of layers which are able to perform operations on the inputs. There are different internal structures of neural networks

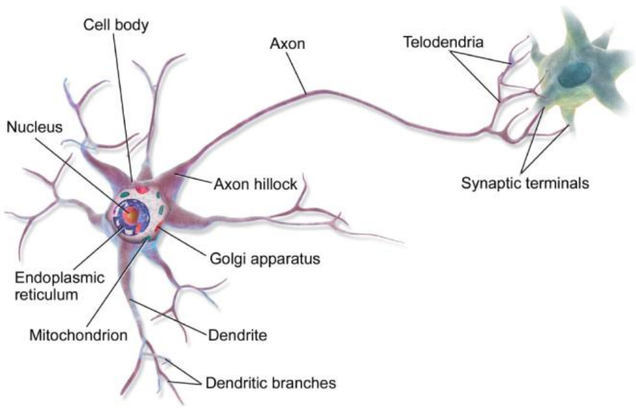
- Feed forward neural networks
  In these the information moves sequentially in the neural network and reaches the output after a fixed amounts of steps
- Recurrent neural networks
  In these topology each node is connected not only with the next layer but also within the same layer



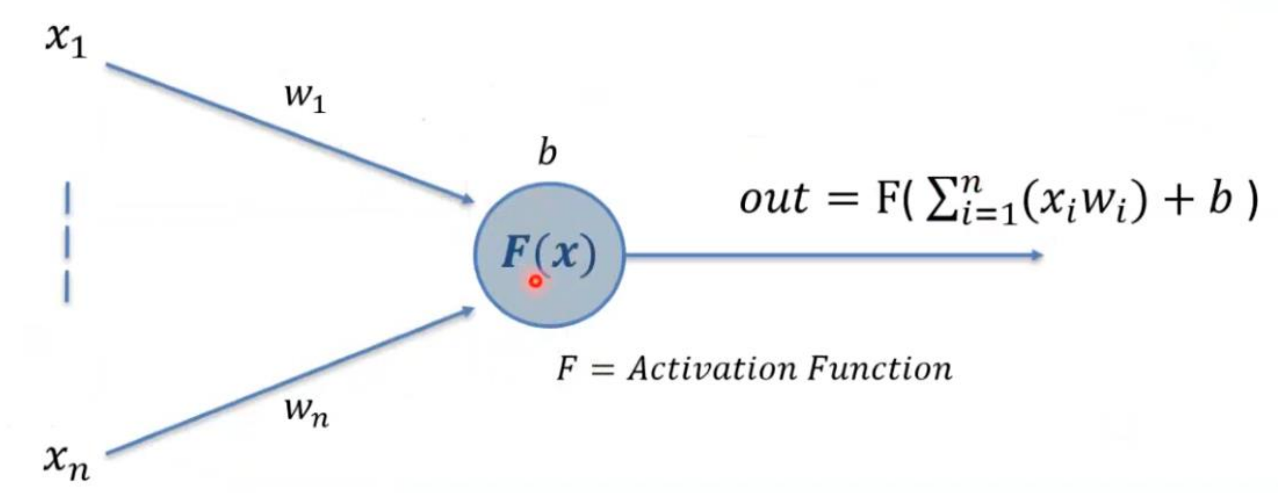


Structure of a neuron

each neuron has different dendrites which can allow it to transfer signals to other neurons, the strength of the signals depends on the dendrite.



This is the base of artificial neural networks, the most basic elements in a neural network is also called neuron and it is made by a series of connections which provide it with the signal coming from previous neurons each weight appropriately.
The neuron makes the weighted sum of the incoming signal and additionally we also ad a constant bias.
Finally we have an activation function F which tells us how to elaborate the total sum before transferring it to the next neuron.



Complete architecture of a neural network

First layer

Receives the input form the photosensor channels each neurons receives the input of one photosensor input layer size is equal to the size of the elements in each observation:

Size of an observation

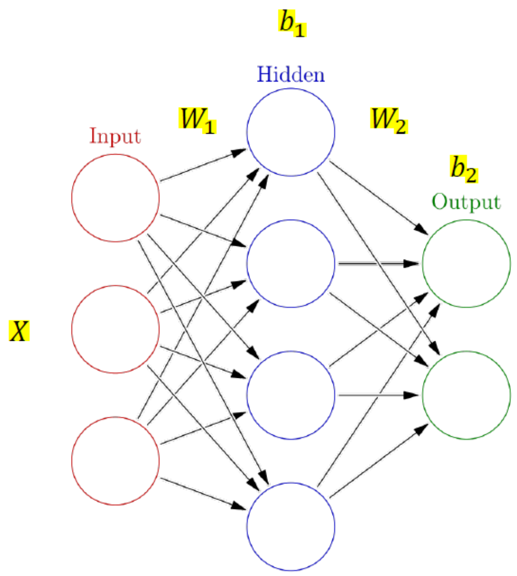
We have 10 photosensors and 500 events the observation matrix will be 500 rows and 10 columns

Hidden layer

Second layer of the neural network connected with the input through different weights, each neuron of the hidden layer is connected to all the neurons of the input.
Size of the weight array will be M number of the hidden layer neurons times D number of the input neurons.
Then we have connections to dent the results to the output again each hidden neurons needs to connect to each output neuron.

Output layer

The number of output neurons is equal to the number of classes in which we want to divide our observations.



$$size(X) = (N, D);$$

$$size(W_1) = (D, M);$$

$$size(b_1) = (M, 1);$$

$$size(W_2) = (M, K);$$

$$size(b_2) = (K, 1);$$

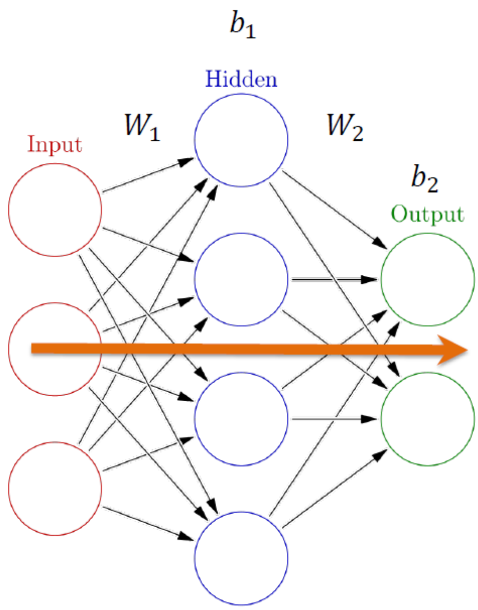
$N = number\ of\ samples;$
$D = number\ of\ features = n°\ input\ neurons;$
$M = number\ of\ neuron\ (hidden\ layer);$
$K = number\ of\ classes = n°\ output\ neuros;$

Operation

The input observations are analyzed and at the output we obtain the probability for each observation to belong to each class.

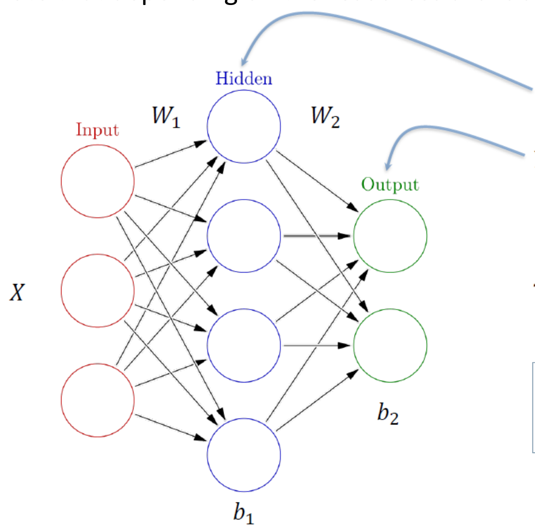
x_1	x_2	x_3
6,548	5,265	3,249
0,649	0,146	8,154
1,986	8,296	4,236

X



Class 1	*Class 2*
0,7	0,3
0,55	0,45
0,2	0,8

Note that depending on the resources available we can process one or more observations in parallel.



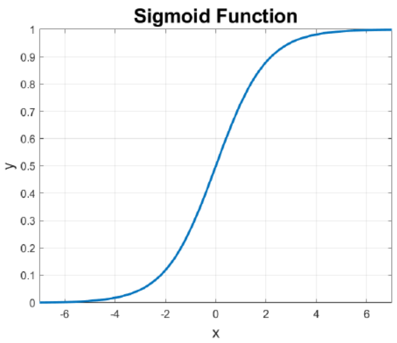
$Z = F(X \cdot W_1 + b_1) \rightarrow size(Z) = (N, M);$

$Y = Softmax(Z \cdot W_2 + b_2) \rightarrow size(Y) = (N, K);$

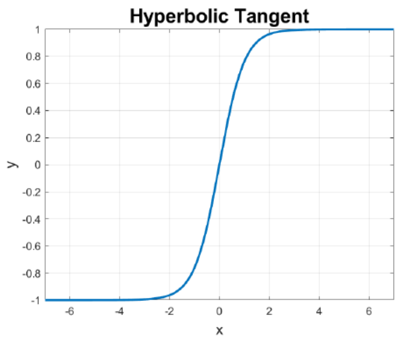
$F = Generic\ Hidden\ Layer\ Activation\ Function;$
$Softmax = Output\ Layers\ Activation\ Function;$

Warning: $X \cdot W_1 + b_1$ is not correct because of dimensions:
$(N, \cancel{D}) \cdot (\cancel{D}, M) + (M, 1) = (N, M) + (M, 1).$
But this works in Python and MATLAB.

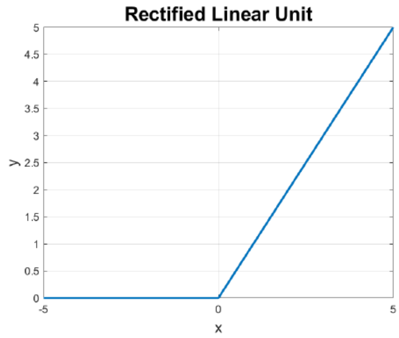
Most common activations functions for hidden layer



$$f(x) = \frac{1}{1+e^{-x}}$$



$$f(x) = \frac{2}{1+e^{-2x}} - 1$$



$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Sigmund function

The higher the value tend to saturate while values close to 0 are transmitted linearly.

Hyperbolic tangent

Similar to the Sigmund function, the result can also be negative

Rectified linear unit

This is one of the most commonly it can only be positive, negative values are set to zero, and it never saturates.

Note

The activations functions can have regions for which they send out the same value even for different inputs

Activation function for output layer

The output layer operates the SoftMax function when we have more than one output class

$$\sigma(z)_i = \frac{e^{Z_i}}{\sum_{j=1}^{K} e^{Z_j}}$$

The activation function uses the value of each neuron in the output layer to define the output value of the neurons and not only that of the specific neuron like in the hidden layer, this is necessary because the output is a probability so the sum of all the outputs of the neurons should be 1.

Example

Example: size(output) = (N, K) = (3,3)

RED	GREEN	BLUE
5	4	2
4	2	8
4	4	1

$\xrightarrow{Softmax}$

RED	GREEN	BLUE
0.705	0.259	0.035
0.0179	0.0024	0.979
0.4878	0.4878	0.024

$$p(RED|x) = \frac{e^5}{e^5 + e^4 + e^2} = 0.705; \quad p(GREEN|x) = \frac{e^4}{e^5 + e^4 + e^2} = 0.259; \quad p(BLU|x) = \frac{e^2}{e^5 + e^4 + e^2} = 0.035;$$

Each row indicates an observation, we can see that the probability for each class is computed with the inputs of all 3 output neurons.

Loss functions

These are the functions which indicates how much the estimation of our network corresponds to the real answer.
We can define the loss function in different ways depending on the problem specific

$$Mean\ Square\ Error = \frac{\sum_{n=1}^{N}(t_n - y_n)^2}{N};$$ Regression

$$Binary\ Cross\ Entropy\ Loss = -\sum_{n=1}^{N}[\,t_n \log y_n + (1-t_n)\log(1-y_n)];$$ Binary Classification

$$Categorical\ Cross\ Entropy\ Loss = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\log(y_{nk});$$ **Multi-class Classification**

Categorical cross entropy loss

Let's consider the probability output y_{nk} of our network (n indicate the index of the event while k indicates the class.

k

RED	GREEN	BLUE
0.705	0.259	0.035
0.0179	0.0024	0.979
0.4878	0.4878	0.024

n

This is t_{nk} : Target Matrix

k

RED	GREEN	BLUE
1	0	0
0	1	0
0	0	1

n

Target matrix

We define the target matrix t_{nk} the probability we can make if we already know the class to which the events belong to so just 1 for the correct class and 0 for the others.

What we do now is the following operation element by element

$$Categorical\ Cross\ Entropy\ Loss = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\log(y_{nk});$$

$$t_{nk}\log(y_{nk})$$

RED	GREEN	BLUE
1	0	0
0	1	0
0	0	1

$\odot$ log(

RED	GREEN	BLUE
0.705	0.259	0.035
0.0179	0.0024	0.979
0.4878	0.4878	0.024

)

The final result will be

$$= -\sum_{n=1}^{N}\sum_{k=1}^{K}\left(\begin{array}{|c|c|c|}\hline \text{RED} & \text{GREEN} & \text{BLUE} \\ \hline -0.3495 & 0 & 0 \\ \hline 0 & -6.032 & 0 \\ \hline 0 & 0 & -3.7297 \\ \hline \end{array}\right) = -\sum_{n=1}^{N}\left(\begin{array}{|c|}\hline \\ \hline -0.3495 \\ \hline -6.032 \\ \hline -3.7297 \\ \hline \end{array}\right) =$$

$$= -(-0.3495 \; -6.032 \; -3.7297) = 10.1112;$$

The final number is the overall quality for all our input observations.

Analysis

When considering the numbers in the result matrix we can see that the values are higher when the probability assigned by the network to the correct result is smaller and vice versa

- Red in the first case has a 70% probability so the value is just $-0{,}3$
- Green in the second case has a probability of $0{,}0024$ so the value is much higher at $-6{,}32$

Training

There are 3 ways to train a neural network.

Supervised learning

We provide data for which we know the results

Unsupervised learning

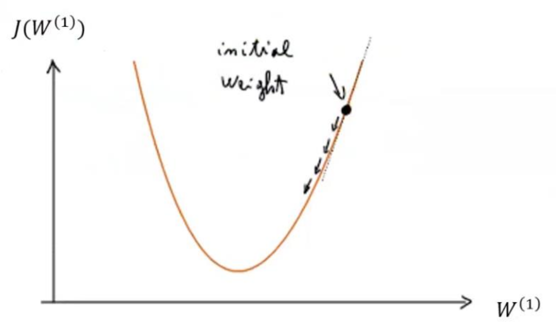
We provide data without knowing the exact result but let the detector organize them as we know that they come approximately from the same position.

Reinforcement learning

We want to find the values of the weights and biases which minimize the loss function so training is a minimization problem, the difficulty of this problem is that we have a huge number of coefficient so we can not use an analytical solution but instead we need to implement numerical processes to optimize the values.

Gradient descent technique

In this solution we calculate ethe gradient of the loss function and tells us in which direction move the coefficients to minimize it, the easiest case is the one where the loss function is a parable with a single clearly defined minimum.



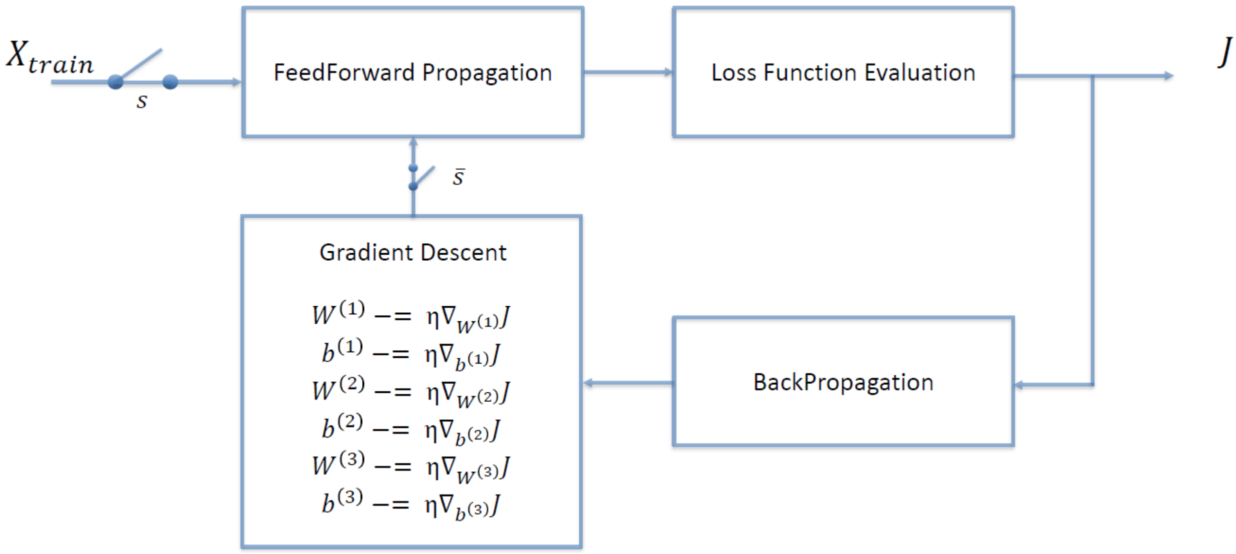
In this algorithm we determine how much the step of each iteration with the coefficient η called learning rate

- If η is too small we might require to many iterations or remain locked in a local minimum
- If η is too small the operation may diverge

Back propagation algorithm

When we want to update a coefficient then we need to evaluate the partial derivative of the function for that parameter for each step of the function.
Increasing the layer number increases exponentially the cost of training the network.

Note that if we use an activation function which saturates we may end up in a region where the derivative is almost flat, this means that the optimization of the weighs becomes stuck



Optimizers

These are the algorithms we use to update our coefficients, we have

Standard gradient descent SGD

The one seen before

Gradient descent with momentum

Upgrade on SGD, we add an inertia so that if we adapt the step depending on the gradient on the previous iterations, so that if we were descending fast and we find a small local minimum we can surpass it

$$ParamMomentum = \beta * ParamMomentum + (1 - \beta) * ParamGradient$$

$$ParamUpdates -= \eta * ParamMomentum$$

RMS prop

Root mean square propagation, in this case we define a cache fore each iteration and we update the parameters utilizing the cache value which is an exponentially weighted average of the gradients of all iterations.

Adam adaptive momentum

We unite RMS and momentum

$$ParamMomentum = \beta * ParamMomentum + (1 - \beta) * ParamGradient$$
$$cache = \rho * cache + (1 - \rho) * (ParamGradient)^2$$

$$ParamUpdates -= \frac{\eta}{\sqrt{cache} + \varepsilon} * ParamMomentum$$

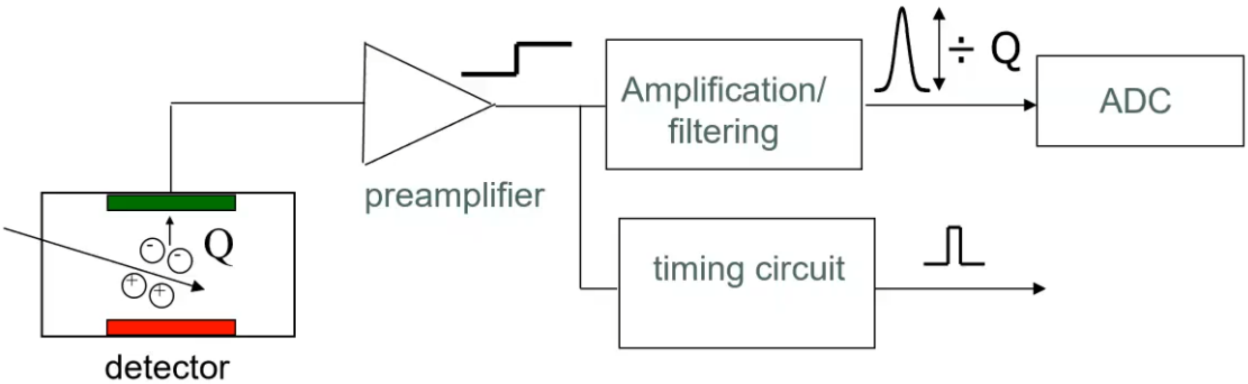
Parte 2 di elettronica

Electronics for the readout and filtering of signals from radiation detectors

A detector provides us with a charge, we need to convert this charge into a voltage amplify it so that we can measure

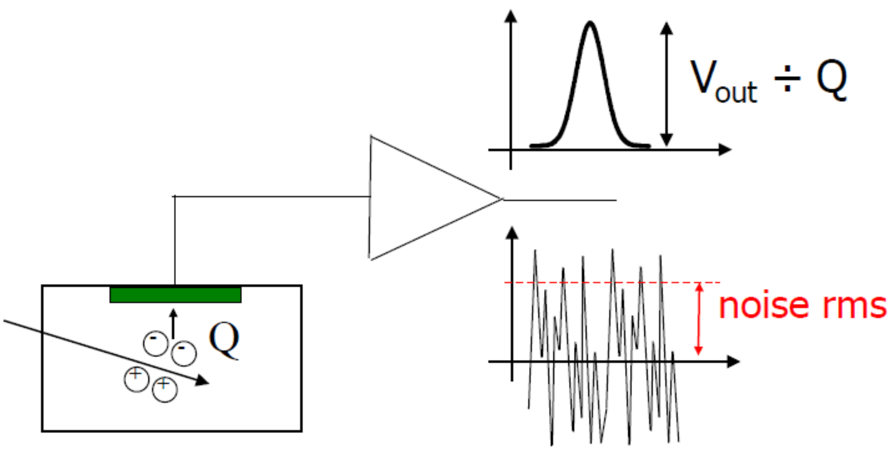
- The energy released by the radiation to for example discriminate form the 511keV of the PET radiation vs other gamma rays
- The arrival time of the event
- The position of interaction: example of the anger camera to determine the arrival postion

a classic circuit chain is the following



Equivalent noise charge

The ENC is the charge that the detector needs to release to produce an output SNR equal to 1, this is useful because it becomes possible for us to express it as the minimum number of electrons that we can detect



$$\boxed{Q = ENC \quad \Rightarrow \quad S/N=1}$$

ENC: the charge to be provided by the detector to produce a S/N at the output of the system equal to 1

Signal to noise ratio formula as number of electrons

$$SNR=NSNSF+ENC2M2$$

Signal

The signal is simply indicated by the number of electrons

Noise

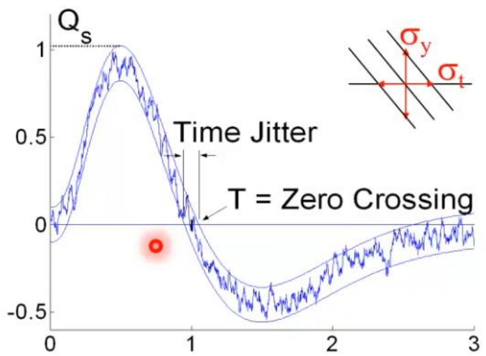
The noise is represented as the square sum of 2 variances

First term

This is the intrinsic fluctuation of the charge.
According with Poisson statistics if we have a number of electron Ns we have a variance equal to the square root of the number of electrons.
Additionally, we have the excess noise factor F.

Second term

This term indicates the noise introduced by the circuit, thanks to the ENC we can sum directly the fluctuation of the charge with the noise contribution as now both express a variation in charge, this would not have been possible if we were to express the noise with its voltage or current spectral density.



Time measurement and jitter

A typical technique to measure the time stamp of a signal is to select a bipolar pulse and measure when the pulse crosses zero.

Since we have noise overlapped to the pulse we have a random shift of the zero crossing timing.
This jitter is particularly significant in PET.

Time jitter value

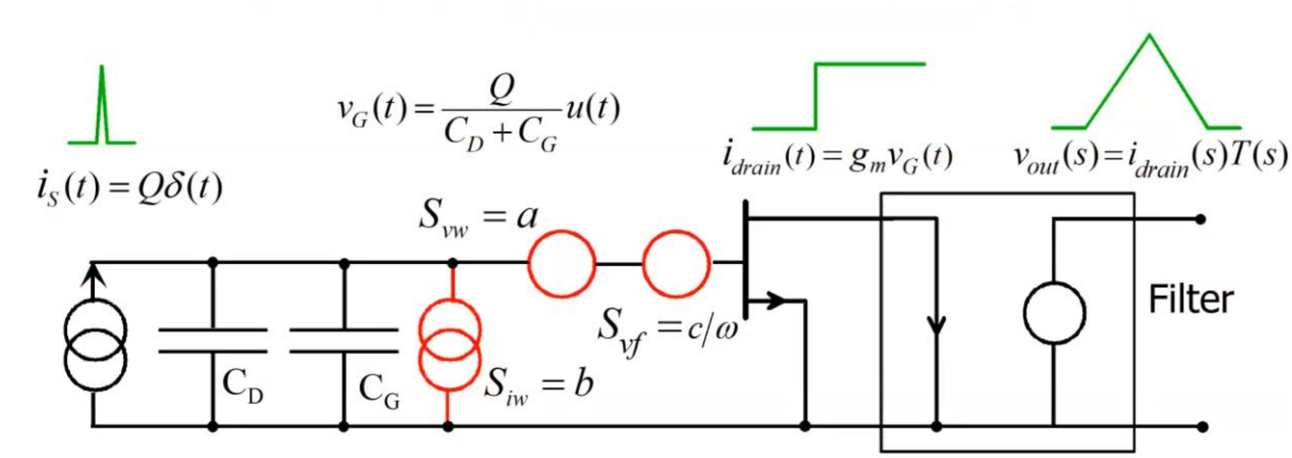
The time jitter is equal to the vertical jitter (the noise) divided by the derivative of the signal

$$t=yy'$$

So to reduce time jitter we can either

- Reduce the noise
- Increase the steepness of the signal

Readout scheme of a voltage pre amplifier



1. In this configuration the charge is not integrated on the feedback capacitor but in the input capacitance (sum of detector capacitance and gate capacitance and possibly the wire capacitance) which creates a Heaviside step.
2. This voltage step turns o the transistor which generates a current proportional to the input voltage
3. The current then passes through a filter which turns the step current into a voltage step

In red we have the noise sources

- An equivalent current generator indicating the shot noise of the dark current and gate current of the transistor (if present)
- And 2 voltage generator (divided in 2 for simplicity one indicates the thermal noise while the other shows the 1\/f noise)

Note

All the circuit is expressed in the time domain except for the filter which is expressed in the Laplace domain

Noise sources in the circuit

Thermal noise of the input FET

SVW=2kTgm=2kTCg1T

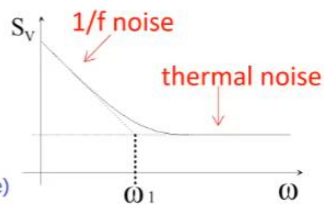
where α≃23 for a short channel FET and T=gmCG

Shot noise of the leakage current of the detector

SiD=qID

Shot noise of the gate current of the FET

SiG=qIG



Thermal noise of a possible resistances

There resistances could be connected at the input for the discharge of the signal and of the leakage current

SiR=2kTR=qIReq

1f noise of the FET

Svf=12Aff=Af

where 1 is the angular frequency of the 1f noise where it matches the thermal noise

Signal to noise ratio at the output of the filter

We need to compute the ratio between the peak of the triangle and the root mean square of the noise

SN=Vso peakvno2

We express with respect to the unitary output signal Vso ut which indicates the value of the output signal for a unitary amount of charge, and the we multiply for the actual charge Q

SN=QMaxVso utvno2

This way we can express the ENC by simply setting the SNR to 1

ENC=vno2MaxVsout

ENC as a function of the noise sources of the filter

We can obtain that the value of the equivalent noise charge for the circuit is given by

$$ENC = (C_D + C_{G2}) \cdot A_1 \cdot a + (C_D + C_{G2}) \cdot A_2 \cdot c + A_3 \cdot b$$

Where

- a is the voltage generator of the series white noise
- b is the parallel current generator of the series white noise
- c is the series voltage generator of the 1f noise

We can see that the terms related to a and c depend on the square of the input capacitance

Parameters of the filters

The important parameters that define the filter are

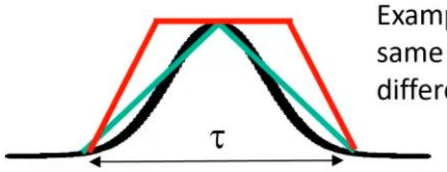
Shaping time

is the shaping time which is the characteristic time related to the width of the pulse at the output of the filter

Filter coefficients

A1,A2,A3 which are the coefficients that determine the shape of the pulse at the output of the filter (they are independent form)

These parameters are necessary because we could have pulses with same shape but different width or vice versa

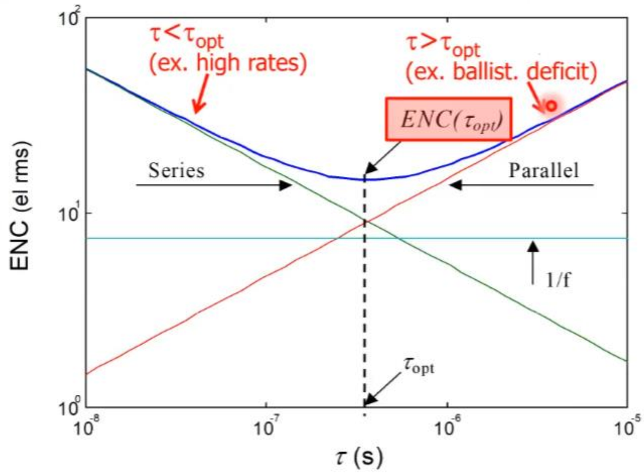


Example: pulses of same width but different shapes

Example: pulses of same shape but different widths

Choice of the optimum shaping time

On the right we have the graph of the ENC as changes.



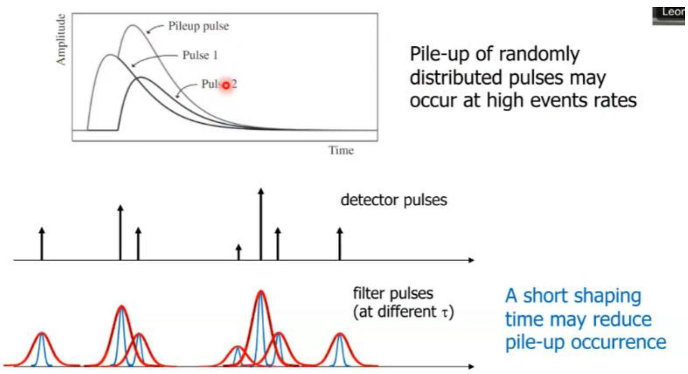
$$ENC^2 = (C_D + C_G)^2 \, a \frac{1}{\tau} A_1 + (C_D + C_G)^2 \, c A_2 + b \tau A_3$$

Since we have that

- The series white noise (a) has an inverse dependance with the shaping time
- The parallel noise b is instead directly dependent with the shaping time
- The 1f noise c is independent from the shaping time

We will have an optimum value of for which the noise transferred is minimized.

Note that often there are conditions where external factors prevents us to use the ideal shaping time, this will of course worsen the performance of the filter but it is sometimes unavoidable

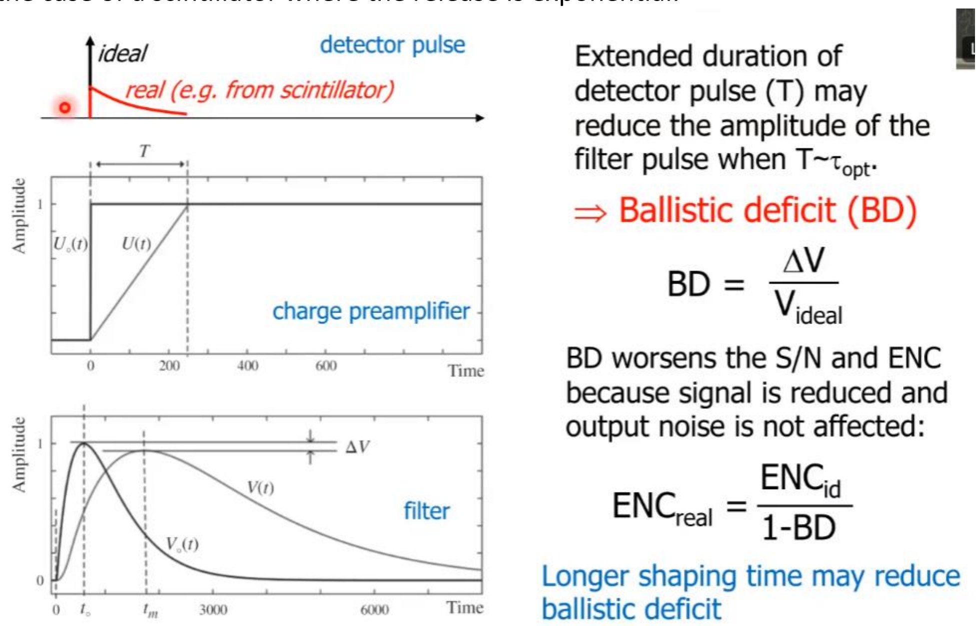


Pile up (shorter shaping time)

An example of a cause that might cause us to select a different shaping time to the optimal one is pile up. Sometimes we may need to select a shorter than opt to avoid having multiple consecutive pulses superimpose to one another making it impossible to read the signal.

No delta detector pulse (longer shaping time)

If the detector release of the photons is too slow, we cannot consider the incoming charge as a delta, this is the case of a scintillator where the release is exponential.



Since the charge is generated in response to the scintillator release it will also present an exponential shape meaning that at the output of the filter we have a slower signal since the input is no longer a Heaviside step.

Since the area of the output pulses remains the same when the duration increases the height of the pulse decreases so we have a reduction in amplitude of the output signal called **ballistic deficit**

BD=VVideal

BD worsens the SN and ENC because the signal is reduced

ENCreal=ENCid1-BD

Longer shaping time may reduce the ballistic deficit as a longer will make the exponential appear like a delta for the filter.

Filter coefficients VS definition of the shaping time

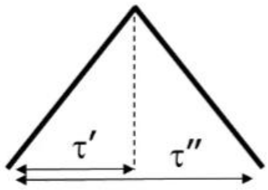
$$ENC^2 = (C_D + C_G)^2 a \frac{1}{\tau} A_1 + (C_D + C_G)^2 2\pi a_f A_2 + b \tau A_3$$

Given a filter, if we change the definition of τ, also the factors A_1, A_2, A_3 changes according to the formula:

$$\boxed{\tau'' = k\tau'}$$

$$\boxed{\begin{aligned} A_1(\tau'') &= k A_1(\tau') \\ A_2(\tau'') &= A_2(\tau') \\ A_3(\tau'') &= \frac{1}{k} A_3(\tau') \end{aligned}}$$

Example: triangular filter



$A_1(\tau'') = 2 \cdot A_1(\tau')$

$A_2(\tau'') = A_2(\tau')$

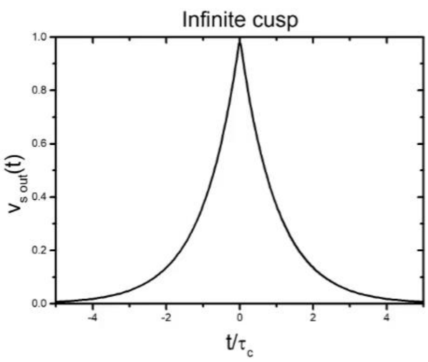
$A_3(\tau'') = 0.5 \cdot A_3(\tau')$

Useful transformation to compare different filters performances at an equivalent shaping time.

There is not a single definition for the shaping time as we can see above we can either select the tiem from the begging of the pulse to the peak ' or the entire duration of the pulse ''.
The parameters follow the shaping time definition as we see above

Optimizing the filter shape

The optimum filter in presence of white noise generators, considering negligible the $\frac{1}{f}$ noise is an infinite cuspid, the problem is that this kind of filter can be obtained only with digital pulse processors, which can not be utilized in medical applications because we have a huge number of channels and each digital filter requires specific and expensive hardware.



$$v_{so}(t) = \exp\left(-\frac{|t|}{\tau_C}\right)$$

$$\tau_c = (C_D + C_G)\sqrt{\frac{a}{b}}$$

In general the optimum filter is the one whose 3 parameters A_1, A_2, A_3 provide the minimum amount of noise at a fixed shaping time.

Here a table with the different parameters and the corresponding shapes, important to note that the waveforms in the table are defined with different definitions of shaping time. To compare them we need to set a single definition like for example the duration of the pulse (in case of infinite shapes we need to set a limit amplitude after which we consider the pulse over).

Table 1
Behaviour of A_1, A_2 and A_3 for different $h(f)$ functions

	Shaping	h (t) Function	A_2	$\sqrt{A_1 A_3}$	$\frac{A_2}{\sqrt{A_1 A_3}}$	A_1	A_3	$\sqrt{\frac{A_1}{A_3}}$
1	indefinite cusp	$e^{t/\tau}$, $e^{-t/\tau}$; O	0.64 $\left(\frac{2}{\pi}\right)$	1	0.64	1	1	1
2	truncated cusp	$k = \tau'/\tau$; $-\tau'$ $-\tau$ O τ τ'; k=1	0.77	1.04	0.74	2.16	0.51	2.06
		k=2	0.70	1.01	0.69	1.31	0.78	1.30
		k=3	0.67	1	0.67	1.31	0.91	1.10
3	triangular	O τ 2τ	0.88 $\left(\frac{4}{\pi}\ln 2\right)$	1.15 $\left(\frac{2}{\sqrt{3}}\right)$	0.76	2	0.67 $\left(\frac{2}{3}\right)$	1.73
4	trapezoidal	O τ 2τ 3τ	1.38	1.83	0.76	2	1.67	1.09
5	piecewise parabolic	O τ 2τ	1.15	1.43	0.80	2.67	0.77	1.86
6	sinusoidal lobe	O τ 2τ	1.22	1.57	0.78	2.47	1	1.57
7	RC-CR	O τ	1.18	1.85	0.64	1.85	1.85	1
8	semigaussian (n = 4)	O 4τ	1.04	1.35	0.77	0.51	3.58	0.38
9	gaussian	$e^{-1/2(t/\tau)^2}$; O	1	1.26	0.79	0.89	1.77	0.71
10	clipped approximate integrator	O 0.5τ	0.85	1.34	0.63	2.54	0.71	1.89
11	bipolar triangular	O τ 2τ	2	2.31	0.87	4	1.33	1.73

Summary for ENC minimization

The components of the formula can be divided into 3 categories

- The noise sources which are independent form the filter
- The shaping time
- The shaping parameters.

$$ENC^2 = (C_D + C_G)^2 \, a \frac{1}{\tau} A_1 + (C_D + C_G)^2 \, c A_2 + b \tau A_3$$

minimize noise sources
(vs. detector parameters)

chose the best filter
(vs. other requirements)

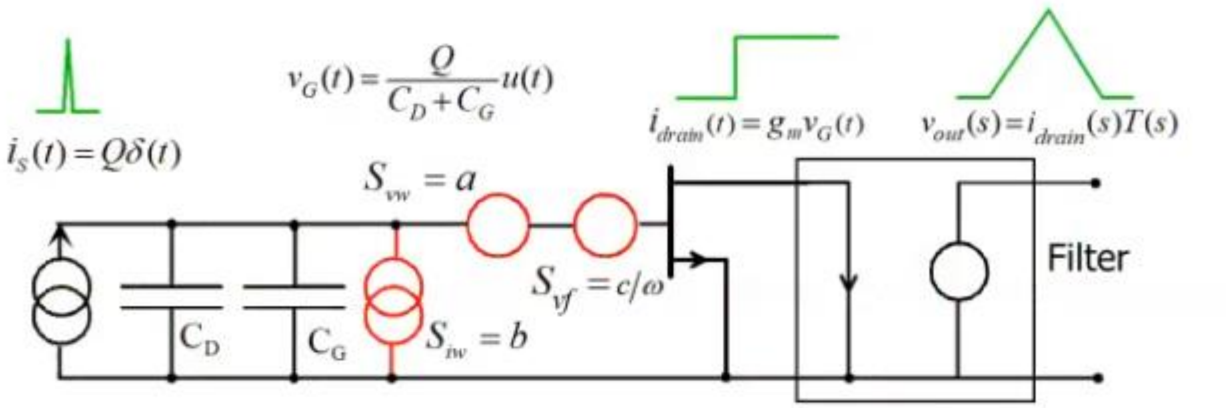
chose the best shaping time
(vs. det./appl. parameters,
e.g. count rate, ballistic def., …)

How to use this formula

1. First we minimize the sources of noise
2. Then we select the best filter (parameter choices)
3. Finally we select the best shaping time

How this formula is calculated

Note that this is not required for the exam
Our circuit is the following



we make the computations moving back and forth between time and frequency domain, below we have the formulas and their equivalent

	Time domain	Laplace domain
Input delta	$i_S(t) = Q\delta(t)$	$I_S(s) = Q$
Step at the filter input	$i_{DRAIN}(t) = \frac{Q}{C_D + C_G} g_m u(t)$	$I_{DRAIN}(s) = \frac{1}{s}\frac{Q}{C_D + C_G} g_m$
Output of the filter as product of the filter transfer function and the input step	$v_{SO}(t) = \frac{Q}{C_D + C_G} g_m \cdot L^{-1}\left[\frac{1}{S}T(s)\right]$	$V_{SO}(s) = \frac{Q}{C_D + C_G} g_m \frac{1}{S} T(s)$

We can repeat to obtain the noise

$$S_{no}(\omega) = \left(a + \frac{2\pi a_f}{|\omega|} + \frac{b}{\omega^2 (C_D + C_G)^2}\right) g_m^2 \left|T(j\omega)\right|^2$$ Power noise spectrum at the filter output

$$\langle v_{no}^2 \rangle = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{no}(\omega)\, d\omega$$ noise at the filter output (rms²)

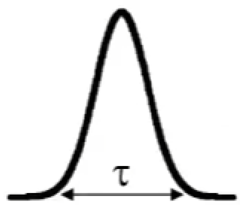
$$\Downarrow$$

$$ENC^2 = \frac{\langle v_{no}^2 \rangle}{Max^2(v_{so\,u\delta})} = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(a + \frac{2\pi a_f}{|\omega|} + \frac{b}{\omega^2 (C_D + C_G)^2}\right) g_m^2 |T(j\omega)|^2\, d\omega}{\frac{g_m^2}{(C_D + C_G)^2} Max^2\left(L^{-1}\left[\frac{1}{s}T(s)\right]\right)}$$

First simplification

We can see that at the denominator we have the maximum of the signal, we arbitrary set it to 1, this can be done without loss of generality because it is always possible to modify the amplifier gain to obtain this result.

Second simplification



The integrals are a function of ω so are connected to the frequency, what we can do is define a new variable $\boldsymbol{x} = \boldsymbol{\omega\tau}$ this variable is now dimensionless and connected to the duration of the pulse.

$$\mathrm{ENC}_{SERIES}^2 = \mathrm{a}(C_D + C_G)^2 \frac{1}{2\pi}\int_{-\infty}^{+\infty} |T(j\omega)|^2 d\omega + \mathrm{a_f} 2\pi (C_D + C_G)^2 \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{|x|}|T(x)|^2 dx + b\tau \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{x^2}|T(x)|^2 dx$$

Now it becomes possible (once selected the filter) to separate the effects of the filter shape and its width (shaping time) we can extract the shaping factors form the equation

$$A_1 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |T(x)|^2 dx$$

$$A_2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{|x|}|T(x)|^2 dx$$

$$A_1 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{x^2}|T(x)|^2 dx$$

Expression in the time domain

Let's start considering the contribution of the series noise (first term of the total equation)

$$\mathrm{ENC}_{SERIES}^2 = \mathrm{a}(C_D + C_G)^2 \frac{1}{2\pi}\int_{-\infty}^{+\infty} |T(j\omega)|^2 d\omega = \mathrm{a}C_T^2 \frac{1}{2\pi}\int_{-\infty}^{+\infty} |T(j\omega)|^2 d\omega =$$

$$= \boldsymbol{aC_T^2 \frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(j\omega)|^2 \omega^2 d\omega}$$

Where we have indicated with $H(s)$ the product of the Laplace transfer function of the charge preamplifier which is an integrator and thus has a transfer function $\frac{1}{s}$ and of the filter which has a transfer function $T(s)$, for this reason in the last term we have ω^2.

We know that by definition $H(s)\cdot s \to h'$ so additionally the Parseval theorem tells us that the integral of the functions squared in the frequency domain is equal to the integral of the corresponding function in the time domain so we can write

$$\boldsymbol{\mathrm{ENC}_{SERIES}^2 = aC_T^2 \int_{-\infty}^{\infty} [h'(t)]^2 dt}$$

Then we define a new dimensionless variable $t' = \frac{t}{\tau}$ which indicates the time normalized to the shaping time τ.

The formula becomes

$$\mathrm{ENC}_{SERIES}^2 = aC_T^2 \frac{1}{\tau}\int_{-\infty}^{\infty} [h'(t')]^2 dt'$$

From this we can extract the A_1 coefficient

$$\boldsymbol{A_1 = \frac{1}{2\pi}\int_{\infty}^{+\infty} [h'(t')]^2 dt'}$$

Conclusion

The A_1 coefficient is related to the area of the derivative of $h(t)$, this means that this factor prefers smoother pulses because a smoother pulses implies lower derivatives to be integrated in the formula. (remember that if the factor increases than also the ENC increases so we want to keep them to a minimum)

Repeat for the parallel noise and obtain A_3

$$ENC^2{}_{parallel} = b1/2\pi \int_{-\infty}^{+\infty} \frac{|T(j\omega)|^2}{\omega^2} d\omega =$$

$$= b1/2\pi \int_{-\infty}^{+\infty} |H(j\omega)|^2 d\omega = [Parseval\ theorem] = b \int_{-\infty}^{+\infty} [h(t)]^2 dt$$

$$t' = t/\tau \qquad dt = dt' \cdot \tau \qquad ENC^2{}_{parallel} = b\,\tau \int_{-\infty}^{+\infty} [h(t')]^2 dt'$$

$$A_3 = 1/2\pi \int_{-\infty}^{+\infty} \frac{|T(j\omega)|^2}{\omega^2} d\omega = \int_{-\infty}^{+\infty} [h(t')]^2 dt'$$

$\Rightarrow A_3$ is related to the area of h(t)

We obtain that A_3 is instead proportional to the area of the pulse, so if we compare 2 filters with the same τ the one with the larger area will have a larger A_3.

Repeat for $\frac{1}{f}$

In this case the computation is harder and we will obtain a dependance on the half derivative

$$ENC^2{}_{1/f} = afC_T{}^2 \int_{-\infty}^{+\infty} \frac{|T(j\omega)|^2}{|\omega|} d\omega = a_f C_T{}^2 \int_{-\infty}^{+\infty} |\omega||H(j\omega)|^2 d\omega$$

$$|\omega| = \sqrt{j\omega}\sqrt{-j\omega}$$

$$|\omega||H(j\omega)|^2 = \sqrt{j\omega}H(j\omega)\overline{\sqrt{j\omega}H(j\omega)} = Z(j\omega)\overline{Z(j\omega)} = |Z(j\omega)|^2$$

$with\ Z(j\omega) = \sqrt{j\omega}H(j\omega)$

$we\ will\ apply\ the\ Parseval\ theorem\ to\ Z(j\omega)$

$note\!:\ the\ counterpart\ of\ operator\ j\omega\ in\ the\ time\ domain\ is\ the\ derivative\ of\ order\ 1\!:$

$$F^{-1}(j\omega H(j\omega)) = h'(t)$$

$$Z(j\omega) = \sqrt{j\omega}H(j\omega)$$

$the\ counterpart\ of\ operator\ \sqrt{j\omega}\ in\ the\ time\ domain\ is\ the\ derivative\ of\ order$ ½

$$\Rightarrow F^{-1}\left(\sqrt{j\omega}H(j\omega)\right) = h^{(1/2)}(t)$$

$$ENC_{1/f} = afC_T{}^2 \int_{-\infty}^{+\infty} |\omega||H(j\omega)|^2 d\omega = [Parseval\ theorem] =$$

$$= a_f C_T{}^2 2\pi \int_{-\infty}^{+\infty} [h^{(1/2)}]^2 dt$$

considering:

$$ENC^2 = (C_D + C_G)^2 a \frac{1}{\tau} A_1 + (C_D + C_G)^2 2\pi a_f A_2 + b\tau A_3$$

$$\Rightarrow A_2 = \int_{-\infty}^{+\infty} [h^{(1/2)}]^2 dt$$

(it can be verified that the integral does not depend from the time scale, i.e. $t' = t/\tau$)

$\Rightarrow A_2$ is related to the area of $h^{1/2}(t)$

The derivative of half order can be computed as

$$Z(j\omega) = \sqrt{j\omega} H(j\omega) = \frac{1}{\sqrt{j\omega}} j\omega H(j\omega)$$

$$h^{(1/2)}(t) = F^{-1}\left(\frac{1}{\sqrt{j\omega}}\right) * F^{-1}\big(j\omega H(j\omega)\big) = g(t) * h'(t)$$

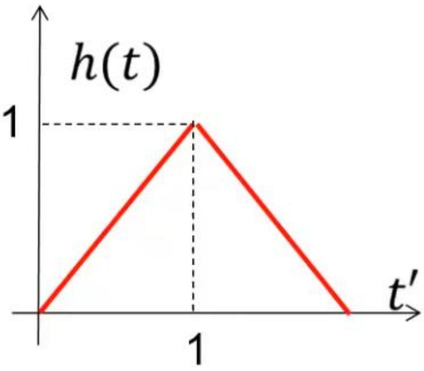
$$with \quad g(t) = \begin{cases} 1/\sqrt{\pi t}, & t > 0, \\ 0, & t < 0, \end{cases}$$

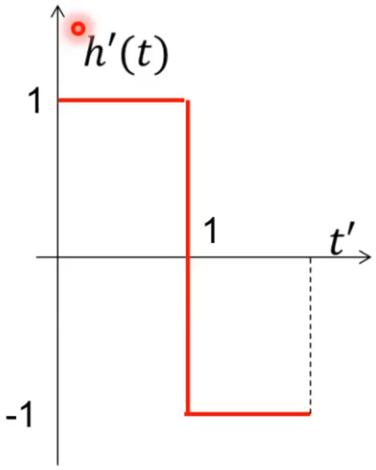
Summary

- A_1 is the integral of the derivative of first order so it increases if the signal is steeper
- A_3 is the integral of the derivative of zero order so it increases if the signal has a larger area
- A_2 is the integral of the derivative of half order

Example

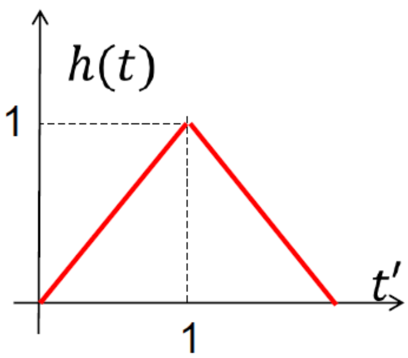
We have a triangle and we want to calculate the shaping factors A_1, A_2, A_3 of the triangle.

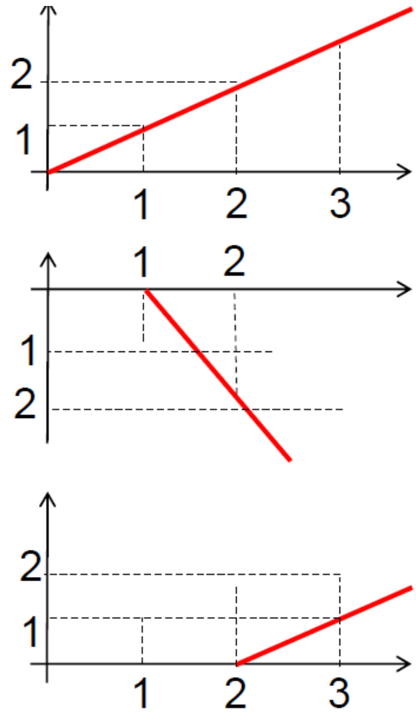




We start from A_3 which is equal to the integral of the square of the time derivative of the triangle

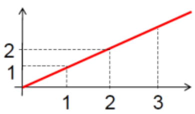
To compute A_2 we divide the triangle in 4 ramps





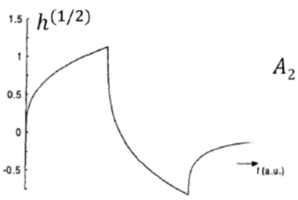
The total half derivative will be the sum of the half derivatives of the ramps

calculation of $h^{(1/2)}$ *for a single ramp*:



$$\begin{cases} \sqrt{4\frac{t}{\pi}}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

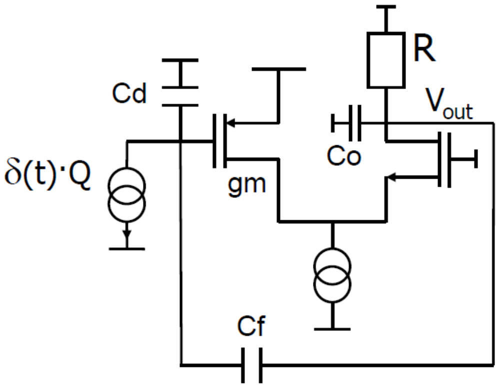
calculation of $h^{(1/2)}$ *for the triangular filter*:



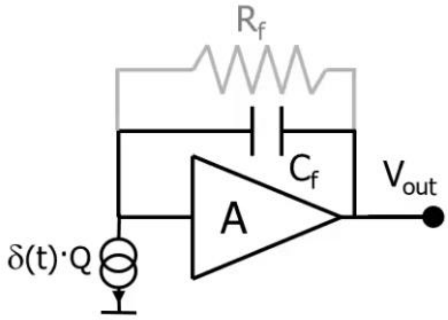
$$A_2 = \int_{-\infty}^{+\infty} [h^{(1/2)}]^2 dt = \frac{4}{\pi} ln2 = 0.88$$

Charge preamplifier

Basic cascode configuration



This is the simplest configuration of a charge preamplifier
We have a common source transistor flowing in a transistor in folded cascode configuration.
This way the current is forced out of the drain of the cascode transistor into a resistance R, this is what provides the gain to allow for the negative feedback.



We connect a capacitor to create negative feedback and thus an integrator, a resistor is placed in parallel with the capacitor to allow for it to discharge more quickly.

- V_{out}: voltage step of amplitude Q/Cf
- A = -gm·R, tau_{pole} = R·Co
- GBWP ~ A/tau_{pole}= gm/Co
- Gloop = A·Cf/(Cd+Cf)
- tau_{cl} of the closed loop dominant pole
  ~ tau_{pole}/Gloop=Co/gm·(Cd+Cf)/Cf

Value of C_f

We want C_f to have a good closed loop gain, however we need a big G_{loop} so we can not lower it too much as the G_{loop} would also decrease

Why do we not use an OPAMP with an input differential pair

If we were to use an OPAMP or in general an amplifier with a differential input than the noise introduced would also be double as we would have 2 separate inputs applying noise without it being mitigated by the feedback.

1/f noise model

Let's consider a MOSFET, there are 2 models we can consider to represent the noise

$$S_{I(\Delta\mu)} = \left(\frac{q\sqrt{2\mu}\,\alpha_H}{\sqrt{C'_{ox}}}\right)\left(\frac{1}{\sqrt{W L^3}}\right) I^{3/2}\frac{1}{f} \quad [\mathrm{A}^2/\,\mathrm{Hz}]$$

Hooge model
(carrier mobility fluctuations)

$$S_{I(\Delta N)} = \left(\frac{q^2 \mu\, kT\, N_T}{\gamma C'_{ox}}\right)\left(\frac{1}{L^2}\right) I\frac{1}{f} \quad [\mathrm{A}^2/\mathrm{Hz}]$$

McWorther model
(channel conductivity modulation caused by trapping/detrapping)

Hooge model

In this model the $\frac{1}{f}$ noise is attributed to fluctuations in the mobility in the MOSFET channel.

We have a dependency with the current $I^{\frac{3}{2}}$

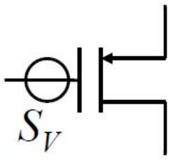
Mc Worther model

In this model the $\frac{1}{f}$ noise is attribute to the presence of traps in the MOSFET crystal structure which capture and release carriers causing fluctuations in the current flow in the channel
There is a dependency with he current I

When we refer the formula to the inputs we can see how in the Mc Worther model the dependance on the current disappears, while in the Hooge formula we have a dependance with the square root

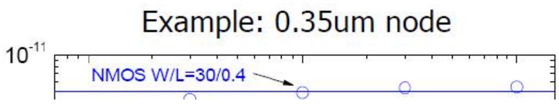
$$g_m^2 = 2\mu C'_{ox} \frac{W}{L} I$$

$$S_{V(\Delta\mu)} = \left(\frac{q\, \alpha_H}{\sqrt{2\mu}\sqrt{C'^3_{ox}}} \right) \left(\frac{1}{\sqrt{W^3 L}} \right) \sqrt{I} \frac{1}{f} \quad [\mathrm{V^2/Hz}]$$



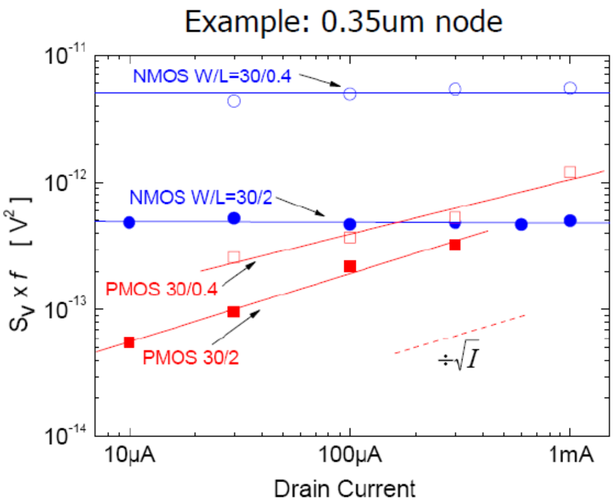
$$S_{V(\Delta N)} = \left(\frac{q^2 k T N_T}{2\gamma C'^2_{ox}} \right) \left(\frac{1}{W L} \right) \frac{1}{f} \quad [\mathrm{V^2/Hz}]$$

$$S_{V(\Delta N)} = \left(\frac{K_F}{C'_{ox} W L} \right) \frac{1}{f} = \left(\frac{K_F}{C_G} \right) \frac{1}{f} \quad [\mathrm{V^2/Hz}]$$



To decide which one is correct we need to look at experimental data, for example below we see a technology for which we can use

- Mc Worther for the nMOS
- The Hooge formula for the pMOS



$$S_{V(\Delta N)} = \left(\frac{K_F}{C'_{ox} W L} \right) \frac{1}{f} = \left(\frac{K_F}{C_G} \right) \frac{1}{f} \quad [\mathrm{V^2/Hz}]$$

- PMOS follows $\Delta\mu$
- NMOS follows ΔN

But McWorther model seems more suitable for both NMOS and PMOS for more scaled technologies:

$$S_V = K_f / C_G \cdot 1/f$$

Rewriting the Mc Worther formula

$$S_{V(\Delta N)} = \left(\frac{q^2\,k\,T\,N_T}{2\,\gamma\,C'^2_{ox}}\right)\left(\frac{1}{W\,L}\right)\frac{1}{f} \quad [\mathrm{V^2/Hz}]$$

$$S_{V(\Delta N)} = \left(\frac{K_F}{C'_{ox}\,W\,L}\right)\frac{1}{f} = \left(\frac{K_F}{C_G}\right)\frac{1}{f} \quad [\mathrm{V^2/Hz}]$$

we see that the $\frac{1}{f}$ noise is inversely proportional to the gate capacitance this create a problem since we wish to minimize the gate capacitance to improve the ENC which instead increases if we increase C_G we are going to need to find a compromise.

Optimizing the input MOSFET

Let's consider the part of the ENC formula connected with the $\frac{1}{f}$ noise

$$\boldsymbol{ENC^2_{1/f} = \frac{2\pi}{q^2} A_2 A_f 2\pi (C_{IL} + C_G)^2}$$

- C_{IL} indicates the total capacitance (detector feedback test parasitic)
- A_F indicates the parameters specific to the $\frac{1}{f}$ model we use so we can substitute and obtain

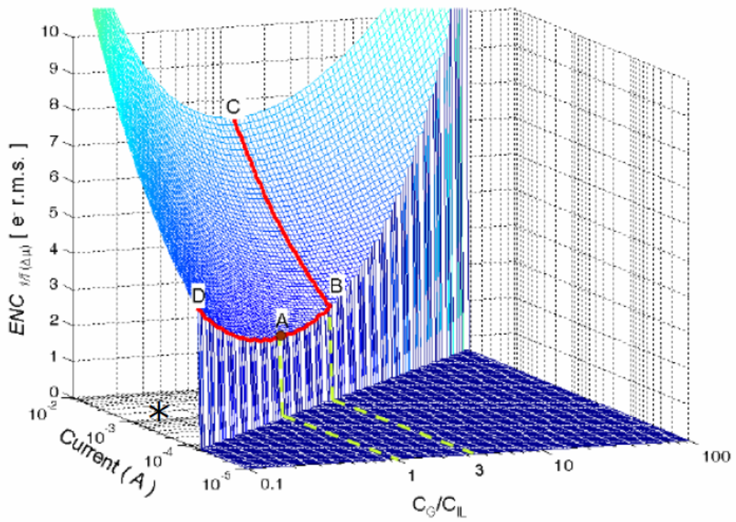
Let's focus on the first formula

- L is at the numerator so we want to minimize it
- I the current (at least in the Hooge model) needs to be minimized, the minimum current is the current for which the transistor is at the edge of strong inversion

$$ENC^2_{1/f(\Delta\mu)} = 2\pi\,A_2\left(\frac{\alpha_H\,L}{q\sqrt{2\mu}}\right)\left[\frac{(C_{IL}+C_G)^2}{\sqrt{C_G{}^3}}\right]\sqrt{I}$$

$$ENC^2_{1/f(\Delta N)} = 2\pi\,A_2\left(\frac{k\,T\,N_T}{2\,\gamma\,C'_{ox}}\right)\left[\frac{(C_{IL}+C_G)^2}{C_G}\right]$$

Since we already set $L = L_{min}$ what we can do now is set either W or C_G.
This is more complex since C_G appears both at the numerator and at the denominator, this can be done utilizing the following plot (**see the results in the slide text**)



If we are free to choose I:
⇒ absolute minimum ($I=I_{min}$)
for $C_G=C_{IL}$

If we are not free to choose I
(I fixed, e.g. for Gloop and speed):
⇒ minimum for I fixed
for $C_G=3C_{IL}$

We might not utilize the minimum current because this lowers the transconductance and thus the loop gain, in this case we are limited to the red line in the graph and thus point B in the plot.

Let's consider all of the noise contribution

Parallel noise

The parallel noise is not related to the transistor but only on

- The dark current
- The transistor gate current which should be zero in a MOSFET

So these choices do not affect the parallel noise

$$ENC_{wp}^2 = \frac{A_3}{q^2} S_I \tau$$

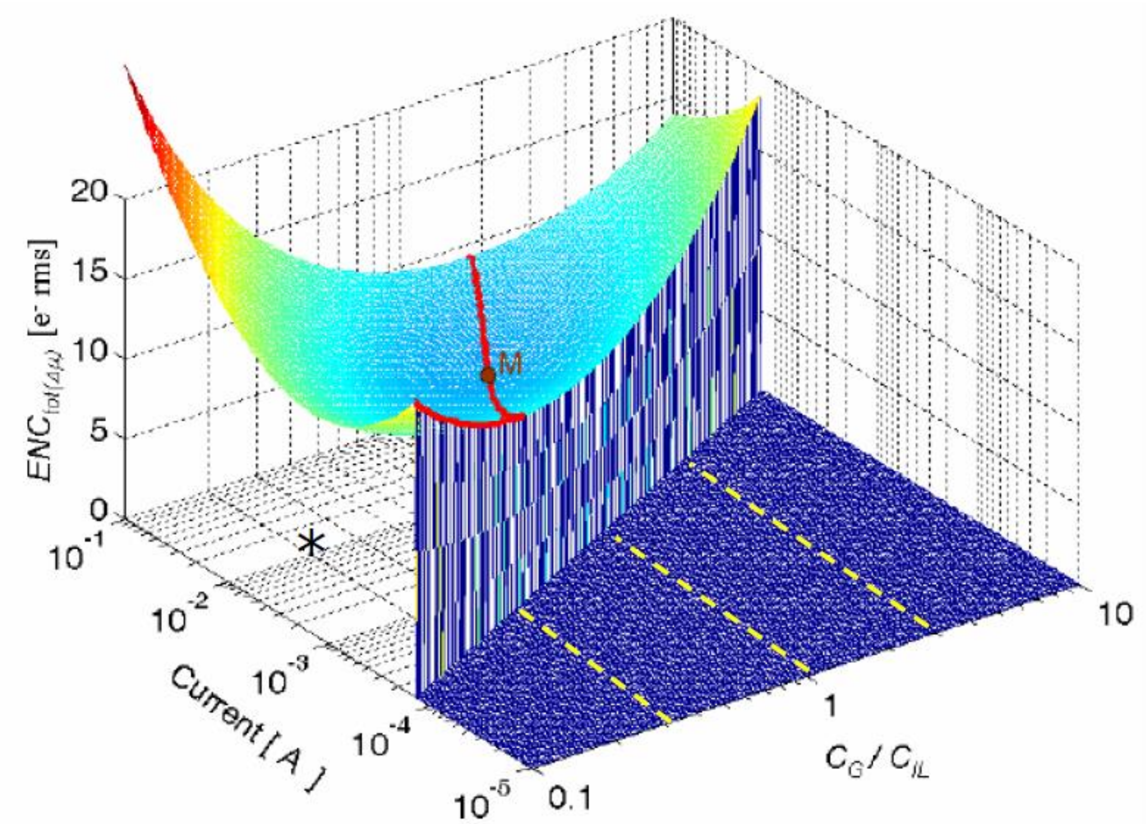
Series white noise

This noise is dependent on the transistor thermal noise so the choices we made affect its value directly

$$\rightarrow ENC_{ws}^2 = \frac{A_1}{q^2} S_V (C_{IL} + C_G)^2 \frac{1}{\tau} \qquad S_V = \beta \frac{4kT}{g_m} = \frac{\beta\, 4kT}{\sqrt{2\mu C'_{ox} \frac{W}{L}} \sqrt{I}}$$

If we minimize the current then we maximize the white parallel noise since the transconductance is reduced and thus the resistance and its thermal noise is increased.

Now the new minimum is not at the minimum current but at an average point M



Note that in this condition **the best condition for the size of the C_G is still a ratio 1:1 with the detector capacitance in any case so the matching condition remains valid**.

We can still not be able to use this current and require

- A larger current if we need it for the loop gain in this case the optimum gate capacitance is between

$$C_{IL} < C_G < 3C_{IL}$$

- A smaller current if we have power or thermal limitations in this case the gate capacitance optimum value is between

$$\frac{1}{3C_{IL}} < C_G < C_{IL}$$

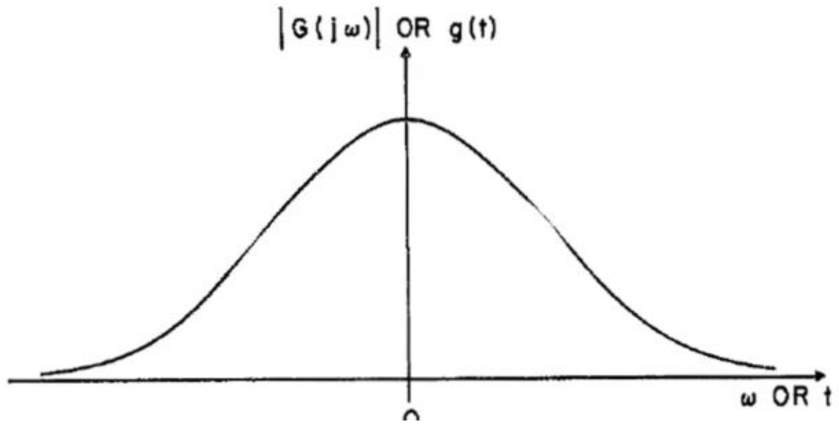
Overall the gate capacitance should be between

$$\boldsymbol{\frac{1}{3C_{IL}} < C_G < 3C_{IL}}$$

The semi gaussian filter with conjugate complex poles

We want to approximate a gaussian filter which is a good approximation of the optimum filter but can actually be implemented in practice.
In particular we can implement a semi gaussian filter which can be implemented with $R - C$ networks

$$f(t) = a_0\, e^{-\frac{1}{2}t^2/\sigma^2} \qquad F(\omega) = a_0\sqrt{(2\pi)}\,\sigma\, e^{-\frac{1}{2}\sigma^2\omega^2}$$



We want to approximate the Gaussian with an almost identical approximation but with a more easily implementable filter, so we are looking for a transfer function $H(s)$ with this form

$$H(s) = \frac{H_0}{Q(s)}$$

Where

- H_0 is a constant, we do not need to have any zeroes
- $Q(s)$ is a polynomial

we want the modulus of our function to be equal to that of the Gaussian $F(\omega)$ so we use the definition of module and write

$$H(j\omega) \cdot H(-j\omega) = |H(j\omega)|^2 = |F(\omega)|^2$$

We can substitute with the expression for the Gaussian $a_0\sqrt{2\pi}\sigma e^{-\frac{1}{2}\sigma^2\omega^2}$ and get

$$\frac{H_0}{Q(j\omega)} \cdot \frac{H_0}{Q(-j\omega)} = \left[a_0\sqrt{2\pi}\sigma e^{-\frac{1}{2}\sigma^2\omega^2}\right]^2$$

$$\frac{H_0^2}{Q(j\omega)Q(-j\omega)} = (a_0\sigma)^2 2\pi e^{-\sigma^2\omega^2}$$

Which becomes

$$Q(j\omega) \cdot Q(-j\omega) = \left(\frac{H_0^2}{a_0\sigma}\right)^2 \frac{1}{2\pi} e^{\sigma^2\omega^2}$$

We substitute with $s = j\omega$ and we obtain

$$Q(j\omega)\cdot Q(-j\omega)=\left(\frac{H_0^2}{a_0\sigma}\right)^2\frac{1}{2\pi}e^{-\sigma^2\omega^2}\qquad only\ change\ the\ sign\ of\ the\ exponential$$

Finally we normalize and substitute with the dimensionless variable $p=\sigma s$

$$\boldsymbol{Q(p)\cdot Q(-p)=e^{-p^2}}$$

This problem has not a precise mathematical solution, so we utilize a Taylor approximation.

$$Q(p)\cdot Q(-p)=1-p^2+\frac{p^4}{2!}-\frac{p^6}{3!}+\ldots+(-1)^n\frac{p^{2n}}{n!}$$

we need to decide at which point to stop the Taylor series:

$\boldsymbol{n=1}$

If we stop at $1-p^2$ which is a very rough approximation, then we just need to factorize and we obtain

$$Q(p)\cdot Q(-p)=1-p^2=(1+p)\,(1-p)$$

$$\Rightarrow\ Q(p)=1+p$$

Remember that $Q(p)$ is the denominator so the equation is

$$H(p)=\frac{1}{1+p}$$

If we make the reverse transformation we obtain an exponential, obviously this is not a good approximation of the gaussian

$\mathbf{n=2}$

Now the solution is more complex, and the denominator becomes an equation with 2 complex conjugated poles

Example: n=2

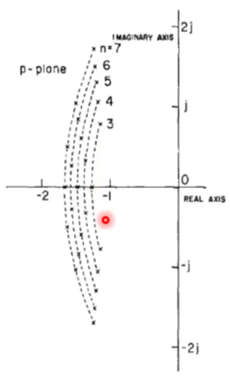
$$Q(p)\cdot Q(-p)=1-p^2+\frac{p^4}{2!}=\tfrac{1}{2}[\sqrt{2}+\sqrt{(2+2\sqrt{2})}\,p+p^2]\times$$
$$\times(\sqrt{2}-\sqrt{2+2\sqrt{2})}\,p+p^2).$$

$$\Rightarrow\ Q(p)=\frac{1}{\sqrt{2}}(\sqrt{2}+\sqrt{(2+2\sqrt{2})}\,p+p^2)$$

We are closer but the solution is strongly asymmetric.

Note that as we expand more and more the Taylor expansions the more poles we will need to obtain the solution

for n≥3 the zeros of Q(p) (poles of the filter) are determined numerically:



Pole locations of the Gaussian filters.

	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
A_0	1.2633573		1.4766878		1.6610245
A_1	1.1490948	1.3553576	1.4166647	1.5601279	1.6229725
W_1	0.7864188	0.3277948	0.5978596	0.2686793	0.5007975
A_2		1.1810803	1.2036832	1.4613750	1.4949993
W_2		1.0603749	1.2994843	0.8329565	1.0454546
A_3				1.2207388	1.2344141
W_3				1.5145343	1.7113028

Note: A_i and W_i are independent from the pulse width (shaping time) but depend just on the filter order.

The general formula for the semi gaussian is

$$H(s) = \frac{A_0 \prod_{i=1}^{k} \{A_i^2 + W_i^2\}}{(\sigma s + A_0) \prod_{i=1}^{k} \{(\sigma s + A_i)^2 + W_i^2\}}$$

n. odd poles

1 real pole

couples of complex conjugate poles

$$H(s) = \frac{\prod_{i=1}^{k} \{A_i^2 + W_i^2\}}{\prod_{i=1}^{k} \{(\sigma s + A_i)^2 + W_i^2\}}$$

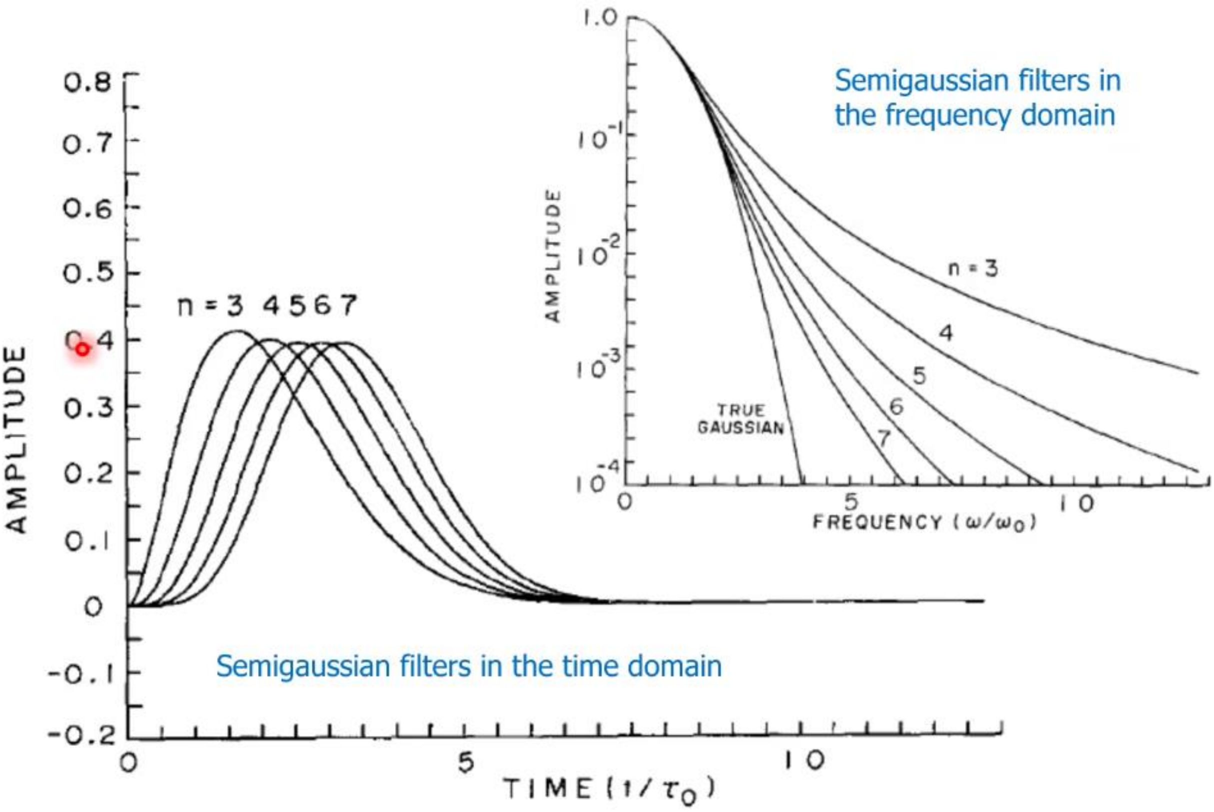
n. even poles

$s = j\omega$

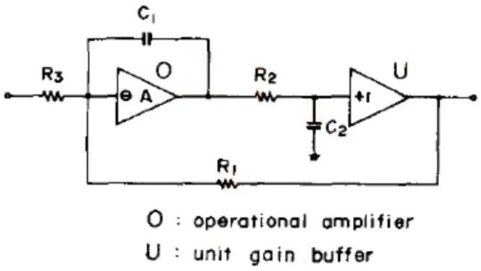
σ of the Gaussian

couples of complex conjugate poles

Note that at the numerator we do not have zeroes but we do have the coefficients while at the denominator we have the sigma.



Elementary cell for complex poles implementation



With this configuration the transfer function is the following

$$H(s) = -\frac{R_1}{R_3}\frac{1}{1 + sC_1R_1 + s^2C_1R_1C_2R_2}$$

Wem have 2 complex conjugated poles

$$\frac{1}{2C_2R_2}\left[-1 \pm j\sqrt{\frac{4C_2R_2}{C_1R_1} - 1}\right]$$

We can then select the components to match the value indicated by the Taylor series, note that in the formula we do not have only the capacitance and the resistance but also τ_0 which indicates the shaping time of the pulse.

This means hat if we change the shaping time we need to change the product C_2R_2 to maintain the same coefficient value for A_i and W_i

$$A_i = \frac{\sigma_0\tau_0}{2C_2R_2},$$

$$W_i = \left(\frac{\sigma_0\tau_0}{2C_2R_2}\right)\sqrt{\left(\frac{4C_2R_2}{C_1R_1} - 1\right)}$$

Pole locations of the Gaussian filters.

	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
A_0	1.2633573		1.4766878		1.6610245
A_1	1.1490948	1.3553576	1.4166647	1.5601279	1.6229725
W_1	0.7864188	0.3277948	0.5978596	0.2686793	0.5007975
A_2		1.1810803	1.2036832	1.4613750	1.4949993
W_2		1.0603749	1.2994843	0.8329565	1.0454546
A_3				1.2207388	1.2344141
W_3				1.5145343	1.7113028

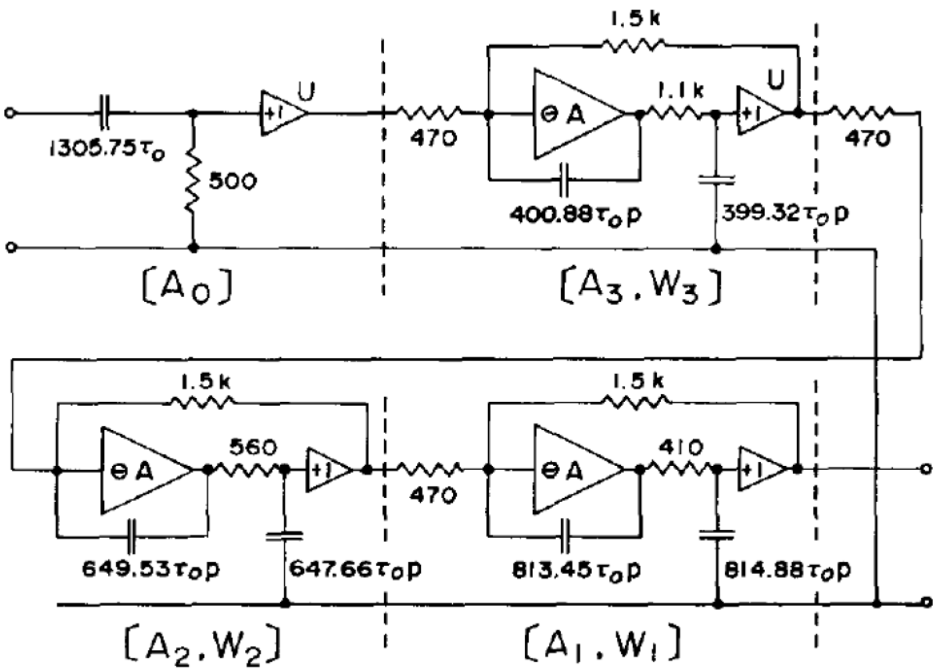
$\sigma_0 = (e/\sqrt{2\pi})$

$\tau_0 = 0.9221\sigma$

Note: the change of the shaping time can be done by changing values of the components, keeping constant A_i and W_i.

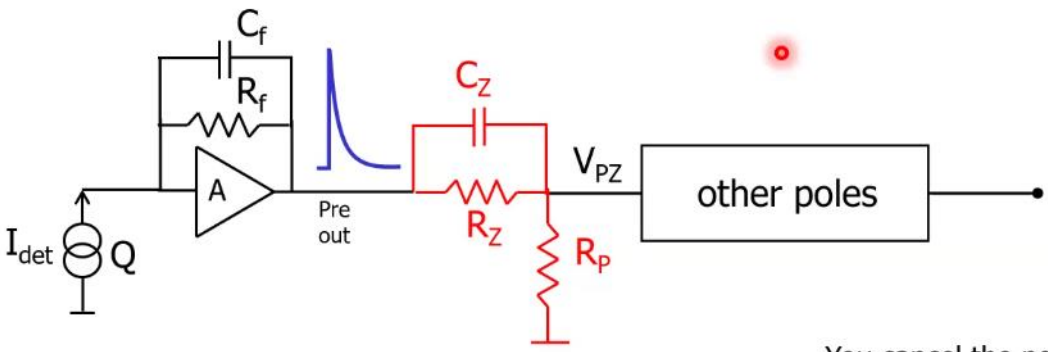
Example of a 7th order filter

has we can see each stage present different sizing to implement the correct parameters, the real pole is created by a simple RC circuit with a buffer connecting it to the second stage to decouple its time constant form the input resistance of the second stage (note we might use a real integrator since we still use an OPAMP and in the real integrator the OPAMP is in the feedback so its non-linearities are compensated.



Network for pole and zero compensation

We want to compensate the pole introduced by the charge preamplifier otherwise the signal coming into the filter will not be what we expect as a delta would become a step.
To do so we add a derivator like in the circuit below



$$V_{PZ} = Q\,\frac{R_f}{\cancel{1+sC_fR_f}}\;\frac{\cancel{1+sC_ZR_Z}}{1+sC_ZR_Z//R_P}\;R_P/(R_Z+R_P)$$

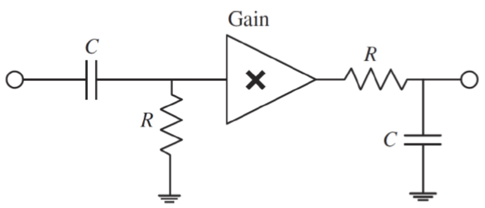
$$\sim R_P$$

(choosing $R_Z >> R_P$)

You cancel the pole of the preamplifier (e.g. constrained by noise requirements) to obtain another pole (e.g. constrained to be the first pole of the filter)

RC CR filters

This is an alternative configuration that can be used to create a filter that while not being a real semi Gaussian filter, it does functionally approximate the Gaussian filter well enough.



With this configuration we get a zero in the origin to compensate for the pole of the recharge amplifier and 2 real poles at the same position

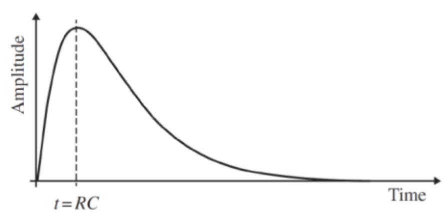
$$\frac{sCR}{(1+sCR)^2}$$

Time domain response

The time domain output response is

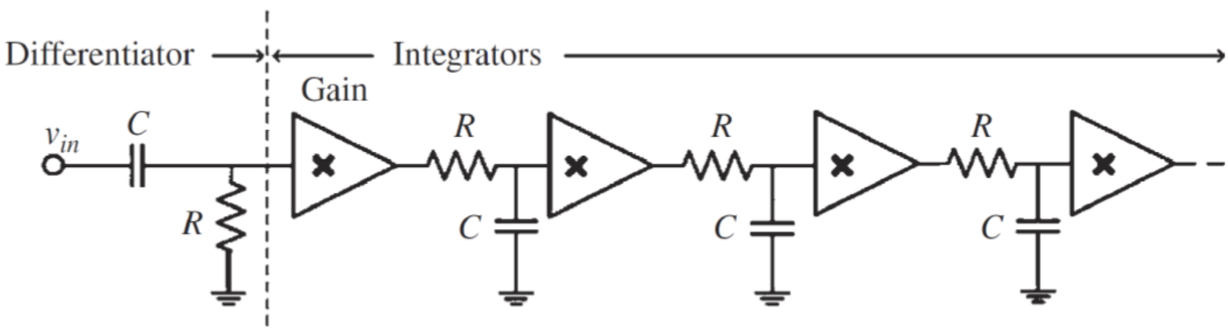
$$v_0(t) = \frac{t}{\tau_0} e^{-\frac{t}{\tau_0}}$$

Note that this pulse presents a peaking time equal to the time constant.

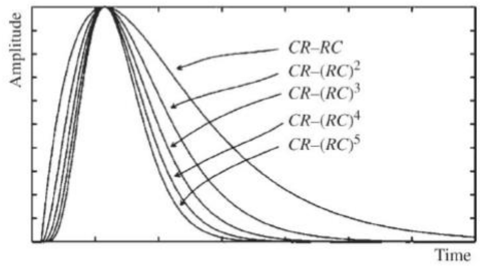


Cascading

If we add more integrators in cascade after the initial differentiator we continue to add poles at the same frequency and the shape becomes more and more like a Gaussian



For a generic number of n integrators the transfer function becomes



$$v_o(t) = \frac{A}{n!}\left(\frac{t}{\tau_s}\right)^n e^{\frac{-nt}{\tau_s}} \qquad \tau_0 = RC \qquad \tau_s = n\,\tau_0 \text{ peaking time}$$

This solution is more resistant to technology variations since the resistance is the values are all the same and do not need to pe carefully selected to match the desired parameters like in the semi Gaussian filter

Definitions of shaping time

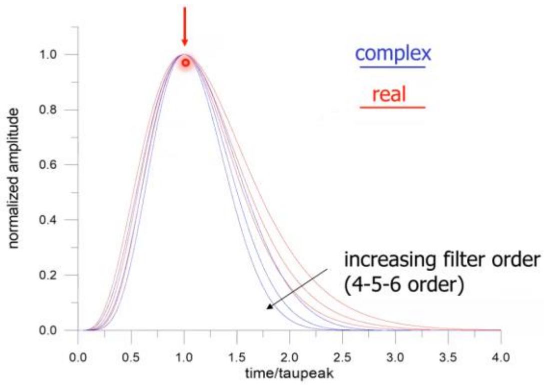
First definition

We define the peaking time as the time required by the filter to reach its peak output value

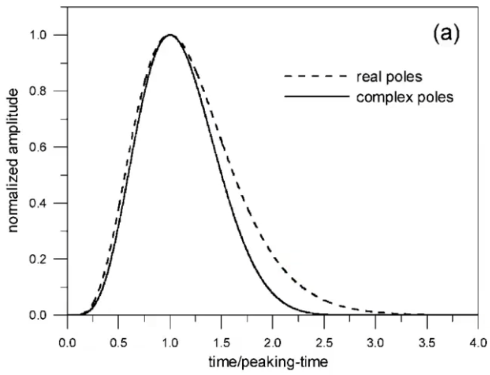
Effect of higher order filters

We can see that using this definition if we compare filters with the same peaking time we get that

A larger order filter will present a narrower pulse.



Real vs complex filters

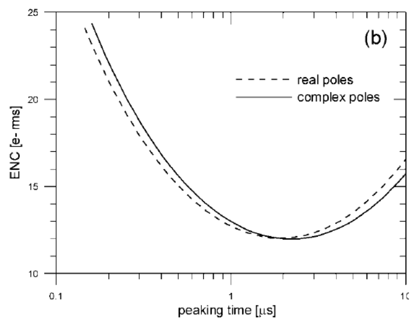


Duration

We can also see that the real filters (the ones implemented with the CR RC method) present a longer tail than the complex filters implemented according to the parameters for the semi Gaussian.

Noise

On the other hand, when considering noise the result is the same for real and complex filters



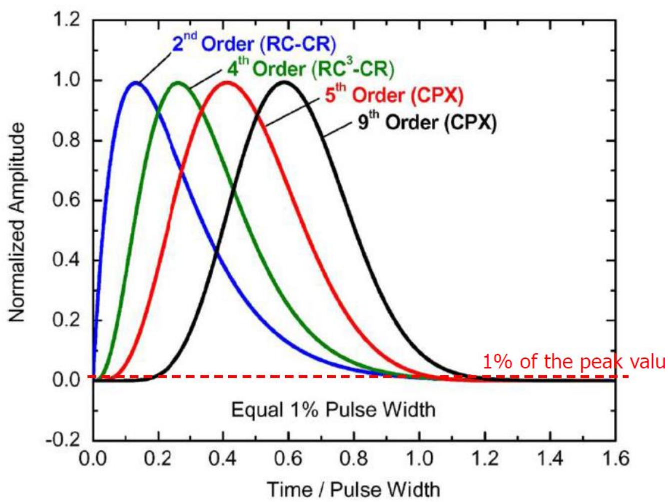
Second definition

We define the shaping time as the time duration of the pulse considering it over when it reaches 1% of its the peak value.

This definition is useful to avoid pile up of pulses.

Effect of higher order filters

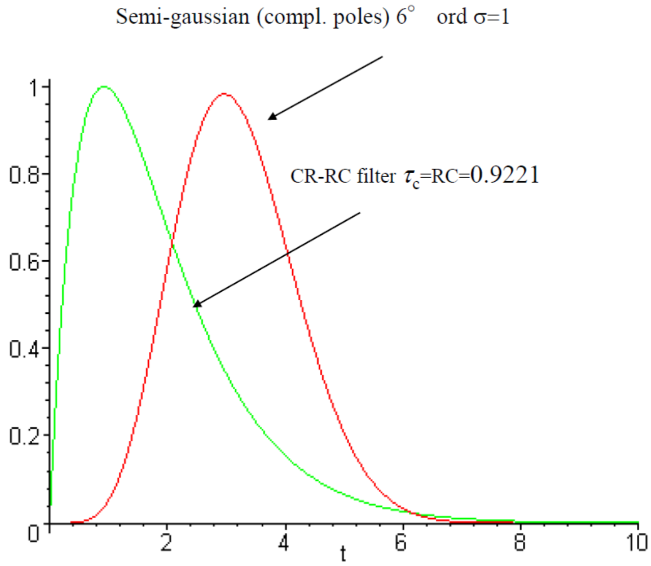
We can see that as the order increases the peak is progressively delayed, this **improves performances regarding the ballistic deficit**. Since the start of the response is delayed the input pulse as more time to provide all of its charge so it is not a problem if the input comes in the form of an exponential and not a delta.



Third definition

The shaping time is defined as the time constant of the CR-RC filter which provides the same peak value and area of the Gaussian approximated by the filter.

Explanation



Given a semi gaussian filter (in red) to find the shaping time I need to find the CR-RC filter which matches it in peak value and area.
The time constant of this filter (in green) is the shaping time of the original filter (red)

Computations

$$f(t) = \left(\frac{t}{CR}\right) e^{-t/CR}$$

$$S_{CR} = e \int_0^{\infty} \left(\frac{t}{CR}\right) e^{-t/CR}\, dt = e \cdot CR$$

area of the pulse CR-RC with amplitude normalized to 1

$$S_G = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2/\sigma^2}\, dt = \sqrt{(2\pi)}\,\sigma$$

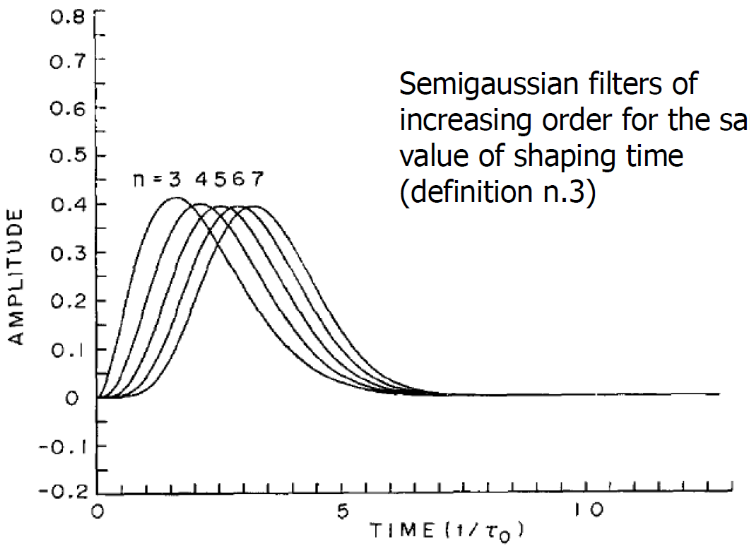
area of the Gaussian with amplitude normalized to 1

$$\boxed{\sigma = \frac{e}{\sqrt{(2\pi)}} \cdot CR = 1.0844\,\tau_0}$$ $$\boxed{\tau_0 = 0.9221\sigma}$$

Remember

$$\boldsymbol{\tau_0 = 0,9221\sigma}$$

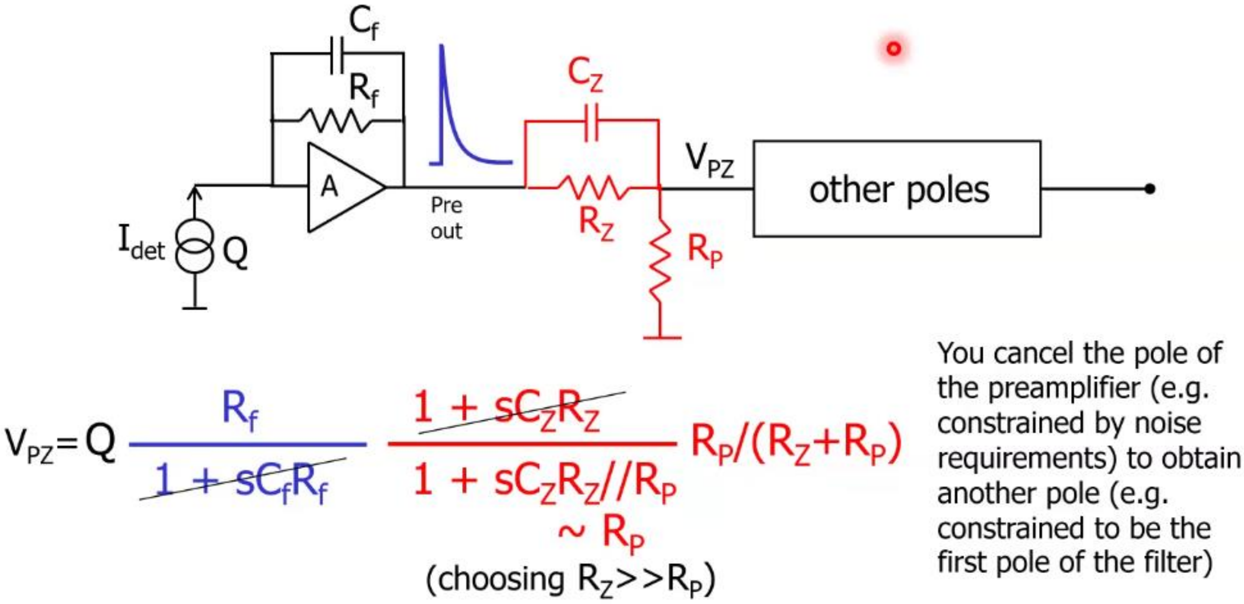
Effects of higher order



Semigaussian filters of increasing order for the same value of shaping time (definition n.3)

Pole zero compensation

We know that often the charge preamplifier will present a feedback resistance R_f in parallel with the feedback capacitance C_f, this will make it so that its pole is not at the origin so we need a more complex solution to compensate for it.



Typically we R_Z is a potentiometer so that it can be adjusted to be used with different detectors.

However we need also to consider that the have a pole, this pole can be used to implement the pole of the shaping amplifier, this however forces me to select an odd shaping of the semi Gaussian filter.

Ideally I would like for the value of the pole to be independent from R_Z since this resistance will be adjusted to match the zero with the pole of the preamplifier, to obtain this result we select $R_p \ll R_z$.

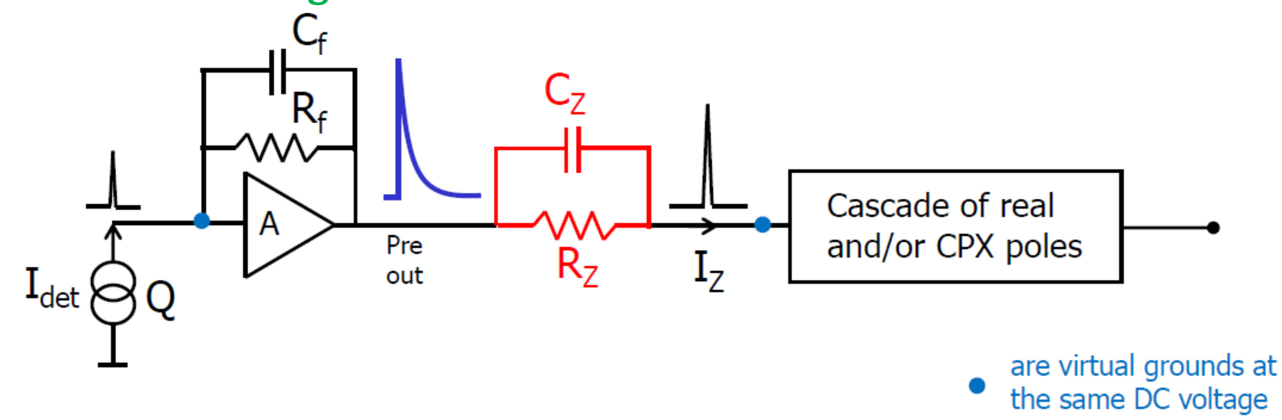
Why do we not just use the pole of the charge preamplifier

This is not a good solution because ethe time constant of the charge preamplifier is typically very long because this allows us to reduce the thermal noise of the resistor R_f, since the time constant of the pole of the filter is related to the width of the output pulse which we prefer to be as much shorter.

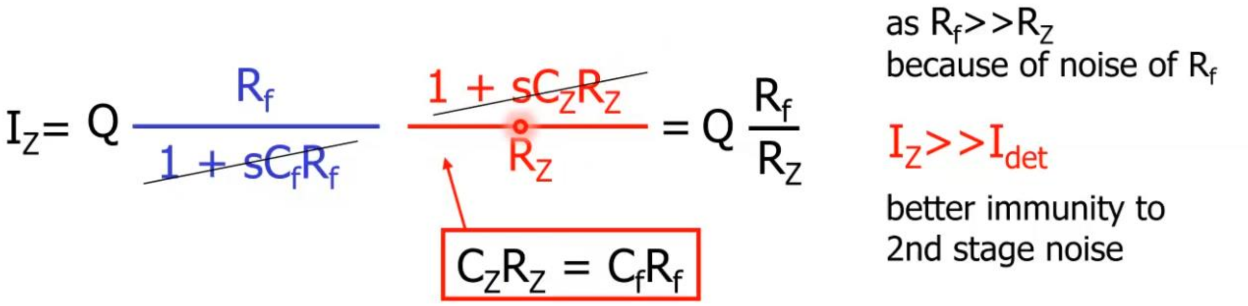
Important

The noise added by the resistance R_p has instead a much less significant impact on the overall signal since it is not amplified by the preamplifier value like the R_f thermal noise would.

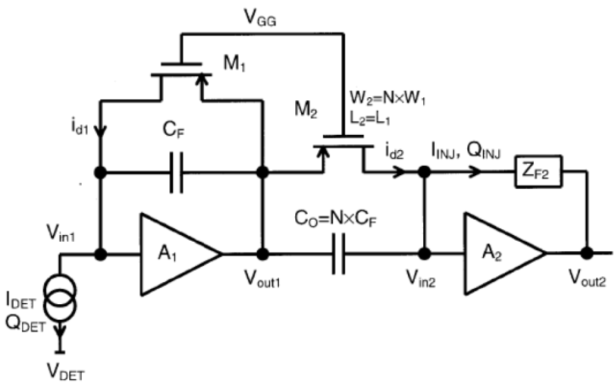
Alternative configuration



With this solution we do not add the pole so we are now able to implement any solution both with odd or even number of poles we just need to assure that the node is a virtual ground.
The input of the filter is now a current not a voltage



Self compensated pole zero network



Resistor design

In this implementation we utilize transistors biased with very low current to create high value resistors.

And to match the condition

$$C_Z R_Z = C_f R_f$$

Since we need to have

$$R_f \gg R_Z$$

What we d pos scale the size of R_f so that W of M_2 is N times larger than M_1.

Important

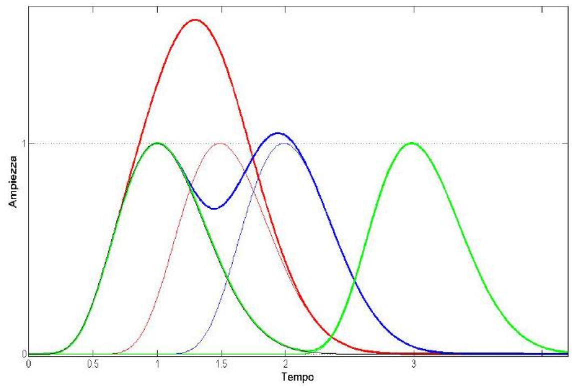
To use a transistor as a resistor for its value to be constant we need for the voltage across it to remain constant. This is not the case for M_2 since of course the output V_{out1} needs to change when an input arrives.
However this change in voltage is equal across both M1 and M2 so the non linearity is compensated.

Capacitor desing

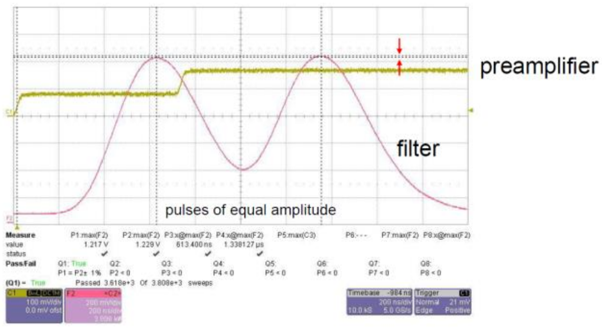
We will need to make C_Z proportionally larger than C_f to compensate for the resistances different values.
It is important to note that C_Z is not implemented by creating a single capacitor with a larger area as otherwise the fringe effects contributions to the total capacitance will affect the overall value.
Instead we connect in parallel N copies of capacitors with size C_f.

Pile up occurrence between neighbor signals

Nearby events produce superimposed signals that can corrupt the peak amplitude of the corresponding pulses, the corrupted signals have to be identified and rejected by a suitable circuit **pile up rejector** or **PUR**.



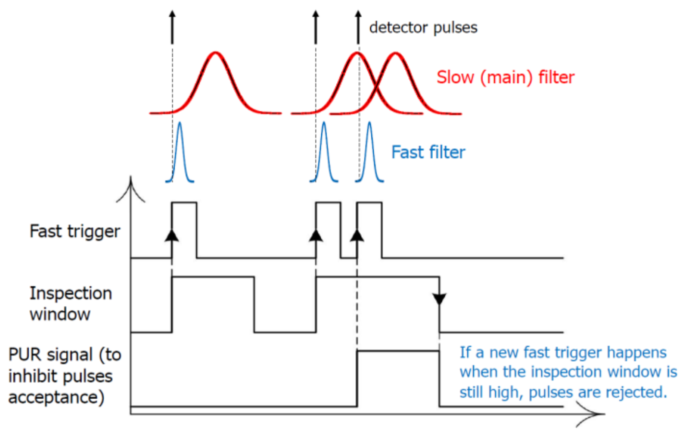
Example



In the graph we can see the condition for which 2 pulses overlap and start to corrupt each other, as we can see because of the overlap the second pulse is higher then the first when coming out of the preamplifier

We can set a limit for which this corruption is acceptable, like for example 1% of the peak amplitude of the signal.

The pile up rejector



We can see that while the first pulse is clean the second and the third are partially overlapped.

We want to be able to reject the overlapped pulses, this can be done through a second filter.

The preamplifier output is sent to both to the main filter but also to a second filter with much faster shaping time.

If the fast filter is narrow enough we are able to distinguish the pulses clearly, then I can use a trigger (a discriminator) connected to the fast filter to obtain a digital signal corresponding to each incoming pulse.

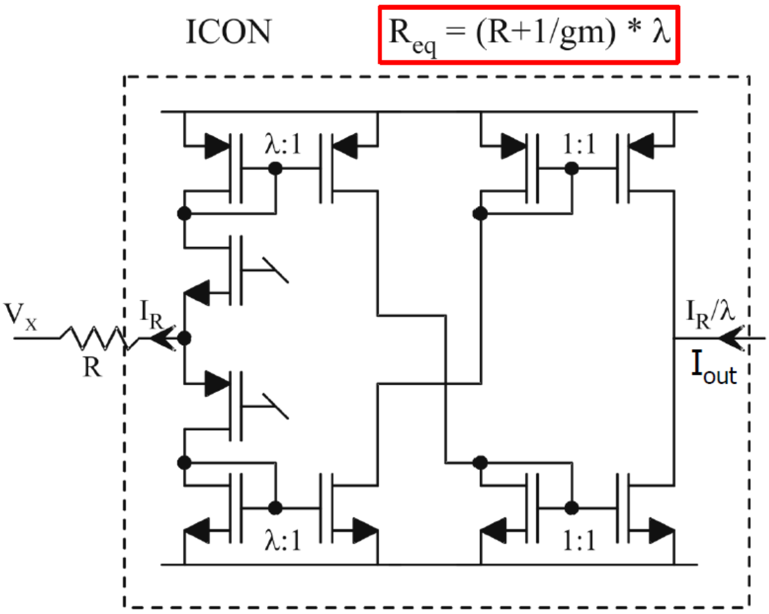
I use the rising edge of the trigger to start another digital pulse called **inspection window** whose duration will correspond to the minimum distance required for 2 pulses to not overlap.
Should the discriminator trigger again while the inspection window is high then the pulses are rejected.

Note

- We can see that after rejection the inspection window is reset because we need of the last pulse to be exhausted before being able to accept a new pulse.
- We do not use directly the fast filter because it is not the optimum filter for the pulses so it would cause a much worse SNR

ICON cell

We have seen more than once how it is difficult to implement a high level resistor with a reliable value.
In this solution we multiply a smaller physical resistor by a factor λ.,



Operation

At the input we have our resistor R, and we apply the input voltage V_x to it.

The other end of the resistance is connected to 2 cascode transistor each connected with a current mirror with a scaling factor of λ/1 so that the current at the output of the mirror will be λ times smaller than I_R. This means that looking from outside we would see a voltage to current ratio and thus a resistance

$$\boldsymbol{R_{tot} = \frac{V_X}{I_{Out}} = R \cdot}$$

Important

At the output we have another series of mirrors to correct the direction of the current otherwise we would have a negative resistance.

Effect of the cascode impedance

We need to remember that the cascode has a non zero input impedance so the real equivalent resistance is

$$R_{eq} = \left(R + \frac{1}{g_m}\right) \cdot \lambda$$

This create a problem because the transconductance changes with he signal, so to compensate we need to bias the branch with a very large current to avoid the signal affecting the bias point of the transistors.
This is not a good solution.

Example of an implementation

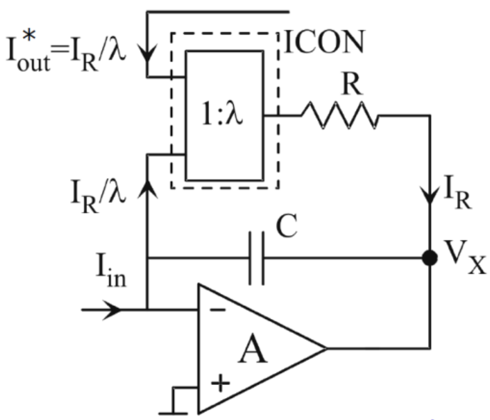
in this implementation we have simply added the ICON cell in series with R.

Use of the output current

The circuit receives an input current and gives

- An output voltage V_X
- An output current

The output current which will be affected by the pole of the integrator can be cascaded into the input of another identical cell, this allows us to create a CR-RC filter.



$$\boxed{\tau = C \cdot (R+1/gm) \cdot \lambda}$$

$$\frac{V_X}{I_{IN}} = \frac{(R+1/gm)\cdot\lambda}{1+s\cdot C\cdot(R+1/gm)\cdot\lambda}$$

$$\approx \frac{R\cdot\lambda}{1+s\cdot C\cdot R\cdot\lambda}$$

* The secondary output I_{out} is obtained by duplicating the output MOSFETs of the ICON

limit: dependence from gm

Implementation of the second output

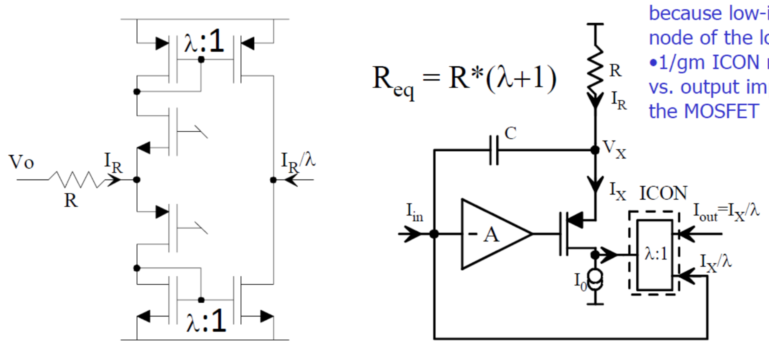
To create this secondary current output we just need to connect other 2 transistors to the output mirror so that the mirror has 2 outputs, we could also think to change for this second output and thus add a gain to the transfer

Alternative implementation

The resistor is not connected in the feedback but between the power supply and the output of the OPAMP, because of the feedback the impedance seen by R at V_x is basically zero so we do not need to worry about the transimpedance of the pMOS.

The current will flow from the resistance R through the pMOS and then through the ICON cell.

An improved ICON solution



- $I_R = V_X/R$ (no effect of 1/gm MOS because low-impedance node of the loop)
- 1/gm ICON negligible vs. output impedance of the MOSFET

Original ICON cell (Chase, Hrisoho, 1998):
R_{eq} determined also from 1/gm of the common-gates and not only by R
(in this example an 'inverting' cell is shown, without the second mirror)

Modified ICON cell:
- R_{eq} determined by R only

$$\boxed{I_{out} = \frac{I_{in}}{1 + s\lambda RC}}$$

We create the current away from the icon cell and then feed it into it through the high impedance of the pMOS drain.
The final transfer function is the following

precise calculation

$$I_{out} = I_{in}\frac{1 + sCR(1+\alpha)}{1 + \beta + sCR[(\lambda+1)(1+\alpha)+\gamma]}$$

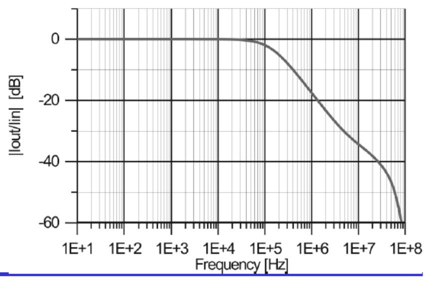
$$\alpha = 1/A \qquad \beta = \frac{\lambda R}{AR_o} + \frac{\lambda}{Ag_{mPMOS}R_o}$$

$$\gamma = \frac{\lambda}{Ag_{mPMOS}R_o} + \frac{\lambda}{Ag_{mPMOS}R}$$

$$A \to \infty$$

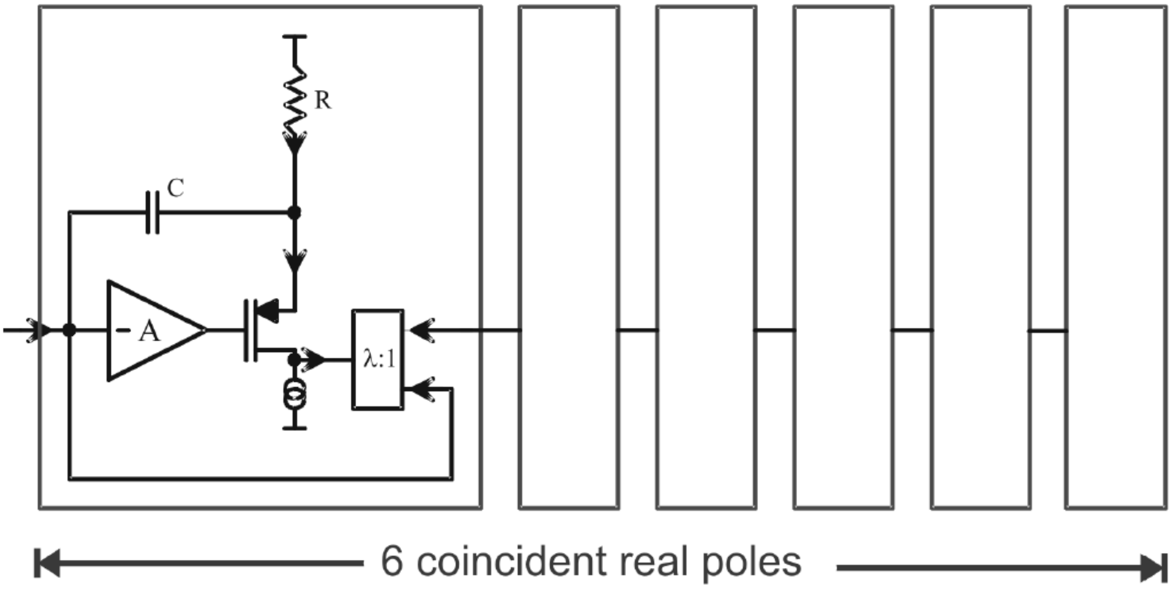
$$\boxed{I_{out} = I_{in}\frac{1 + sCR}{1 + sCR(\lambda+1)}}$$

the zero is λ times faster than the pole
(be aware for small λ!)



We can see that the circuit present snot only a pole but also a zero at a higher frequency, so it is usually neglected as long as λ is large which is our objective so it is a reasonable assumption.

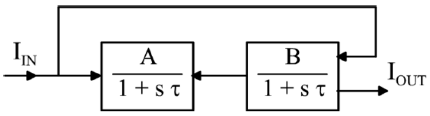
Semi gaussian filter with real poles with ICON cells



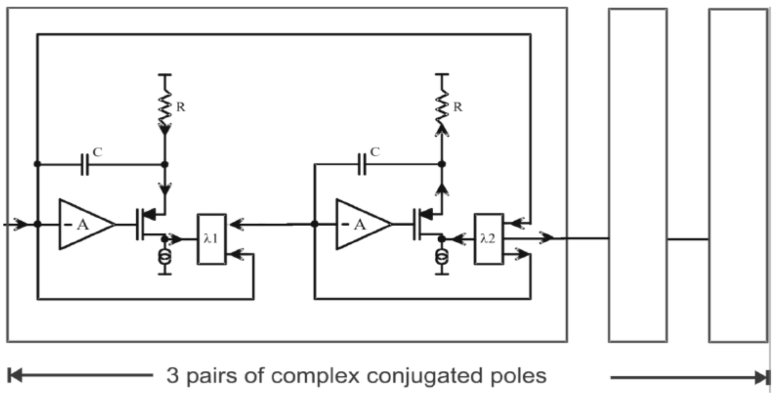
Note: very easy 'copy-and-paste' of the cell layout, in the design of an integrated filter.

This solution has a more stable transition because with this implementation we no longer have $\frac{1}{g_m}$ in the pole time constant.

Implementing complex conjugate poles



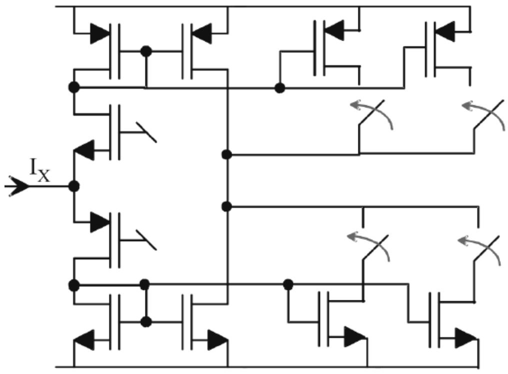
$$\boxed{I_{out} = I_{in} \frac{AB}{(1+s\tau)^2 + AB}}$$



Where A and B are the gain factors at the secondary output of each ICON cell.

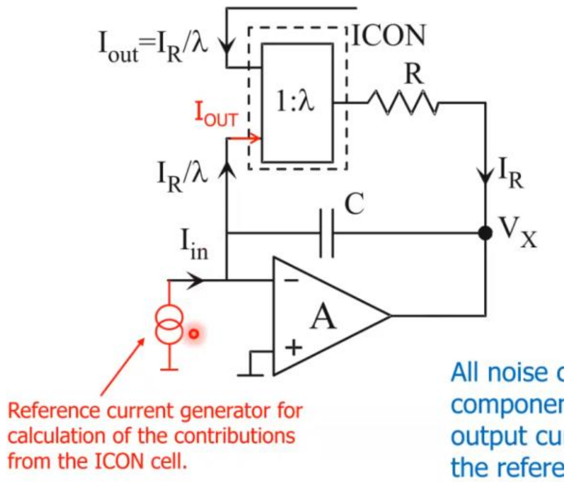
Selecting the shaping time with ICON cells

We may obtain different shaping times if we utilize ICON cells with selectable secondary output transistors for the final mirror which can be connected in parallel thus reducing the equivalent resistance and thus decreasing the shaping time



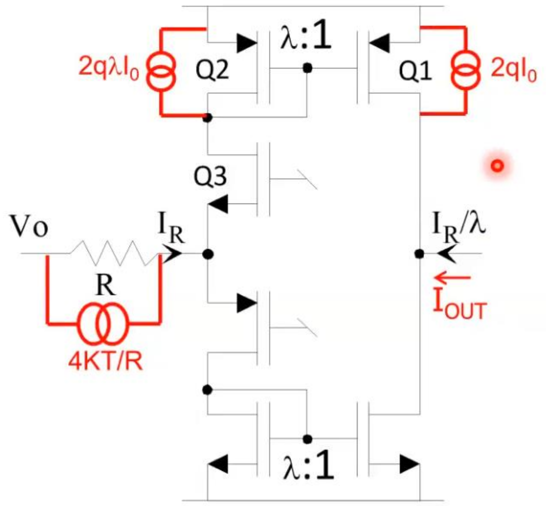
Noise in the ICON cell

we want to refer the total noise to the input current as this makes it easier to compare it with the output current coming in from the previous cell.
To obtain this value we simply need to compute the value of the noise at I_{out}



We have that

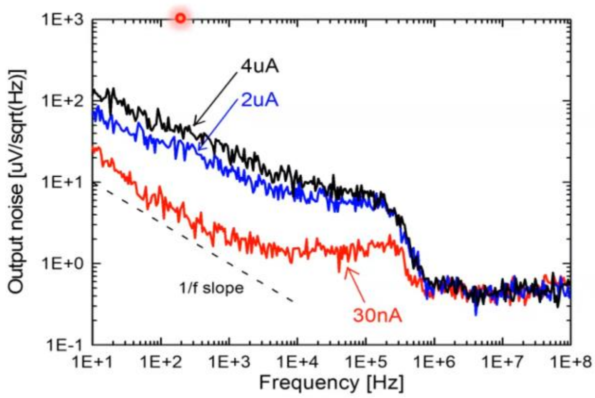
- $Q1$ has its noise directly connected to the output
$$2qI_0$$
the noise is represented using the shot noise because when transistors are operating in weak inversion (low current) than the most appropriate way to indicate the noise is the shot noise.
- $Q2$ its current is λ times larger than Q1 but its noise is divided by λ^2 when referred to the output
$$\frac{2qI_0}{\lambda}$$
- $Q3$ the noise is cancelled by being a cascode configuration
- R needs to be divided by λ^2 so we get
$$\frac{4kT}{R} \cdot \frac{1}{\lambda^2}$$
this means that the total noise will actually be lower by a factor λ than the noise we would get using an equivalent resistor $R_{eq} = \lambda R$



We can repeat the considerations for the bottom transistors.

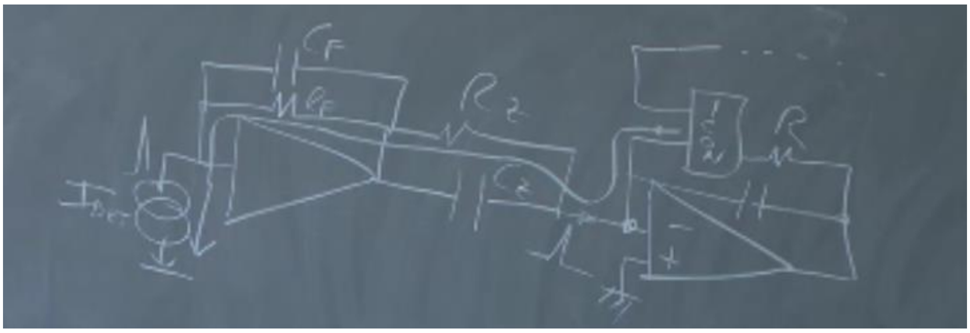
Bias and noise

To maintain a low noise we want to bias the cell with a low current I_0, if we do this the $\frac{1}{g_m}$ will become much larger, this is the reason why we prefer the second implementation which does not include the $\frac{1}{g_m}$ in the total resistance.
This is also why the MSOFETs operate in very weak inversion because as we can see from the graph a lower current causes a much lower noise.



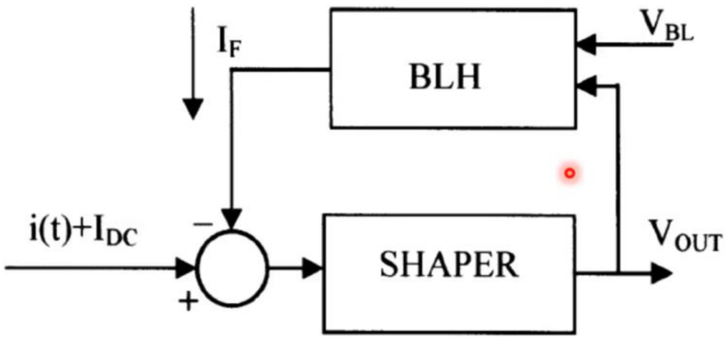
The baseline holder

This blocks stabilizes the DC level of the baseline at the output of the shaper amplifier to V_{BL}.
If we consider the generic acquisition chain we can see that we have a DC path through the resistors and the icon cell this means that the detector receives not only the signal but also the dark current.



This will cause a DC voltage to be present at the output of the integrator

The output of the shaper will be the superposition of a DC value and the signal.



While the DC offset can create a problem causing the signal to exit the range of the ADC, the more serious problem is that this DC voltage depends on the dark current and thus on the temperature, so this value may shift in time.
For this reason we what to introduce the BLH so that we can create a DC feedback which maintains the DC value of the shaper at a constant value.

Important

The circuit must act only at DC and not influence the pulse.

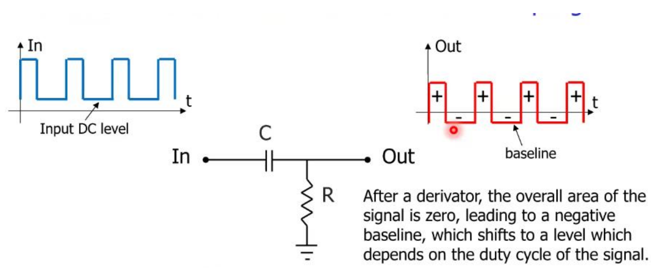
Alternative: AC coupling

this is an alternative solution to using a BLH, we utilize a derivator to remove the DC component.

Problem

Since the average output of a derivator should be zero (we have a zero in DC in the transfer function) if we feed in the input a train of square waves then the output will be based on a level slightly below zero otherwise the net area would be larger than zero.

This is negligible when the pulses are very far apart, so the baseline remains very close to zero, this is not the case for medical applications like pet where we have and extremely high pulse rate with a statistic rate, we would end up having Both a significant baseline shift but also a variable baseline since the rate is not fixed



$$V_{out}(s)= V_{in}(s)\,\frac{sCR}{1 + sCR}$$

On the right we can see the computation for $t \to \infty$ and thus $s \to 0$.

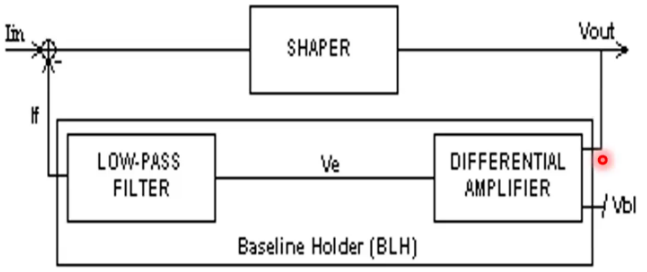
$$\int V_{out}(t)dt = \lim_{s\to 0} \cancel{s}\,\frac{V_{out}(s)}{\cancel{s}} \to 0$$

Because of the zero in the origin from the derivator.
⇒ The signal area is zero

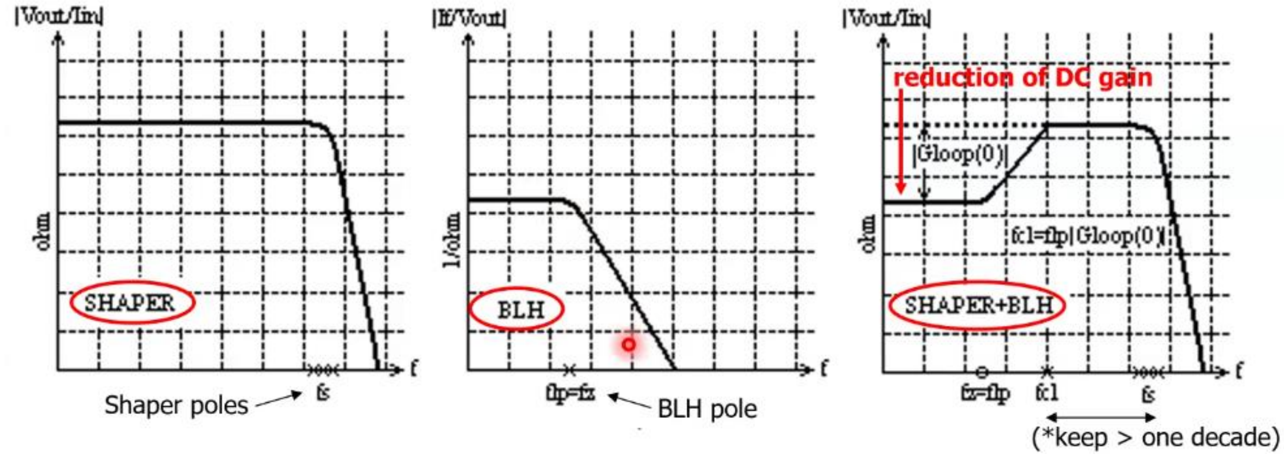
Basic BLH structure

the BLH is placed in parallel with the shaping amplifier, it is composed by 2 basic blocks

- A low pass filter which opens the feedback at high frequency
- A differential amplifier



Bode plots



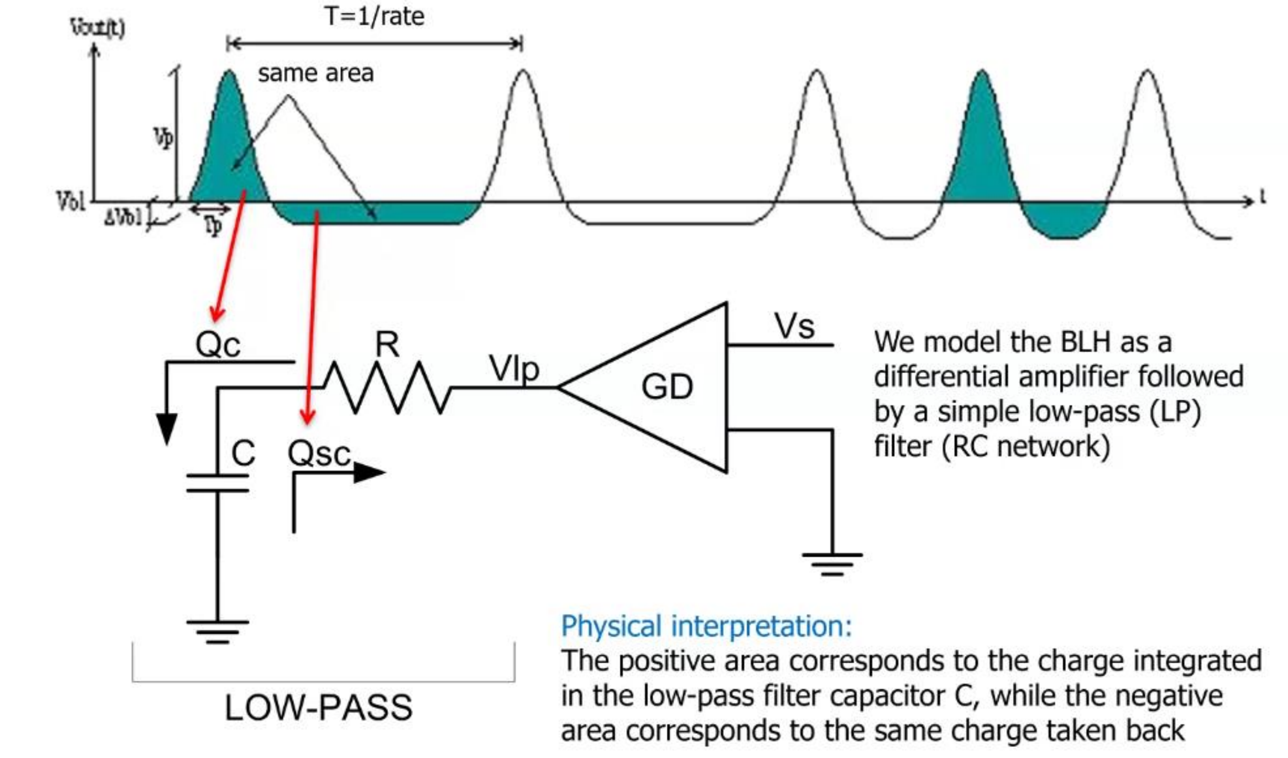
We can see that the DC gain is reduced by the G_{loop}, until we reach eh zero of the BLH at which point the transfer function starts to go back to the original value once the loop gain reaches 1.

G_{loop} and stability

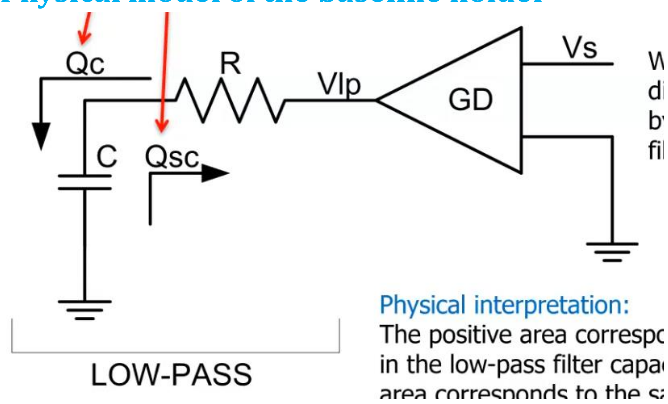
We want for the G_{loop} to be as large as possible so that the DC gain can be reduces as much as we can. We need to consider that the zero introduced by the BLH does not move and can not be moved since we already tried to place it as low as possible, so as the G_{loop} increases we end up in a situation where the pole $f_{C1} = f_z \cdot |G_{loop}(0)|$ eventually reaches the poles of the shaper causing instability so we have a limit to on the maximum G_{loop}.

Baseline shift

The baseline holder is affected by the baseline shift just as the AC coupling solution.



Physical model of the baseline holder

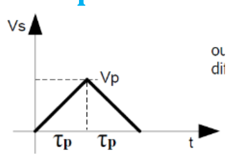


we represent the baseline holder with a differential amplifier with gain G_D,whose output is connected to a RC integrator.

When the input is a positive pulse the output of the amplifier is simply an amplified replica of the pulse, this causes a positive voltage and thus a current flowing in the capacitance C chagrining it.

The negative baseline shift will be caused by the capacitor discharging

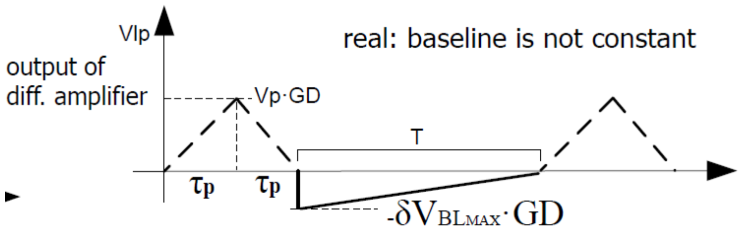
Example



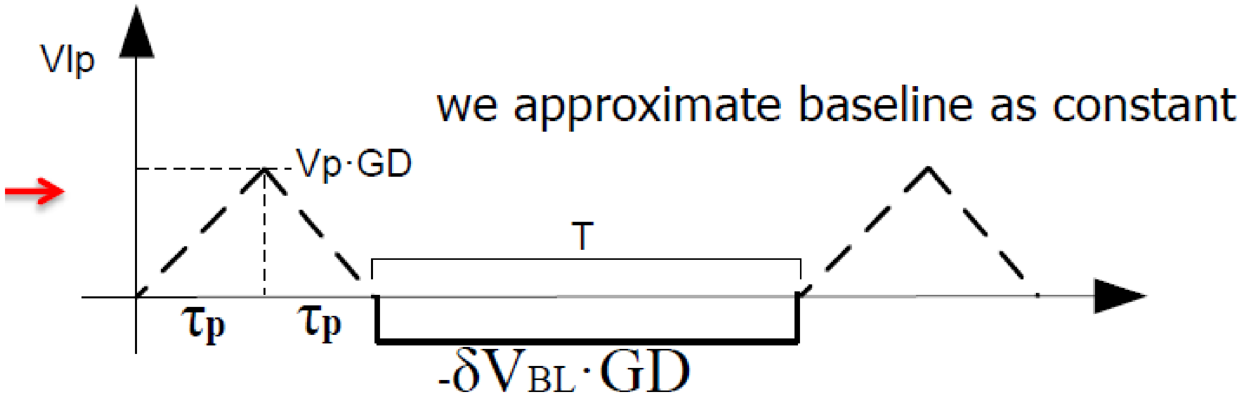
Let’s approximate the pulse with a triangle.

After the amplifier we have the same waveform with a peak amplitude amplified by G_D (gain of the amplifier).

Following it we have the baseline shift as the capacitor exponentially discharges



We approximate and consider a constant baseline variation



Charge balance

We can compute the charge integrated in the capacitor by integrating the current of the pulse

$$Q_C = \int_0^{2\tau_p} \frac{V_{LP}}{R} dt = \frac{1}{R} V_R G_D \tau_p$$

The charge discharged can be calculated in the same way over the duration of the discharge

$$Q_{SC} = \int_0^{T} \frac{|V_{LP}|}{R} dt = \frac{1}{R} \delta V_{BL} G_D T$$

These 2 contribution should match, from this we can obtain the value of the baseline shift

$$\boldsymbol{\delta V_{BL} = \frac{V_p \tau}{T} = V_P \tau_p \cdot pulse\ rate}$$

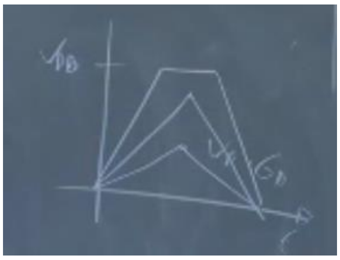
Observations

- The result is independent from the resistance R
- The result is dependent on the pulse rate so if the rate changes the baseline changes

Reducing the baseline variation

First solution

We increase the gain of the differential amplifier so that when the signal arrives its output is instantly saturated, this way the output pulse becomes a rectangle



The charge integrated is now equal to

$$Q_C = \int_0^{2\tau_p} \frac{V_{DD}}{R} dt = \frac{1}{R} V_{DD} 2\tau_p$$

As we can see the charge integrated does not depend on the pulse amplitude and also does not depend on the gain of the differential amplifier.

The computations for the charge that is discharged are the same as before (we assume that the amplifier is not saturating in this case

$$Q_{SC} = \int_0^{T} \frac{|V_{LP}|}{R} dt = \frac{1}{R} \delta V_{BL} G_D T$$

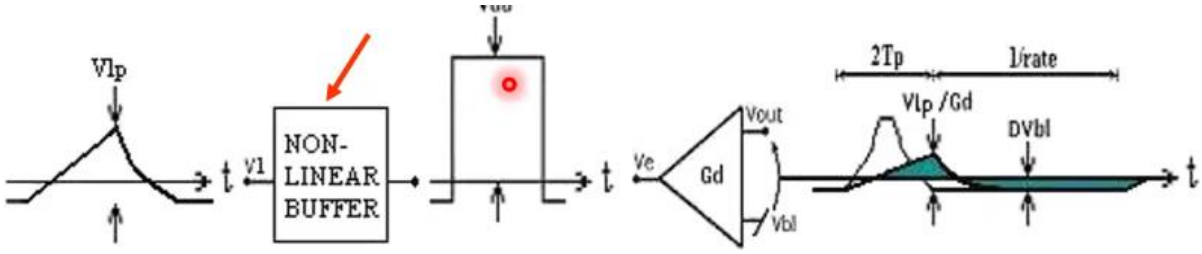
Now we can compute the baseline value and obtain

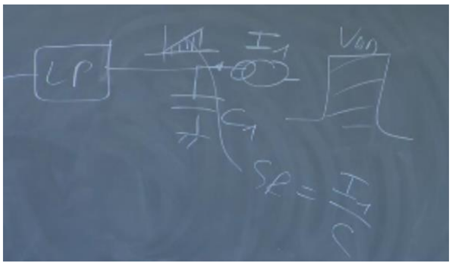
$$\boldsymbol{\delta V_{BL} = \frac{V_{DD} 2\tau_p}{G_D} \cdot rate}$$

We can see that it is now possible to reduce the amplitude of the baseline by simply increasing the gain of the differential amplifier.

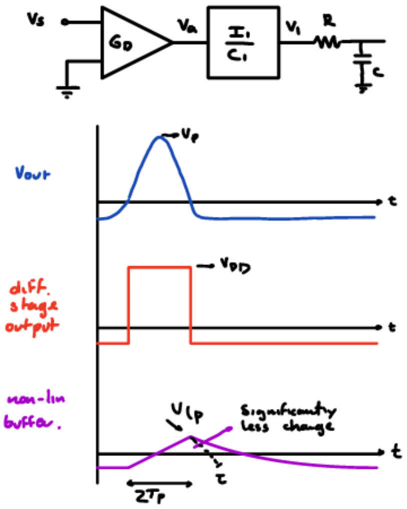
Second solution

We add a buffer with a ramp (a capacitance that needs to be charged) between the amplifier and the low pass filter.





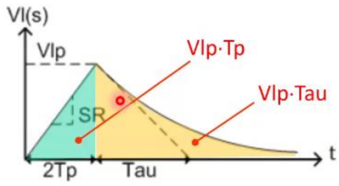
This adds a slew rate which we utilize to slow down the amplifier making it so that rather than integrating a rectangle we integrate a triangle with same width, thus we reduce the amount of charge integrated.



Charge and discharge

When we charge we are connecting the capacitance to a current generator so we have a linear ramp while when we discharge we do so through a resistance so we have and exponential decrease.

Charge balance



The charge in the capacitor first increases linearly with a ramp and then decreases exponentially with a time constant RC.

The charge integrated is

$$Q_C = V_{lp} T_P$$

The total charg eof the waveform is

$$Q = V_{lp} \cdot (T_p + \tau)$$

The base line is given by

After the usual calculation Q_C=Q_{SC}:

$$\Delta(Vbl) \cong \frac{Vlp}{Gd}(Tp+\tau)\cdot rate \cong \frac{(2Vdd)(Tp)(rate)}{Gd}\cdot\frac{(Vlp)(Tp+\tau)}{(2Vdd)(Tp)}$$

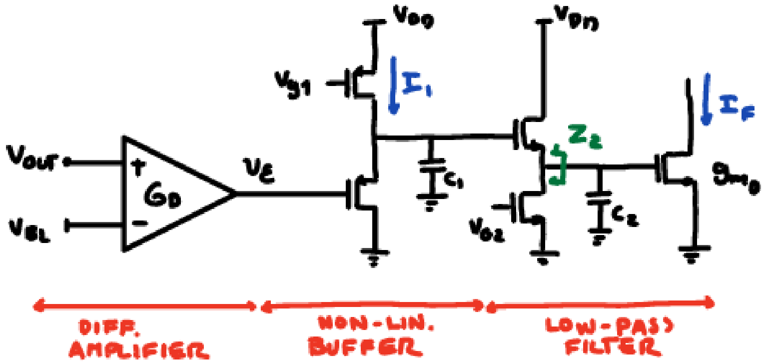
In the formula we have the result of the rectangle scaled by a coefficient K equal to the ratio between the area of the new charge curve and the rectangle area, and that indicates the non-linearity of the buffer. The coefficient depends on V_{LP} which can be controlled by slowing down the slew rate

$$\Delta(Vbl) \cong \underbrace{\boxed{\frac{(2Vdd)(Tp)}{Gd}}}_{\text{saturation}}(rate)\cdot\underbrace{\boxed{K}}_{\text{non-linearity}} \qquad K = \frac{(Vlp)(Tp+\tau)}{(2Vdd)(Tp)}$$

Baseline holder circuit implementation

Differential amplifier

This is a classical OTA, since the circuit needs to operate only at low frequency we do not need to worry about bandwidth but only about offsets as they would be transferred giving us a non zero DC output level which is the objective of the circuit



Non linear buffer

As we said this has the objective of introducing a slew rate, we implement this as a source follower driven by the output of the amplifier.

When the output of the amplifier saturates after receiving an input pulse the connected pMOS turns off and now the capacitor which biased the transistor sends its current in the capacitor C_1 charging it with a constant rate.

We have instead that when there is no pulse the circuit is operating in a linear regime as a normal source follower so it will display a $\frac{1}{g_m}$ impedance through which the capacitor will discharge.

Low pass filter

After the non linear buffer we have a transistor in follower configuration biased with another transistor, a capacitor is connected to this follower thus creating a low pass filter as the capacitor will see the transimpedance $\frac{1}{g_m}$ of the transistor.

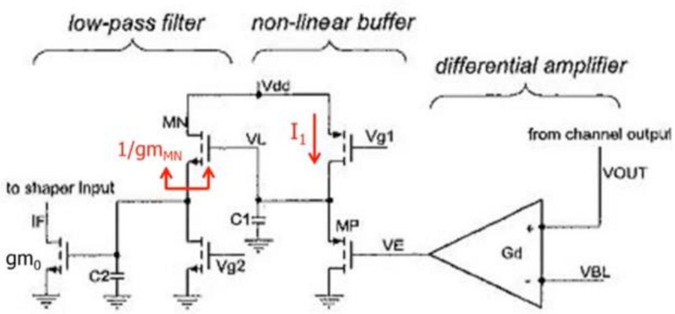
Since we want this filter to have a high time constant we bias the follower with low current.

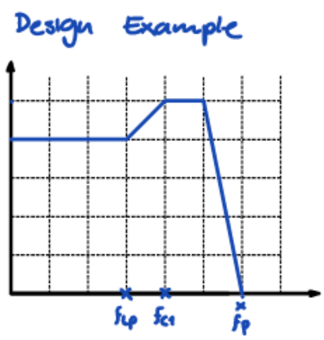
Important

But the baseline holder must supply a current, not a voltage, so we use the output of the LPF to drive a transistor.

Design example

Transfer function





Gain

We start by computing the gain, we multiply the gain of each stage

1. G_d coming dorm the differential amplifier
2. 1 from the first source follower
3. 1 from the second source follower
4. g_{m0} form the final transistor which converts the voltage in current

So the overall gain is

$$G(0) = G_d g_{m0}$$

Poles

We have that each follower operates as a low pass filter with a tau equal $C \cdot \frac{1}{g_m}$, to find these poles we need to consider the formula for the transconductance in a transistor in weak inversion

$$g_m = \frac{I_d}{V_{th} n}$$

- The current divided by the thermal voltage times a coefficient n

We obtain the following transfer function

$$F(s) = \frac{G_d g_{m0}}{\left(1 + s\frac{V_{th} n}{I_d} C_1\right)\left(1 + s\frac{V_{th} n}{I_d} C_2\right)}$$

The first pole introduced by the source follower connected to the amplifier can often be neglected as it will be at a much higher frequency than that of the low pass filter.

Loop gain

We said that while we wish to have a large loop gain we need to pay attention to avoiding instability, as if the gain is too high than the pole of the low pass filter may reach the poles of the shapers and introduce instability.
So let's operate backwards and find the maximum loop gain which avoids instability.
Lets say that we want to have the closed loop pole at least 2 decades before the shaper pole

$$f_{cl} < \frac{f_s}{100}$$

We know that the closed loop pole will be at a frequency equal to

$$f_{cl} = |G_{LOOP}(0)| \cdot f_{lp}$$

Where $\tau_{lp} = \frac{V_{th} n}{I_d} C_2$, so we can set

- $f_s = 100kHz$
- $C_2 = 10pF$
- $G_{Loop} = 100$

And obtain the maximum acceptable bias current for the transistor

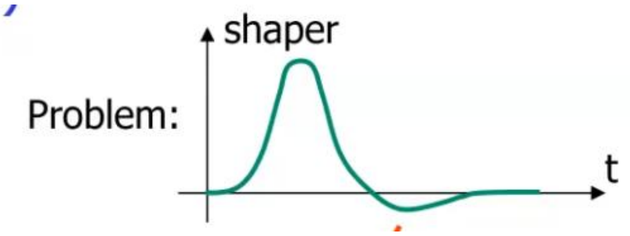
$$I_d \leq 10pA$$

Creating a small current

To create such a small current we need to first create a bigger current (like $10nA$) and then use a mirror with a scaling factor to create ethe smaller current as it is not reliable to generate directly this small current.
Note we might still have variations in the order of 50% but for this kind of operation is fine as long as we do not go wrong by an order of magnitude.

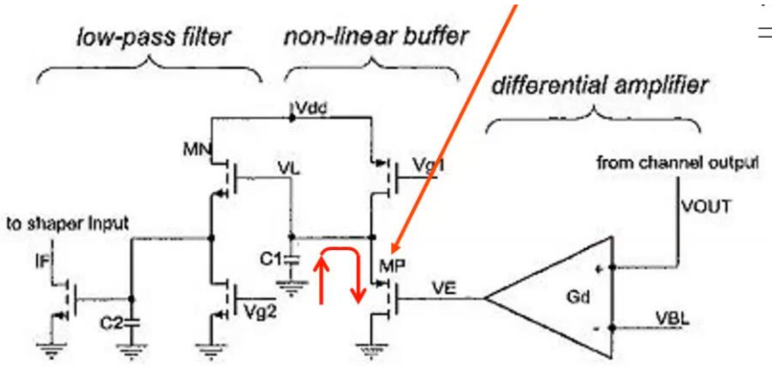
Negative pulses

There are pulses which show a residual negative undershoot as a consequence of a non-perfect positioning of the poles in the constellation.



Current limitation

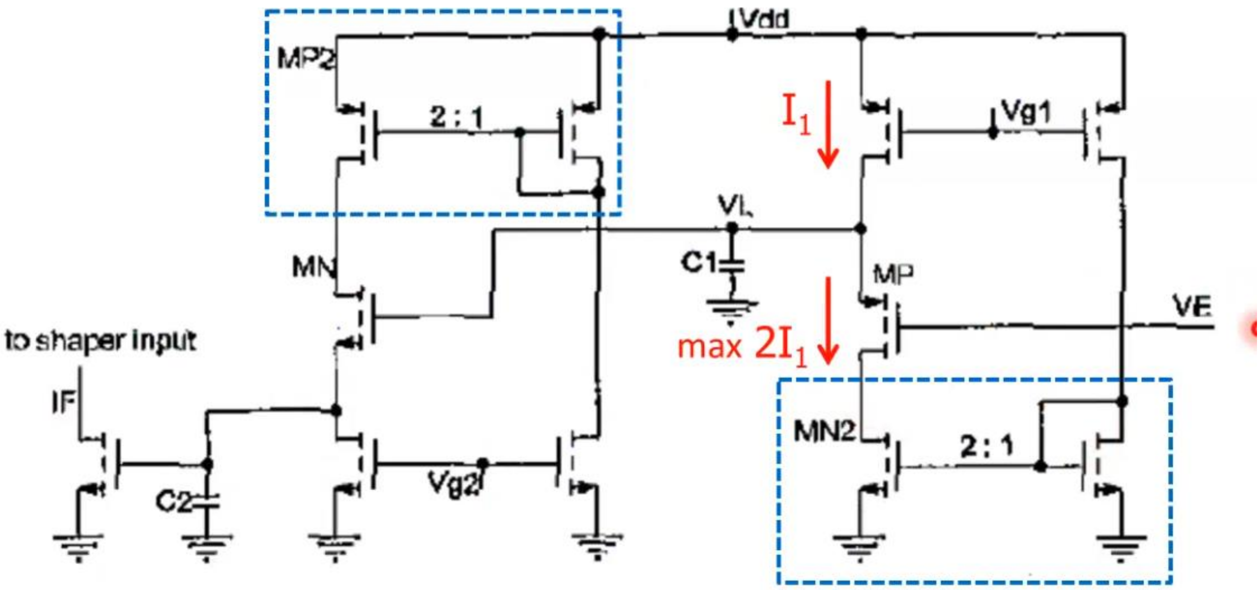
This creates a problem because the output of the amplifier will saturate to the negative value thus the pMOS of the first source follower starts to call a lot more current than the bias which will need to be provided by the capacitor.



We are going to need to refill the charge extracted this will significantly affect operation when the next pulse arrives.

Current limtier

To avoid this we need to introduce a transistor that limits the current in the branches.



Network limiting MP drain current

In this structure the current is limited by a current generator, now the mirror at the drain of the pMOS makes it so that we limit the current to a fixed value.

The reason why this works in DC bias is because the transistor MN2 exits saturation because the voltage at its drain drops when the current of MP is not enough so the mirror is broken.

Base line general formula

Vp=1V Gd=10 Vdd=1.7V

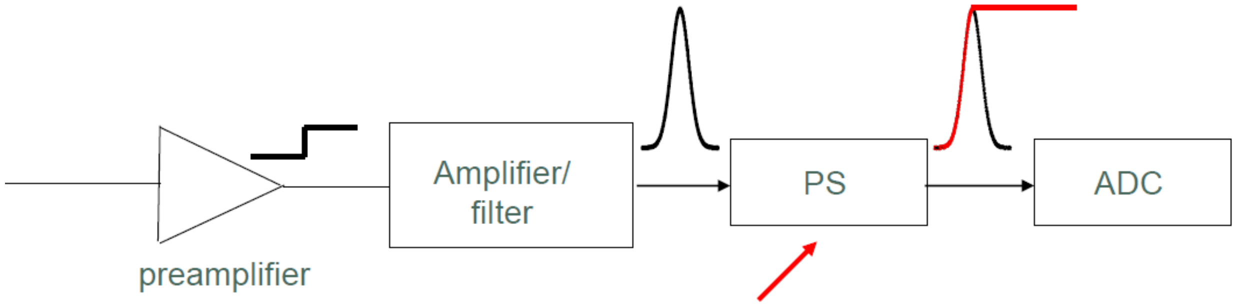
τ_p=4us I1=10nA

rate = 100k counts/s C1=200fF

$$\boxed{\delta V_{BL} = V_P \tau_p rate \cdot K}$$

	K	δV_{BL}
Basic	1	400mV
saturation of Gd	$\frac{2V_{DD}}{V_P G_D}$ = 0.34	136mV
satur.+SR	$\frac{I_1}{C_1}\frac{2\tau_s}{V_P G_D}$ = 0.04	16mV
single stage: sat.+SR+LP	$\frac{2V_{th}}{V_P G_D}$ = 0.005	2mV

Peak stretcher



The peak stretcher operates in 2 phases

1) First it tracks the pulse and operates as a buffer

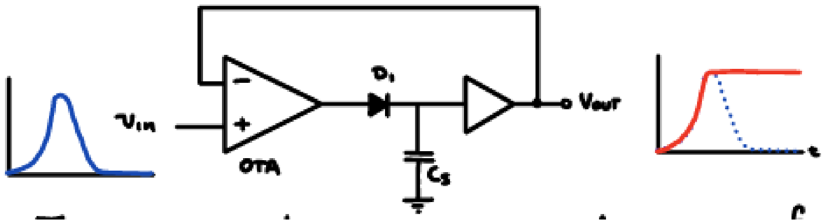
When the pulse reaches the peak and starts to fall down we move towards the second phase

2) The circuit starts operating like a hold circuit maintaining an output voltage equal to the peak voltage of the pulse

This allows the ADC to convert the peak even if the pulse is very short

Structure

We have an OTA followed by a diode a capacitor and then a buffer.
The output of the buffer is sent back to the negative input creating a feedback.



Operation

The feedback works only when the diode is on, if this is true (which requires the OTA to have a positive output) then the negative feedback will adapt so that the voltage at the negative input and thus at the output of the circuit matches that of the positive input.
In particular the capacitance C_S will be charged to the voltage value which gives this output.

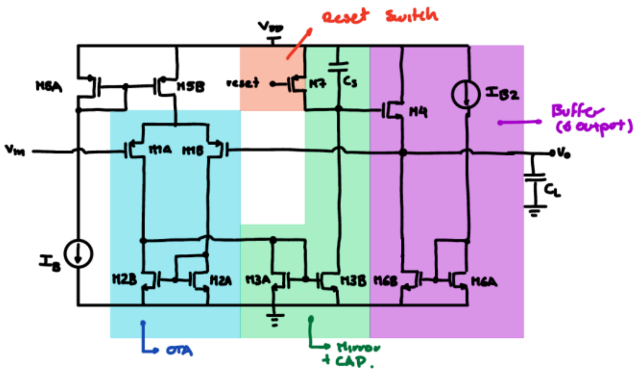
Once the voltage at the input starts to decrease the OTA output becomes negative at this point the diode turns off and the feedback is broken.

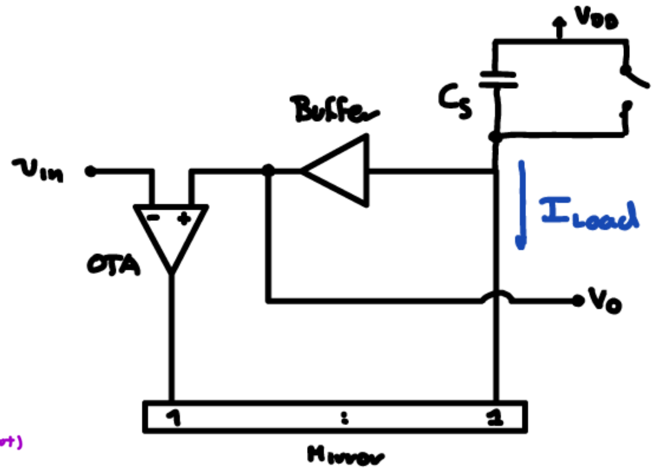
The capacitor is left floating and thus maintains its charge and its voltage so the output of the buffer does not change and remains equal to the maximum reached by the input.

Alternative configuration

Rather than utilizing the diode we utilize a current mirror to connect the OTA with the capacitor.

We have that if the current of the OTA changes direction than the mirror turns off so we have the same behavior but now we do not need to relay on a diode.





Note

We need to remember that the mirror inverts the current correction so we need to close the loop on the positive input rather than the negative one.

Figures of merit

Loop gain

We need for the loop gain to be high in the frequency range of the shaping amplifier as the error of the buffer will be inversely proportional to the loop gain.
The gain is equal to the product of the gain of each single stage

- G_m because of the OTA
- 1 of the mirror
- $\frac{1}{sC_S}$ because of the capacitor

$$G_{loop} = G_M \cdot \frac{1}{sC_S}$$

Droop rate

The capacitor discharges because of parasitic currents, the droop rate indicates the voltage variation overt time of the output and is equal to the ratio between the parasitic currents discharging the capacitor and the capacitor itself

$$\frac{dV}{dt} = \frac{I_{off,M7} - I_{off,M3B}}{C_S}$$

Note

The current subtract themselves since they opposite directions so we can try to match the transistor so that they cancel out as much as possible

Capacitance value

The value of the capacitance is determined by a tradeoff between lop gain and drop rate.

Operation details

Reset values

We can see that since the capacitor is connected to V_{DD} when we reset it the output goes to V_{DD} and not to ground.

The circuit will then quickly settle back to ground, we do not directly set to ground because this helps us reduce the influence of offsets.

Why the output buffer

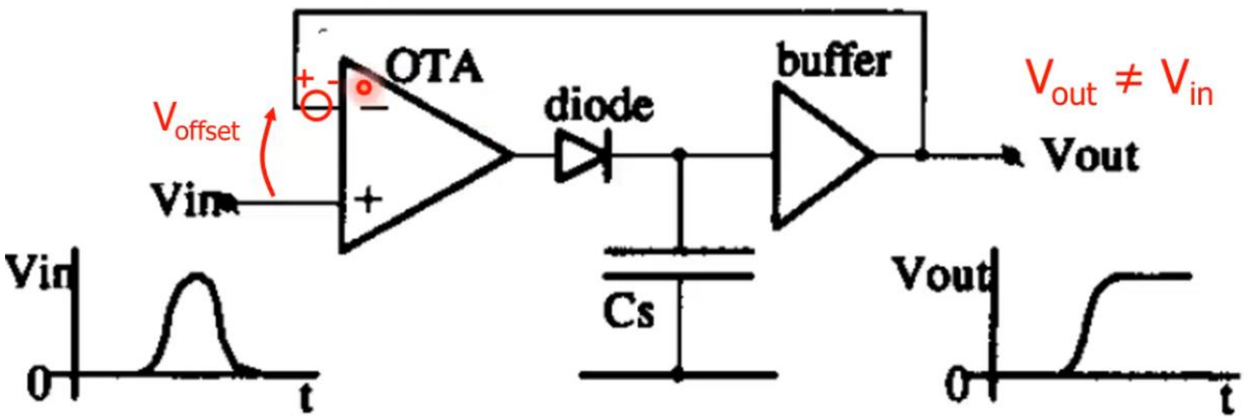
The buffer is not necessary in theory in reality we need it to prevent the capacitor to interact from interacting with the parasitic capacitances of other components.

We keep the buffer inside the loop so that any of its non linearities is compensated by the loop.

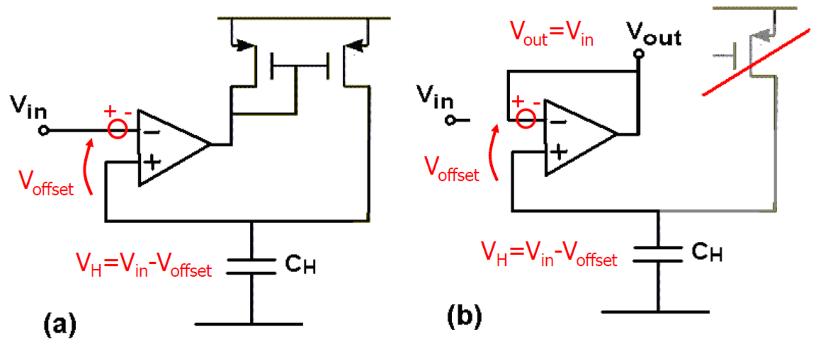
Note we can either use a real buffer or just a transistor is source follower configuration

Effect of the offset of the OTA

The offset is transferred directly to the output, we can either calibrate the device to compensate for the offset or utilize a 2 phase solution peak stretcher.



Two phases peak stretcher (sample and hold)



the circuit uses 2 distinct phases with different topologies to charge and discharge the capacitor.

Sample

During the sample phase we charge the capacitor to a voltage

$$V_H = V_{in} - V_{offset}$$

To do so we use a current mirror to create a negative feedback matching the – and + input of the amplifier.

Hold

During the hold phase the mirror is turned off and instead the feedback is created between the output and the negative pin (disconnecting it form the input).

Now the capacitor applies its voltage to the positive and the output will be

$$V_{out} = V_H + V_{offset} = V_{in} - V_{offset} + V_{offset}$$

$$V_{out} = V_{in}$$

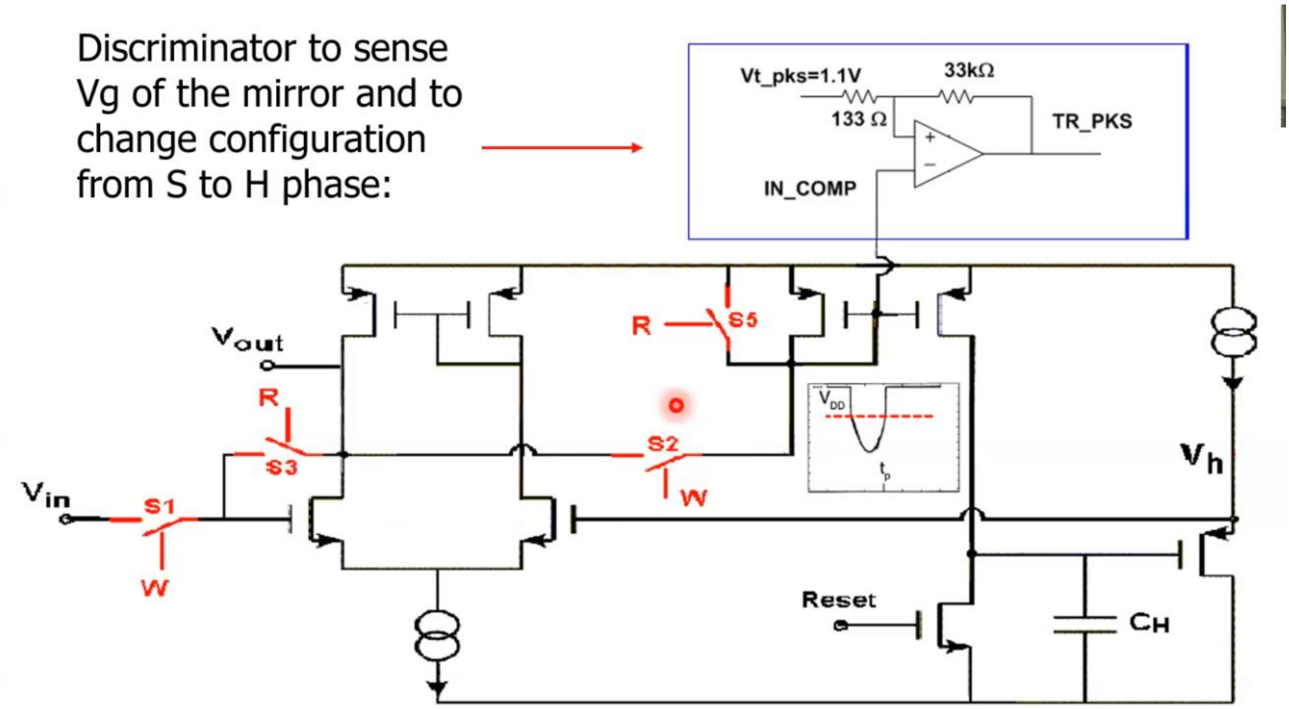
Additional components

To implement this solution we need 2 switches

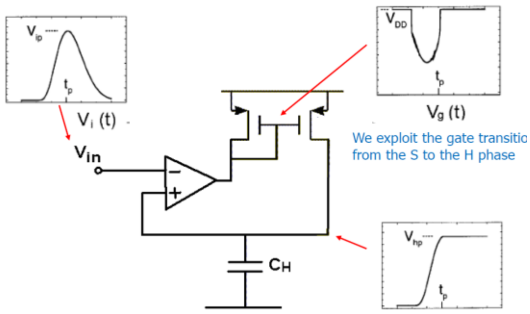
- A switch $S1$ to connect and disconnect the input
- A switch $S3$ to open or close the feedback between output and negative input

Additionally we can add for good practice

- The switch $S2$ to disconnect the output form the mirror even if it is turned off
- The switch S5 to actively switch off the mirror even though it would switch off automatically

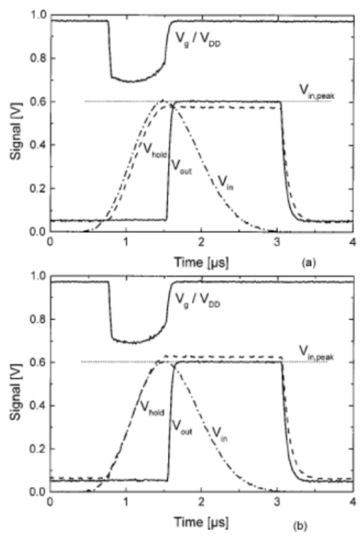


How to drive the switches



We know that the mirror will turn off automatically when we reach the peak.
What we do is connect a Schmitt trigger to the gates of the mirrors thus detecting when it will turn off (the voltage will increase).
Now we can generate a signal which will drive the switches when the peak is reached.

Example of the waveforms in the circuit

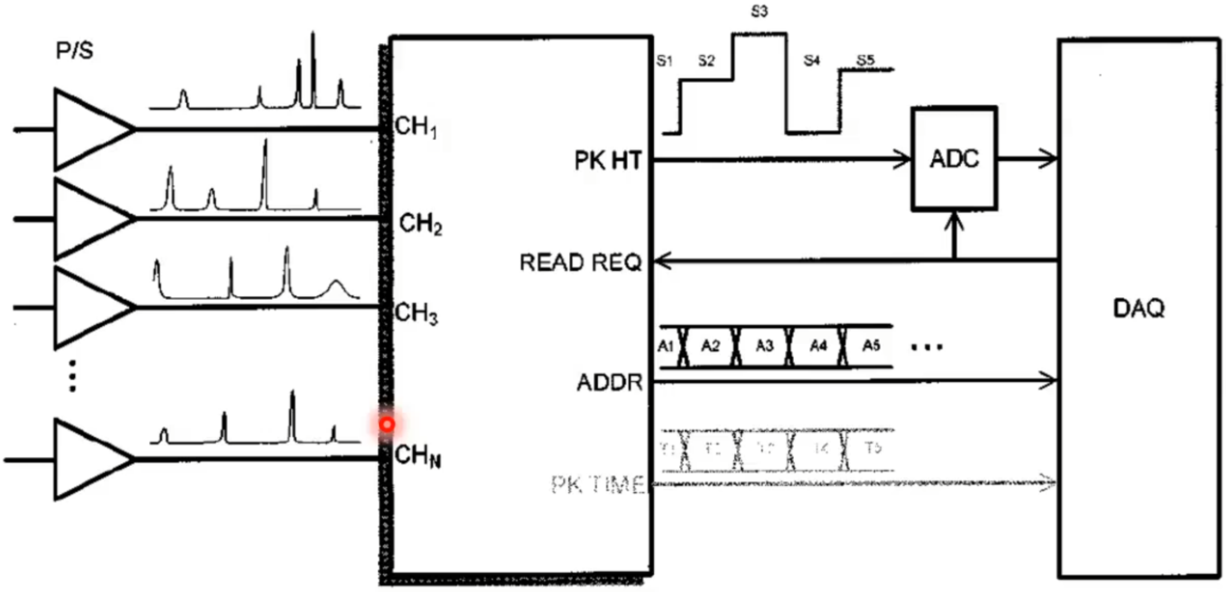


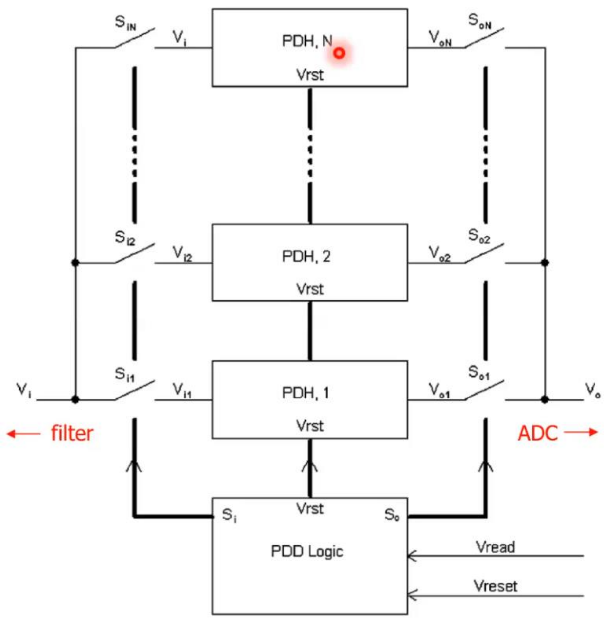
example of response in case of offsets with opposite sign

(V_g: gate voltage of the mirror MOSFETs)

Use of peak stretchers for derandomization of the events

In medical imaging the pulses arrive at a random rate, however an ADC converts signals at a constant rate. This means that we might lose some pulses because they arrive too close to one another.
The solution is a brute force approach: the conversion rate of the ADC is usually selected as 10 time sthe average event rate.





Additionally what we can do is to create a series of independent peak stretchers and utilize them as a form of analog memory to store the pulses while the ADC is busy.

This process is called derandomization: we store random pulses and allow the ADC frequency and the event frequency to match, this allows us to relax the requirements for the ADC frequency which now can match the average pulse rate.

Timing techniques

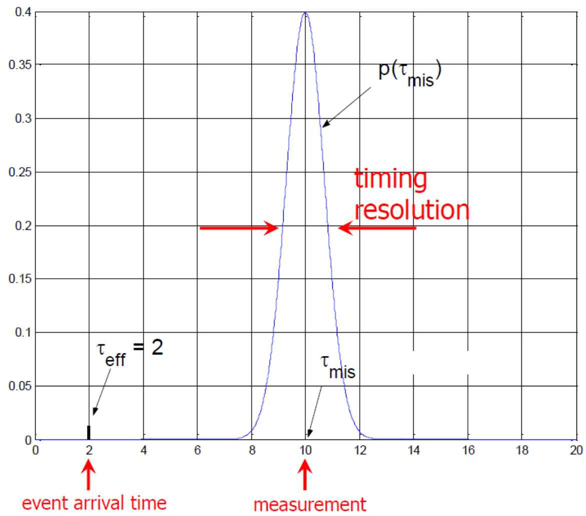
We want to measure the time of occurrence of an event or the time difference between events like in the case of γ rays couples in PET.
We also need to be able to match the time stamp with the amplitude measurement which could be difficult since the amplitude may be measured after a delay, for example in the previous circuit using the derandomize completely cancels the timing information of the pulses, so we would need a separate circuit operating in parallel to record the timing and then an event rebuilder to associate the amplitude and the time.

Result of a timing measurement

Here we can see a result of a typical timing measurement

- The event arrives at the time τ_{eff}
- The measurement is at time τ_{mis} which is delayed with respect to τ_{eff}

If we repeat many time the same measurement the result will not be always equal but will present a statistical distribution. The width of this distribution is the timing resolution.



In PET we are not interested in the delay between τ_{eff} and τ_{mis} as it will be present for all pulses we measure, instead we are much more interested in the timing resolution.

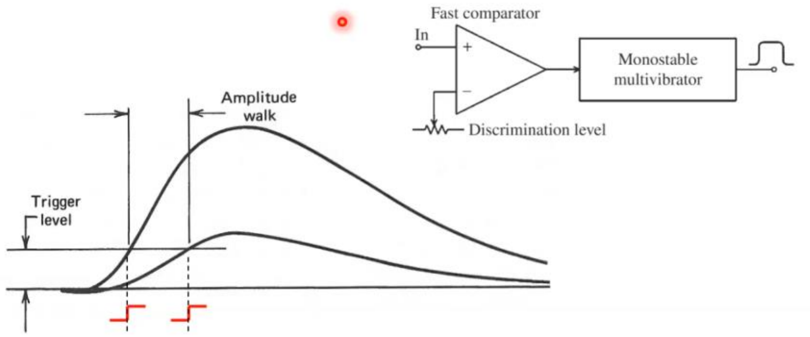
We are going to see 3 techniques for timing measurement

- Leading edge triggering
- Zero cross timing
- Constant fraction timing

The starting signal for these techniques will be the preamplifier Heaviside step because we cannot accept the filtering effect of the shaper since we wish to consider the highest frequency components.

Doing this will however mean that we need to consider the effect of the preamplifier noise which is strongest before filtering.

Leading edge triggering



This is the simplest solution we use a comparator to verify when the input crosses a threshold and thus generate a signal.

Amplitude walk

The leading edge detector will trigger at different moments depending on the signal amplitude, a signal with a higher amplitude will reach the threshold sooner than one with a lower amplitude arriving at the same moment.

This could be corrected utilizing the amplitude information since the 2 are correlated with a deterministic relationship, this can also be done in the case in which we are interested in a single amplitude $511keV$ for PET, so the variation is fixed.

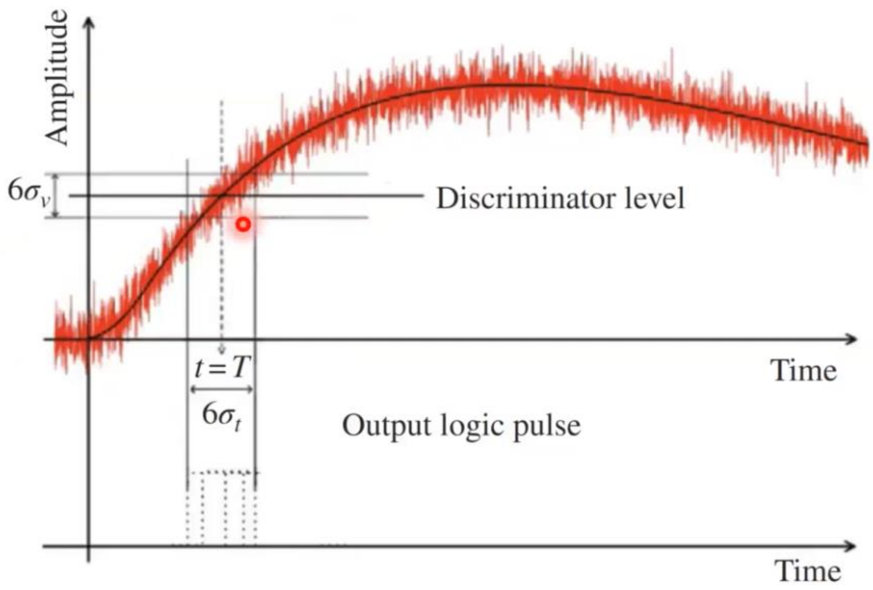
Relationship with different parameters

The amplitude variation δy between different events can be reduced if

- We lower the threshold
  - Problem the noise creates a jitter which becomes more and more significant as the threshold is lowered
- Shorter rise time
- Larger signal

Time jitter

The noise superimposed to the signal will cause a jitter as the signal can cross the threshold in different moments

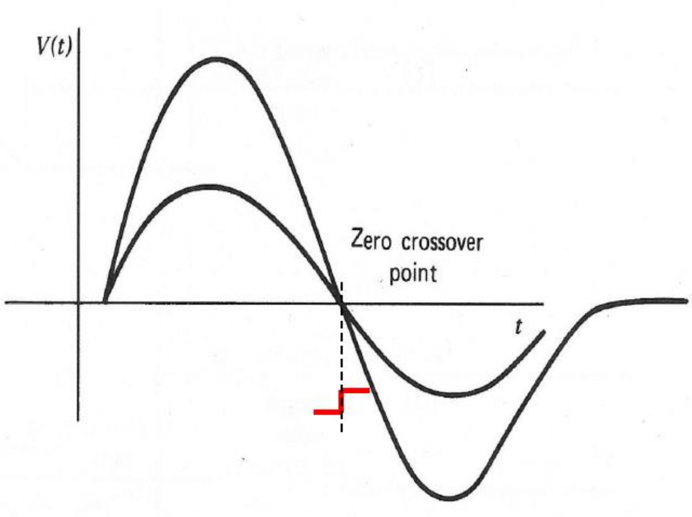


$$\sigma_T = \left.\frac{\sigma_v}{dv/dt}\right|_{t=T}$$

The time jitter becomes less relevant as the slope of the signal becomes steeper.

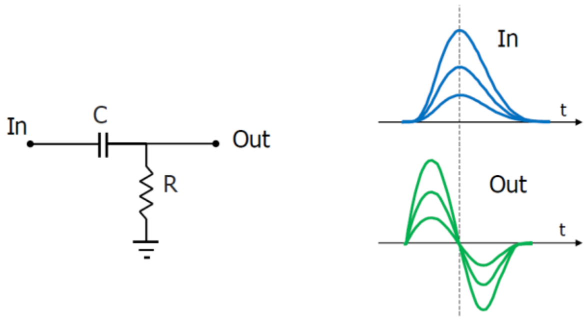
Zero crossing time

This solution eliminates the problem of the amplitude walk, what we do is generate a bipolar pulse rather than a unipolar pulse, and we check not the point in which it passes a threshold but the point in which it crosses the zero, as this point is amplitude independent.



Using a derivator to obtain a bipolar pulse

From the detector we obtain a unipolar pulse to turn it into a bipolar pulse we pass it through a derivator ($C - R$ circuit with a small time constant as this allows us to obtain a pulse as symmetric as possible).



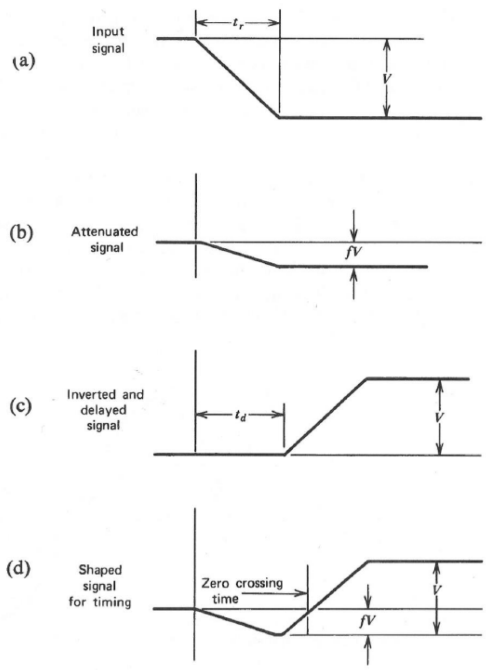
We can see that the peaking time is the same independently form the output, and the derivative is by definition zero at the peak, for this reason the bipolar pulse obtain has a zero crossing always at the same moment.

This solution additionally removes the jitter introduced by the variation in the signal amplitude because of the detector intrinsic variance, the charge at the output is not always constant but fluctuates, since this solution is independent from the amplitude of the pulse this problem disappears.

Constant fraction timing

Similar to the zero crossing except we use a different method to shape the bipolar pulse

1) The input signal $y_{riv}(t)$ is taken and attenuated by a factor f
2) The result is then added to an inverted replica of $y_{riv}(t)$ delayed by a time τ_d, we just need to pass the signal through a coaxial cable long enough (100ns for meter)



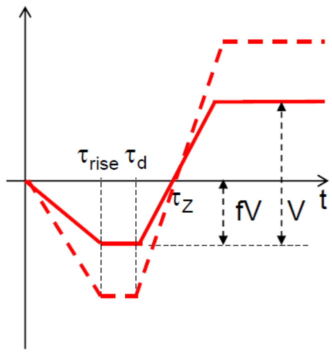
Typically

- The best attenuation factor ranges form 0,1 to 0,2
- The delay time has to be larger than the rise time of the pulse

The resulting pulse is

$$y_{bip}(t) = f \cdot y_{riv}(t) - y_{riv}(t)(t - \tau_d)$$

We can demonstrate that the zero crossing of this resulting pulse will be independent from the amplitude



$$y_{bip}(\tau_z) = 0$$

$$0 = -fV + \frac{V}{\tau_r(\tau_z - \tau_d)}$$

We get that the zero crossing depends on

- The attenuation factor f
- The rise time τ_r
- The delay τ_d

$$\boldsymbol{\tau_z = f\tau_r + \tau_d}$$

Advantage

It is simpler to implement and less expensive than the zero crossing using the derivator

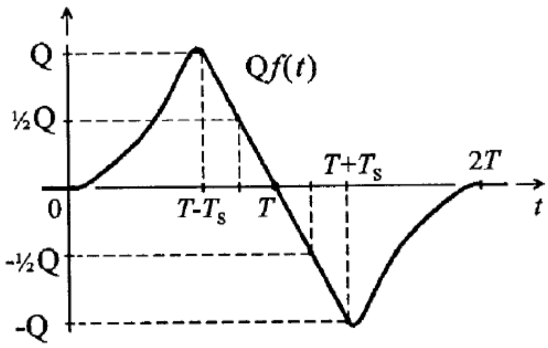
Disadvantage

We have no filtering so we have more noise

Time resolution in the zero crossing technique

We explained how we can obtain a bipolar pulse, now we want to analyze the effect of the noise on the measurement.

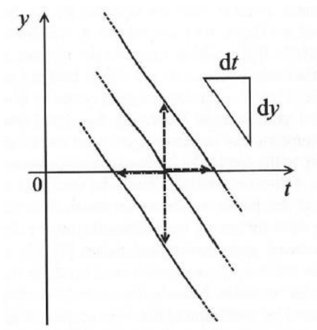
The noise can be considered a statistical fluctuation of the signal which will be consequently translated higher or lower.



We can approximate and say that around the zero crossing point where is a linear relationship between amplitude and timing fluctuations

$$\sigma_\tau = \frac{\sigma_{bip}}{y'_{riv}|_{t=T}}$$

Before starting



- When we use the constant fraction method the noise σ_{bip} remains almost unchanged so the solution is sued for cases where the electronic noise is negligible, like when we have internal gain in the detector (PMT for example)
- When we utilize zero crossing the filter can be modeled to optimize the transfer and reduce σ_{bip}

Important

The formula above is correct for both cases, note that as already explained with this solution the statistical fluctuation of Q is not relevant.

Computing the time jitter

We simply substitute

$$\sigma_\tau = \frac{\sigma_{bip}}{y'_{riv}|_{t=T}} =$$

First thing we do is we normalize by the amplitude of the signal by multiplying and dividing by y_{MAX}

$$\sigma_\tau = \frac{\sigma_{bip}}{y'_{riv}|_{t=T}\frac{y_{MAX}}{y_{MAX}}} =$$

Then I separate the normalized derivative $\frac{y'_{riv}|_{t=T}}{y_{MAX}}$ and obtain

$$\sigma_\tau = \frac{\sigma_{bip}}{y_{MAX}} \cdot \frac{1}{\frac{y'_{riv}|_{t=T}}{y_{MAX}}}$$

I can now repeat and normalize by the zero crossing time T, since in the derivative, the time is indicated at the denominator the normalized derivative is multiplied by the time so we get

$$\sigma_\tau = \frac{\sigma_{bip}}{y_{MAX}} T \cdot \frac{1}{y'_{riv}|_{t=T} \cdot \frac{T}{y_{MAX}}}$$

Now if we group

- $\frac{\sigma_{bip}}{y_{MAX}} = \sigma_{bip}^{\%}$ which indicates the noise normalized by the amplitude
- $y'_{riv}|_{t=T} \cdot \frac{T}{y_{MAX}} = \hat{y}'_{bit}|_{t=T}$ the **normalized derivative**
- T we leave as is

$$\boldsymbol{\sigma_\tau = \sigma_{bip}^{\%} \cdot \frac{1}{\hat{y}'_{bit}|_{t=T}} \cdot T}$$

Note that the normalized noise can be represented as a ratio of charge and thus we can write

$$\boldsymbol{\sigma_\tau = \frac{ENC}{Q_S} \cdot \frac{1}{\hat{y}'_{bit}|_{t=T}} \cdot T}$$

We can see that the time jitter will depend

- Linearly with the noise to signal ratio $\frac{ENC}{Q_S} = \sigma_{bip}^{\%}$
- Inversely with the normalized derivative
- Linearly with the bipolar pulse duration (time for the zero crossing)

Reducing the jitter

To reduce the time jitter we need to consider the 3 terms, let's not ethe formula for the ENC since it is important for our considerations

$$ENC^2 = (C_D + C_G)^2 a \frac{1}{\tau} A_1 + (C_D + C_G)^2 c A_2 + b \tau A_3$$

Where

- a is the series white noise
- b is the series parallel noise
- c is the $\frac{1}{f}$ noise

Increase the signal

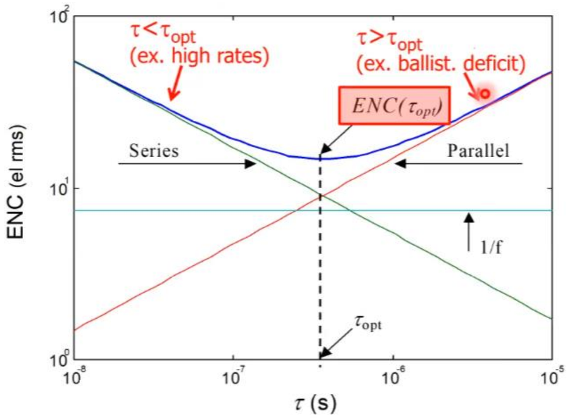
So we select the scintillator so that it releases as much charge as possible, for the same reason we want the photodetector to have a quantum efficiency as high as possible.

Reduce the noise ENC

Minimizing the noise means selecting the optimum shaping time and thus the zero crossing time T, so we need to consider a tradeoff between the increase of the noise when we reduce the shaping time as well as the proportional reduction of T.

The dominating noise contribution for a short shaping time will be the

- Not the parallel noise because it is directly dependent on the shaping time
- Not the $\frac{1}{f}$ noise because it is independent form the shaping time
- **The series noise is inversely dependent on the shaping time so will be the noise component increasing when we reduce it**.



Proportionality

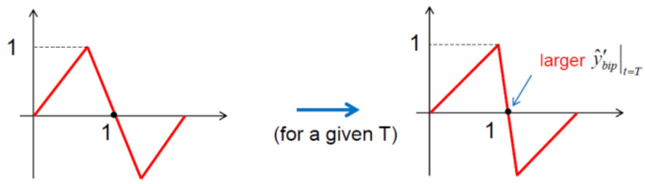
The series noise and thus the ENC increases with the square root of the inverse of the shaping time

$$ENC \propto \frac{1}{\sqrt{\tau}}$$

We have instead a linear dependance when considering the time jitter so overall we will want to minimize the shaping time to minimize the time jitter.

Increasing the normalize derivative

We need to consider that the ENC depends on the shape of the pulse through the filter coefficients, and thus will also depend on the normalized derivative.



We noted that

- A_1 is the integral of the derivative of first order so it increases if the signal is steeper
- A_3 is the integral of the derivative of zero order so it increases if the signal has a larger area
- A_2 is the integral of the derivative of half order

So the coefficient A_1 will increase if we make the system steeper and thus the series noise will increase

Reducing the zero-crossing time VS the preamplifier rise time

For what we said before it appears that there is nothing limiting us form minimizing the shaping time to reduce the jitter.

These considerations have however been made considering an ideal Heaviside step coming form the detector.

We know that this is not true in the real case and this is the reason why we face ballistic deficit, and that the best way to decrease this effect is to increase the shaping time.

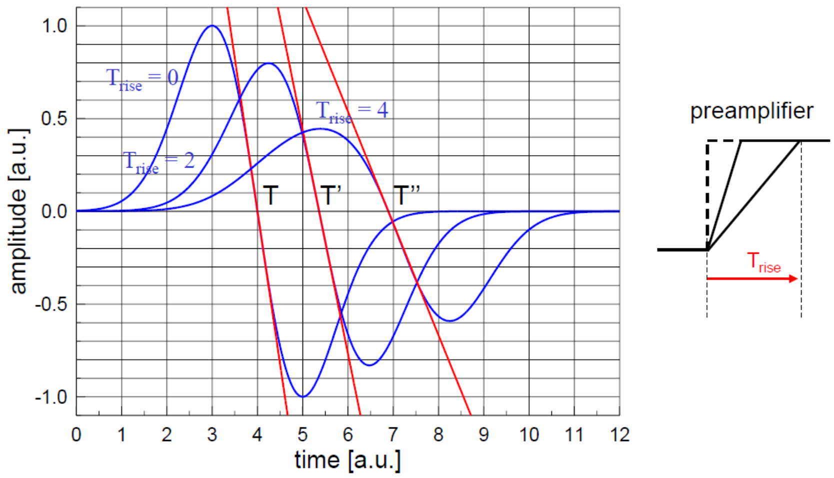
If the preamplifier ramp is not steep enough this will affect the operation of the derivator used to create the bipolar filter.

Effect of a progressively slower preamplifier

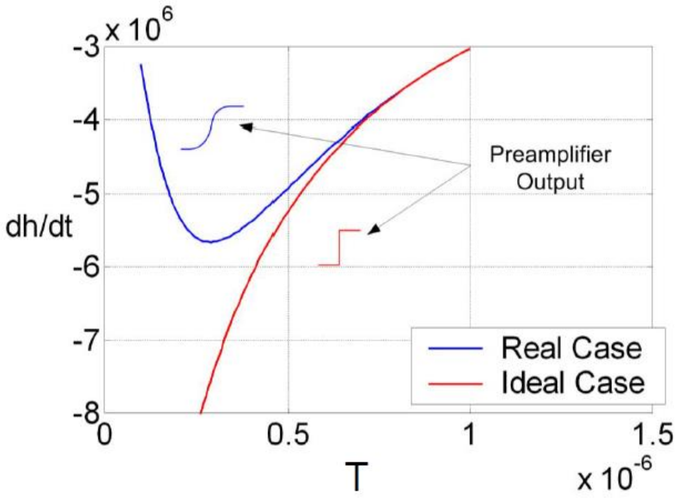
We can see the simulation results of feeding the same bipolar pulse to preamplifiers with a progressively increasing rise time.

The zero crossing slope becomes less steep additionally the zero crossing time T is increasing.

We can expect a much worse time jitter

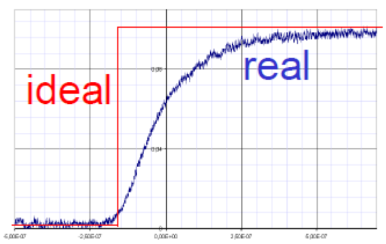


Below we do the opposite we
feed to the same preamplifier pulses with increasing zero crossing time



On the vertical axis we have the derivative at the zero crossing time while on the horizontal axis we have the zero crossing time of the pulse.

In red we have the derivative in the ideal case of an Heaviside step for the preamplifier.



In blue we have a real preamplifier, as we can see at a certain point decreasing the zero crossing causes the derivative to become less steep rather than continuing to go towards $-\infty$.

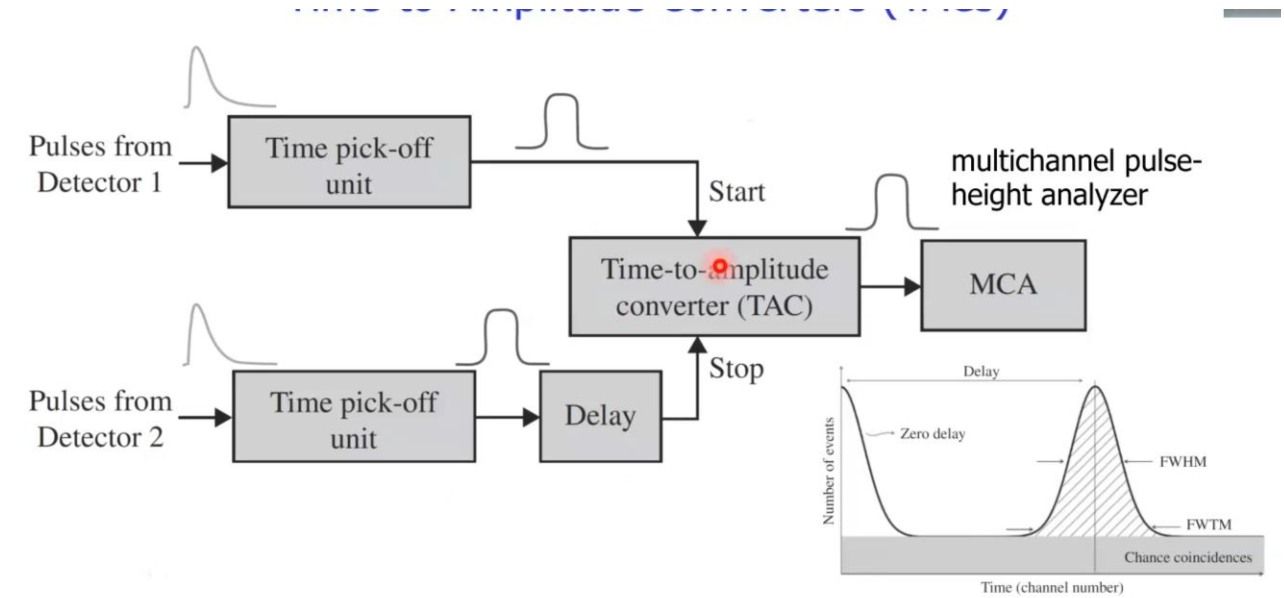
So we have an optimum shaping time also for timing which will differ form the optimum shaping time for amplitude for this reason e have 2 parallel processing circuits.

Time interval measuring techniques

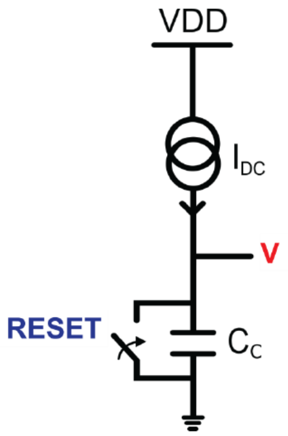
There are 2 techniques to measure coincidences, something that is very important for PET where we need to compare 2 measurements each with its jitter.

TAC time to amplitude converters

A TAC measures the time interval between pulses to its starts and stop inputs and generates an analog output whose amplitude is proportional to the time interval, the stop pulse is normally delayed to ensure that the stop pulse arrives after the start pulse

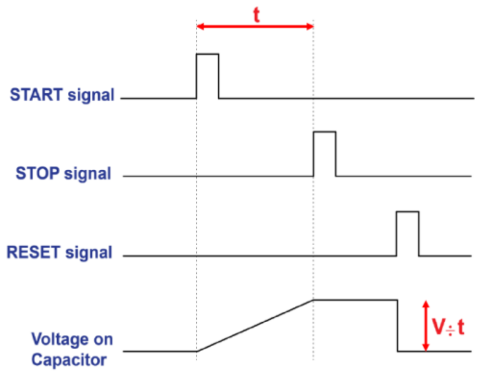


Conversion of the time interval into a pulse amplitude



The start signal opens the reset switch so the capacitor starts to integrate the current and we obtain a ramp.

When the stop signal arrives we stop the current generator so that on the capacitor we are left with a voltage which is linearly proportional to the delay between the 2 timings.

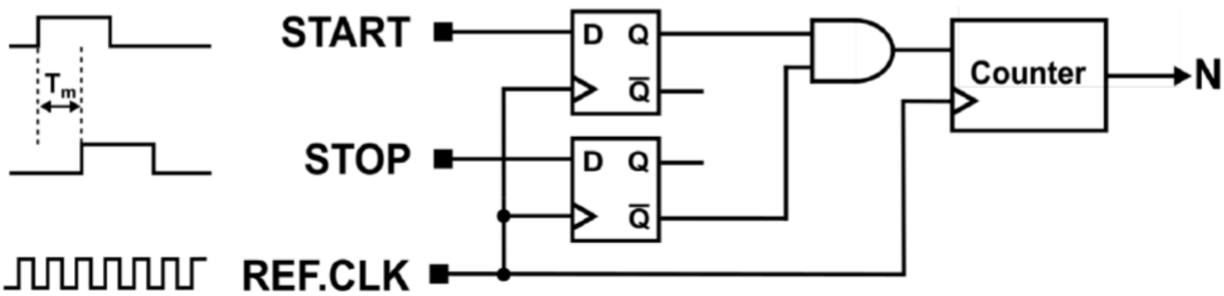


We have an output of zero if the signals are coincident

Critical case

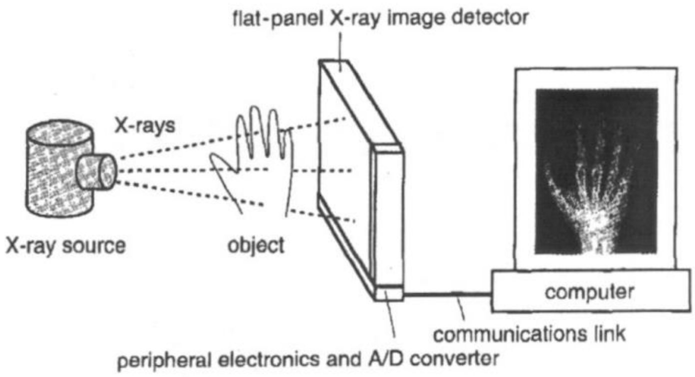
If the 2 signals are coincident but the stop passes through a shorter cable we may incur in the critical case of having the stop before the start, for this reason the stop signal is typically delayed.

Time to digital converter TDC
This circuit measures directly the time by counting clock cycles
- We start counting when the start signal arrives
- We stop counting when the stop signal arrives



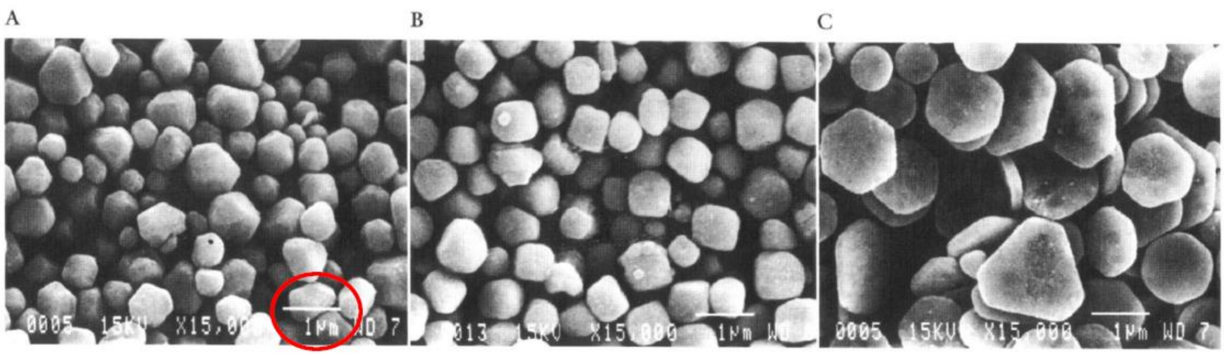
The time resolution matches the clock period.

Digital radiography
Radiography structure



We have an X ray generator and a collector to measure the X ray that pass through the body.

Analog quantization
In the past the radiography were recorded in an analog formula to photographic slaps which had granules inside of them which reacted with x rays change their color like old photographic film, the size of these granules were about $1\mu m$



Photographic emulsion for traditional analog radiography

Digital quantization
Now days measurement is executed utilizing digital converted with a pixel size, this gives us a huge advantage when w eneed to transfer data and we can increase resolution by increasing the number of pixels.

Difference between radiography and SPECT/PET

While SPECT and PET are based on single photon detection, instead in radiography we integrate all the incoming photons without counting them.

The main limitation of a technique based on integration and not photon counting is that we lose the energy information meaning we do not now if we integrate a high amount of energy because a high number of photons or a single high energy photon

Radiographic specifications (remember order of magnitudes)

Clinical Task →	Chest radiology	Mammography	Fluoroscopy
Detector size	35 cm × 43 cm	18 cm × 24 cm	25 cm × 25 cm
Pixel size	200 μm × 200 μm	50 μm × 50 μm	250 μm × 250 μm
Number of pixels	1750 × 2150	3600 × 4800	1000 × 1000
Readout time	~ 1 s	~ 1 s	1/30 s
X-ray spectrum	120 kVp	30 kVp	70 kVp
Mean exposure	300 μR	12 mR	1 μR
Exposure range	30 - 3000 μR	0.6 – 240 mR	0.1 - 10 μR
Radiation (quantum) noise	6 μR	60 μR	0.1 μR

Size

Important give the high size in the order of **tens of centimeters** we can not use integrated silicon technologies as the cost would be too high.

Pixel size

In the order of $100\mu m$

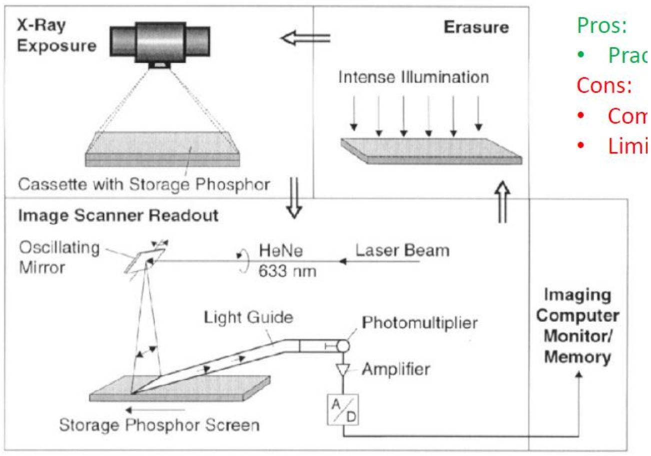
Readout time

This indicates the time that we have available to make the measurement and collect the charge

X ray spectrum

Maximum energy of the X ray photons which allow us to measure the specific body part, denser parts require more energy and vice versa.

Photostimulable phosphor digital radiography system



Operation

exposure

The optical signal is not derived form the light that is emitted in response to the incoming radiation but form subsequent emission when electrons and holes are released from traps in the material.

Raster scanning

The X ray stimulates the electrons which starts filling the traps. This is done pixel by pixel so that we can read the intensity for each one.

Post stimulated luminescence

After we shine red light on the crystal, this raises the trapped electrons to the conduction band. Finally when the electrons relax we have the emission of a shorter wavelength blue light.

After this we heat the slap to free all charges and then restart the process

Pro

Practical and simple

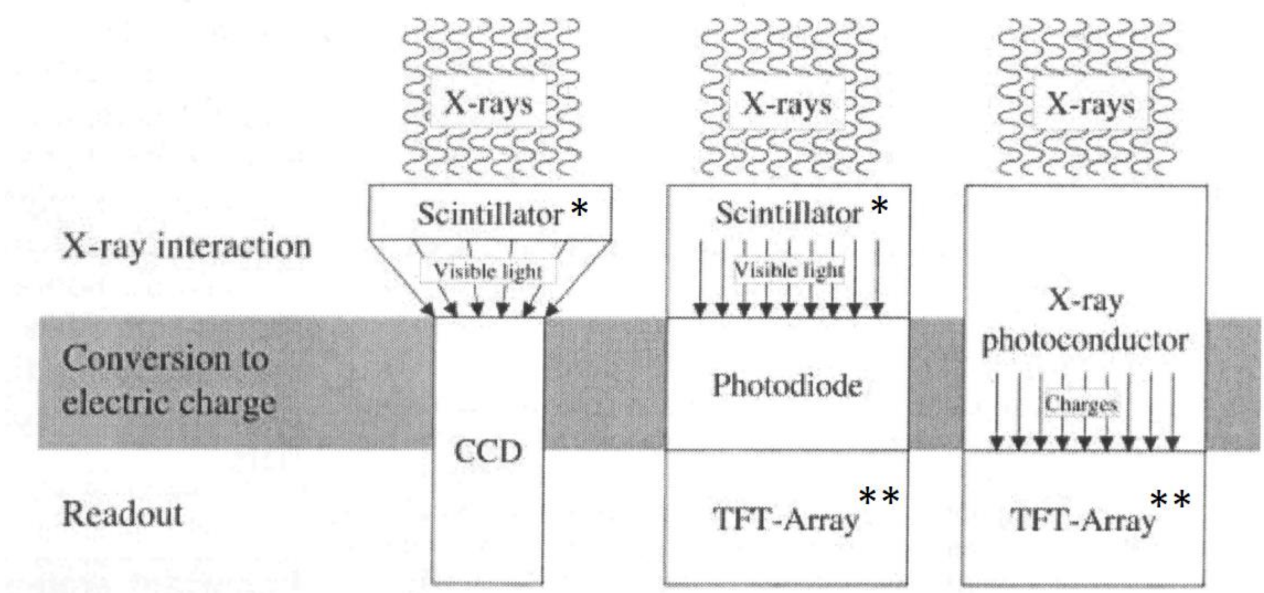
Cons

- Complex readout system
- Limited spatial resolution

Detectors for digital radiography

We know focus on active detectors which allow us to measure the X ray intensity in real time without the complex readout process.

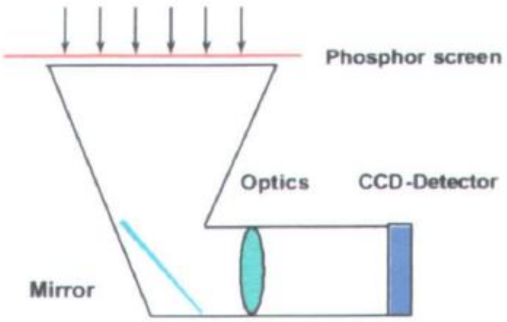
We can divide between indirect conversion using a scintillator and direct conversion

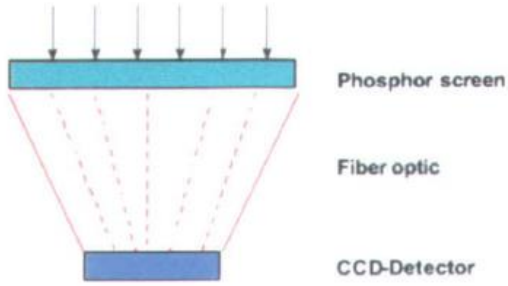


Indirect conversion Direct conversion

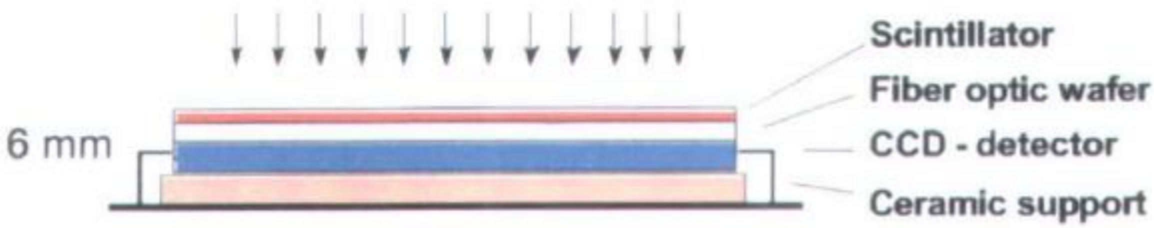
Note CCD is a silicon photodetector and it is relatively expensive for this reason as we can see from the image we make the light converge sot that we need a smaller area CCD to make our measurements.

CCD based readout system



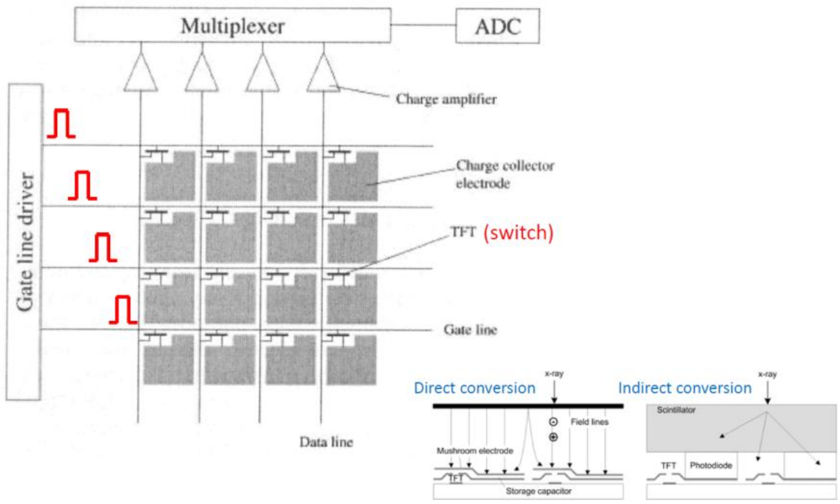


We use optical fibers to squeeze the image of the scintillator onto the CCD smaller area (this has the disadvantage of worsening the resolution).
We can use a 1:1 ratio if the area required is not too high, like for example in the case of mammography.



TFT array (Thin film transistor)

We utilize an matrix of charge collectors to read the charge of the photodiodes, we could use this solution also in direct conversion utilizing a material called photoconductor which converts directly the X ray into charge without the need of combining a scintillator and a photodiode.



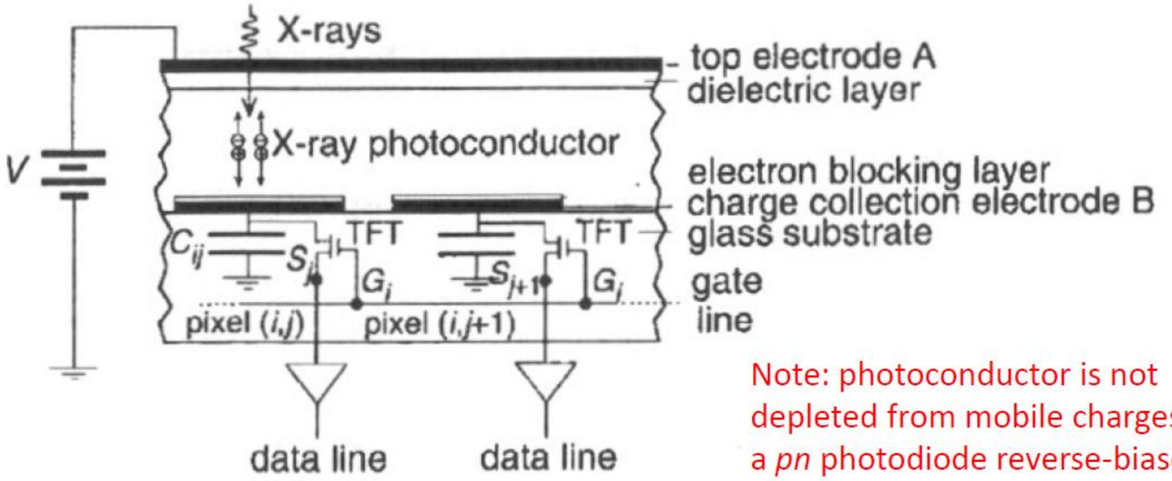
Architecture

A TFT array is a matrix of charge collector electrodes (capacitors)
the X ray supply charge into the capacitors (gray areas in the image) during the exposure time.

Readout

We can not afford to have a dedicated amplifier for each pixel so the readout is done utilizing the same logic as RAM memory.
The cells are divided into rows and columns, we have an amplifier for each column.
Each cell has a switch connecting it to the column and thus to the amplifier.
Readout is performed row by row connecting all the switches of a raw, reading the result form the amplifiers on the columns and then moving on to the next row.

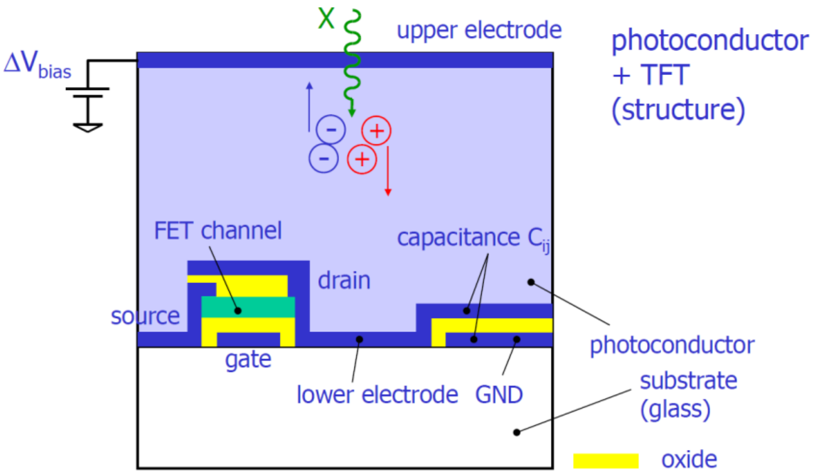
Direct conversion X ray imager (photoconductors)



Note: photoconductor is not depleted from mobile charges (as a *pn* photodiode reverse-biased)

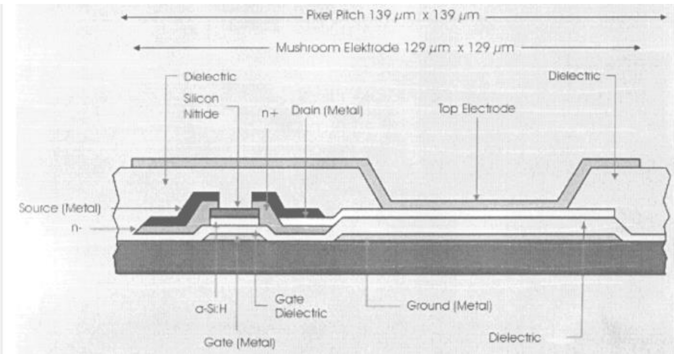
The 2 plates in dark represent the top electrode of the capacitor, the bottom one is instead connected to ground, as we can see we have a transistor to connect it to the data line we can see on the right a more realistic representation.

The drain of the transistor is connected to the top plate of the capacitor while the source is connected to the data line.



TFT technology: Si amorphous hydrogenate (a-Si:H), semiconductor deposited as thin layer from silane gas (SiH_4) in a plasma chamber

Important



We can see that the connection between top of the capacitance and transistor extends over the transistor this is made to extend the area which can collect the charge generated by the photoconductor.
As we can see from the image the top electrode is almost as big as the pixel.

Why do we adopt this Thin transistor technology

We have that the layers in the TFT are all amorphous unlike in the classic cMOS technology.
The production of the device in this case is obtained by spattering layers over layers:

1) First we place the bottom electrode
2) Then we put the oxide
3) Then the channel ecc

Sowe stack the different layers without the need to obtain a specific crystalline structure so the production is much less expensive, we can do this because we are not interested in performances.

Photoconductor

A photoconductor are typically used because when hit by radiation their conductivity changes so we can detect radiation by utilizing them in a voltage divider.
In the case of radiography instead we utilize them because X ray are capable of ionizing them releasing electron hole pairs through photoelectric effect and Compton.

Critical difference with other ionizing detectors

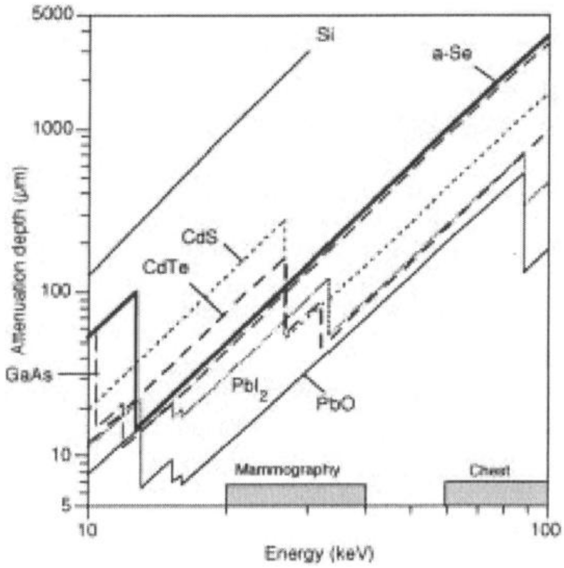
While in a classic pn diode we apply a reverse bias to deplete the volume from charge, the photoconductor is an intrinsically conductive material so we cannot deplete it, so we need to find another way to

- Prevent current from being generated for the bias and not the generation
- Prevent the charges to recombine with the other charges naturally present in the photoconductor

Setting the thickness of the photoconductor

When we decide the thickness L of a photoconductor we face a trade off

- The absorption efficiency decreases if the thickness is too small
$$\eta_{assorb} = 1 - e^{-\mu L}$$
- If the thickness is too large then
  - We increase the probability of recombination/ trapping
  - It is difficult to grow areas with large thickness without defects
  - The bias voltage required to obtain a given electric field F increases



Note on the values

For radiography we have a thickness in the order of a few millimeters while in the case of nuclear imaging we have a thickness of centimeters.

Average travel distance without recombination

To determine the average distance a carrier can travel before recombining and thus the maximum thickness of the photoconductor we consider that

- The average life time is τ
- The carrier speed is equal to the product of the mobility μ and the electric field F

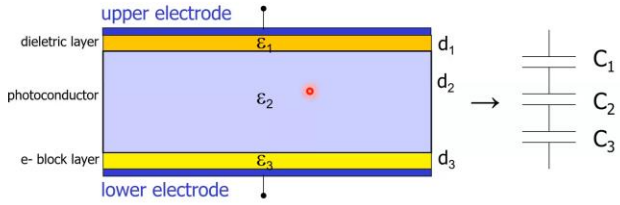
So we get that the maximum thickness is
$$L < \mu F \tau$$

Dark current

We want a larger energy gap to reduce the dark current however to assure a good sensitivity we cannot have a too large energy gap.

Current in the Photoconductor

To minimize the current flow we have a thin layer of dielectric between the contacts of the photoconductor, this addition does not change significantly the behavior of the photoconductor, since at frequency it becomes like having a series of capacitors so the smallest one (the one in the middle with larger thickness) will be dominant.



$1/C_P = 1/C_1 + 1/C_2 + 1/C_3$

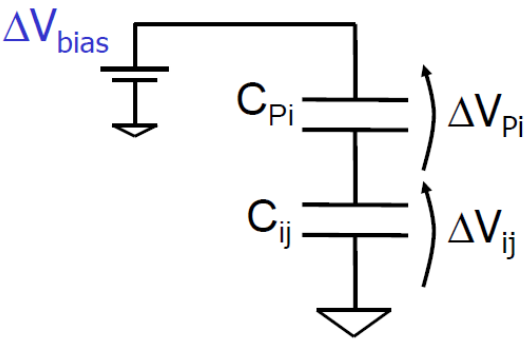
$C_X = e_X S/d_X$ $\Rightarrow$ $\boxed{C_P \sim C_2}$

$\Rightarrow C_1, C_3 >> C_2$

Biasing of the pixel

We have created 2 capacitors in series one is the storage capacitor, and one is the pixel, we want the majority of the voltage to fall across the pixel and not the storage capacitor to do this we need for the storage capacitor to be much larger than the pixel capacitance. This also helps us for the signal transmission since the majority of the charge and thus the signal will be stored in the larger capacitance.

Biasing



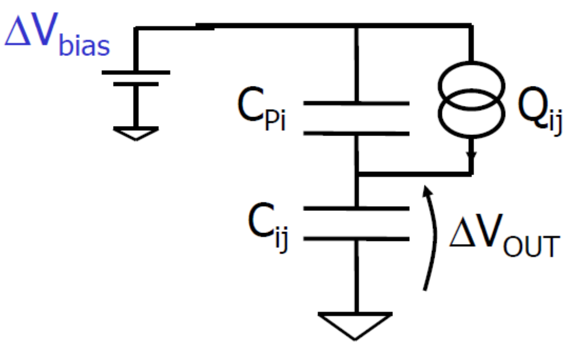
$\Delta V_{Pi} = \Delta V_{bias}\, C_{ij}/(C_{ij}+C_{Pi})$

$C_{ij} >> C_{Pi}$ $(C_{ij} \sim 1\text{pF})$

$\Rightarrow \boxed{\Delta V_{Pi} \sim \Delta V_{bias}}$ $(\sim\text{kV})$

Signal integration

(signal produced by X-ray absorption in the exposure time)

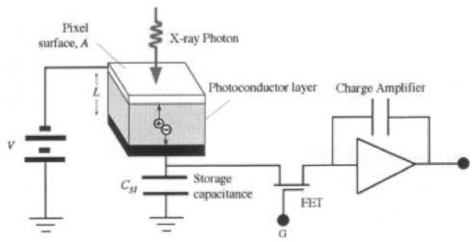


$\Delta V_{OUT} = Q_{ij}/(C_{ij}+C_{Pi})$

$C_{ij} >> C_{Pi}$ $(C_{ij} << C_3)$

$\Rightarrow \boxed{\Delta V_{OUT} \sim Q_{ij}/C_{ij}}$

Readout circuit



when a pixel is connected by switching-on the FET to the common output line connected to the preamplifier:

$V_{OUTpre} = Q_{ij}/C_F$