

Spectral Response of Silicon Image Sensors

Arnaud Darmont, Aphesa (www.aphesa.com), White paper, April 2009

Abstract

Although the general theory of spectral response of silicon photodiodes is known, its detailed response is usually misunderstood or ignored. This paper will describe the simplified general theory and the variations due to various processing technologies.

1. The nature of light and light-matter interaction

Light is an electromagnetic wave presenting wave properties but because of its wave-particle duality it will behave as a particle when entering a material like Silicon. Both the wave and the particle properties of light should be considered when analyzing its interaction with Silicon. The particle associated with light, or with any electromagnetic wave is the photon. A photon is characterized by its energy which is directly related to the wavelength of the light when the radiation is considered as a wave. This relation is

$$E_{ph} = \frac{hc}{\lambda}$$

where h is Planck's constant ($6.626 \cdot 10^{-34}$ Js), c is the speed of light in vacuum and λ is the wavelength. If the photon energy is expressed in electron-volts ($1\text{eV} = 1.602 \cdot 10^{-19}$ J) and if the wavelength is expressed in nanometers then the formula simplifies to

$$E_{ph} [eV] = \frac{1240}{\lambda [nm]} \quad (1).$$

A large amount of photons make a ray of light, this ray of light can be characterized by its wavelength, intensity and the variation of this intensity over time. This translates in the particle domain to the amount of particles crossing an imaginary surface per unit of time.

Particles interact with each other by exchanging particles or energy. A photon interacts with other particles by transferring its energy to the other particle and then disintegrates itself. This exchange of energy happens if the maximum energy that the photon can bring is at least equal to the smallest quantum of energy that the other particle can accept. For a monocrystalline Silicon lattice, this smallest quantum is 1.1eV, referred to as the silicon band gap.

Photons with an energy below 1.1eV will go through silicon without any interaction, the material is transparent. This corresponds, by relation (1), to a wavelength larger than 1125nm.

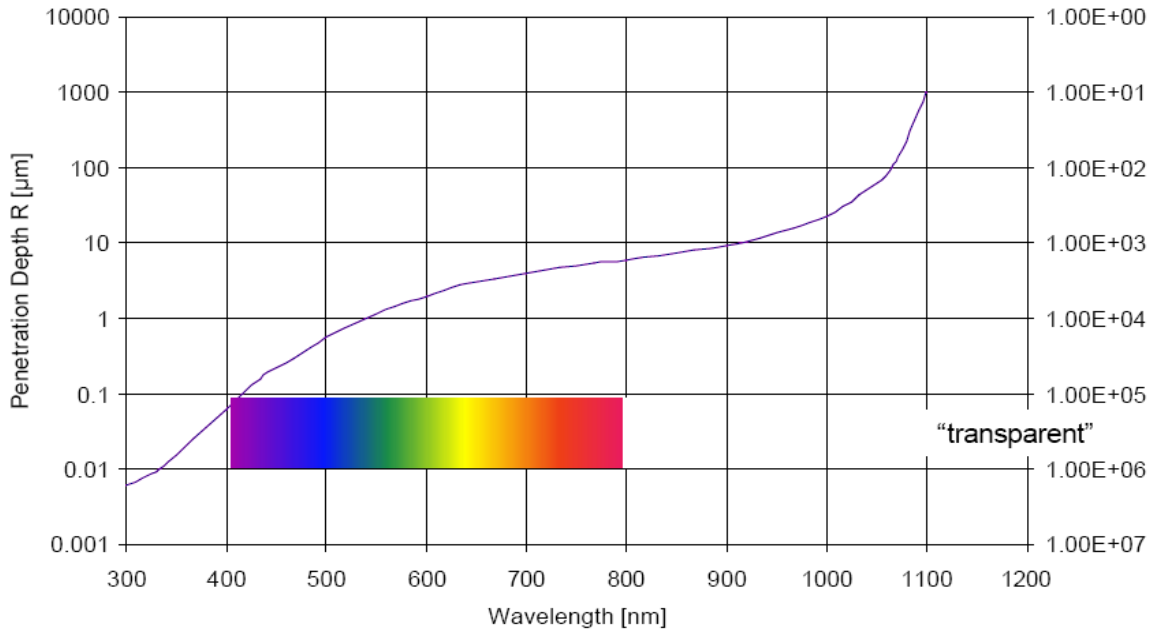
Photons with an energy above 1.1eV will not necessarily interact with silicon. Their probability of interaction depends on how much higher than the 1.1eV band gap the energy of the photon is. This defines an absorption coefficient α [cm^{-1}] depending on the wavelength and the type of material. This absorption coefficient defines which percentage of the photons entering a one centimeter thick material will be absorbed. By reasoning on a infinitesimal thickness of material of thickness dz and integrating over the thickness of the material, one can deduct the Beer-Lambert law giving the

density of photons at a given depth for a given density of electrons entering the material:

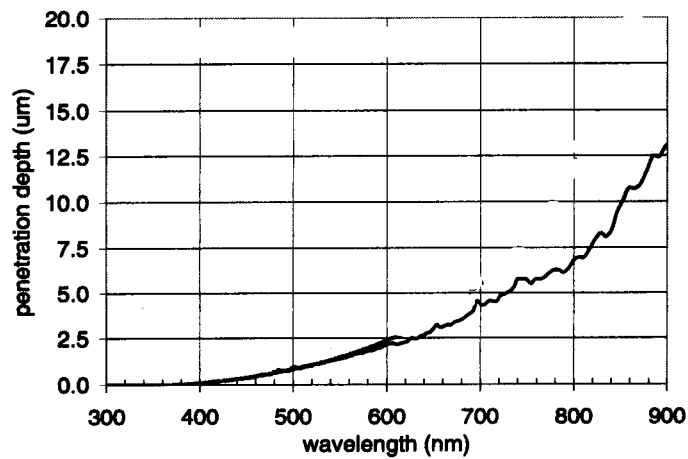
$$I(z) = I_0 e^{-\alpha z}$$

From there, one can define a penetration depth as the thickness of the material required to absorb 1/e (37%) of the incoming radiation.

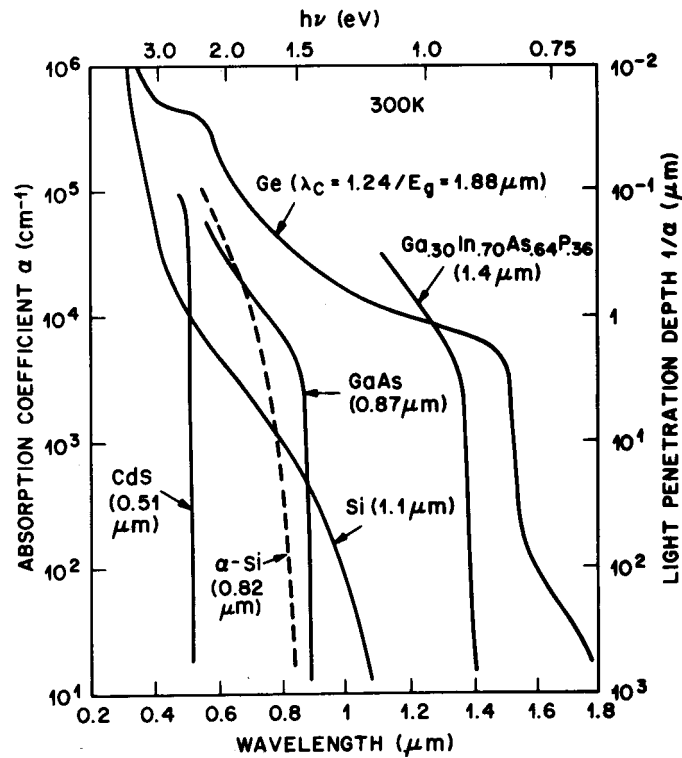
Figure 1 plots the penetration depth of real silicon versus wavelength superposed with the visible light spectrum.



(a)



(b)



(c)

Figure 1: penetration depth of silicon ((b) reproduced from [1] and (c) reproduced from [2]).

Photons that are absorbed by silicon bring energy to the atom they collide with, this energy is used to let one or more electrons escape from their energy band. These electrons become free and can be part of a conduction process. The excess energy heats up the crystal or activates phonon states. The electron slot left behind in the valence band is referred to as a hole and will behave as a heavy positively charged particle.

2. How silicon sees

The freed electrons and holes need to be detected. The most basic detector is the PN junction diode. If light penetrates such a diode, the diode will act as a photodiode, but electron-hole pairs generated outside of the diode junction can also diffuse to the junction and participate in the detection process.

A diode is a junction of two semiconductor materials doped differently. At the junction of the two materials there is a depletion region virtually free of charges. The negative charges drift to the p-side and the positive charges drift to the n-side due to the electric field that builds up between the two materials. This flow of charges is a current that can be measured by an electronic circuit. Figure 2 shows a PN junction in Silicon illuminated from the left side. The most common photodiode is N-well in P-substrate as shown in figure 3.

The total current is the sum of the photocurrent inside the junction, the photocurrent due to carriers that have drifted to the junction and a leakage current which is a function of the absolute temperature.

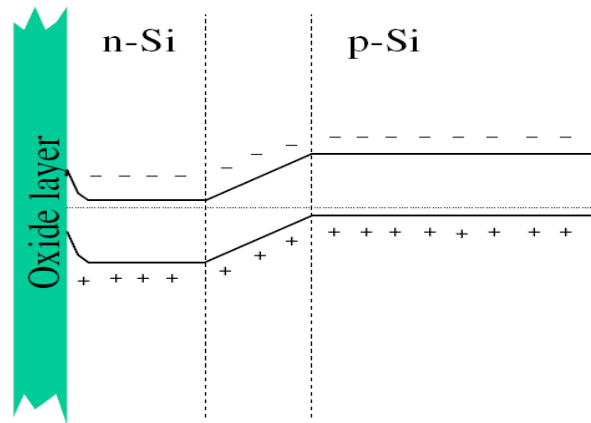


Figure 2: PN junction with energy bands; on the p-side (left) there is a large quantity of holes which makes the Fermi level (dashed line) go close to the valence band; on the n-side (right) there is a large quantity of electrons which make the Fermi level go close to the conduction band; when the two materials are joined the Fermi levels of both sides equalize and the break in the band structure appears at the junction.

Unfortunately the carriers cannot exist forever. When an electron meets a hole they may annihilate each other, this process is called recombination. In Silicon the recombination process is indirect which means that no photon recycling is possible, a recombined pair is lost from the point of view of quantum efficiency. This occurs in both N and P type materials but not in the depletion region. Another important recombination process is due to surface states that exist due to the imperfection of the crystal at its surface and the imperfect matching with the oxidized silicon that grows at the surface of silicon when the material is exposed to the air. Recombination states can also exist due to imperfections anywhere in the crystal.

The farther a carrier from the junction, the more time it takes for it to reach this junction. It then has more chances to be annihilated by recombination. The net contribution of the electron-hole pairs decreases when the distance to the junction increases.

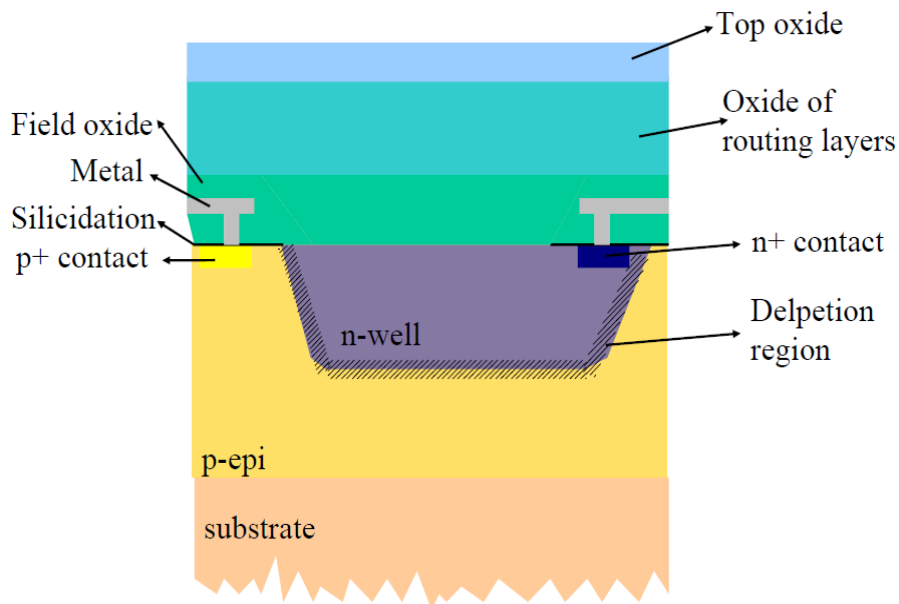


Figure 3: N-well in P-substrate photodiode

Figure 4 represents in green the net contribution of each area of an N-well in P-substrate photodiode. On the left side are the various oxides of the top surface, the depth of the silicon is to

the right. The red curve represents the penetration depth, the deeper you go in the silicon the less electron-hole pairs are generated. Note that the vertical scale is logarithmic.

In the depletion region (region II), the efficiency is 100% as recombination does not occur, all converted photons are useful. In region I and region III, the efficiency decreases when the distance to the depletion region increases. In region I the efficiency suddenly goes to 0% close to the silicon surface. The size of region III is not unlimited, many wafer processes use an epitaxial layer that will limit the size of the region III by adding a deep barrier. This epitaxial layer will cut the responsivity in infrared sooner than the theoretical 1125nm limit from the band gap structure.

As the penetration curve depends on the wavelength, each wavelength will yield to a different plot (figure 5). Measuring the total area of regions I, II and III versus wavelength yields to the plot of figure 6 that represents the light to current conversion or efficiency.

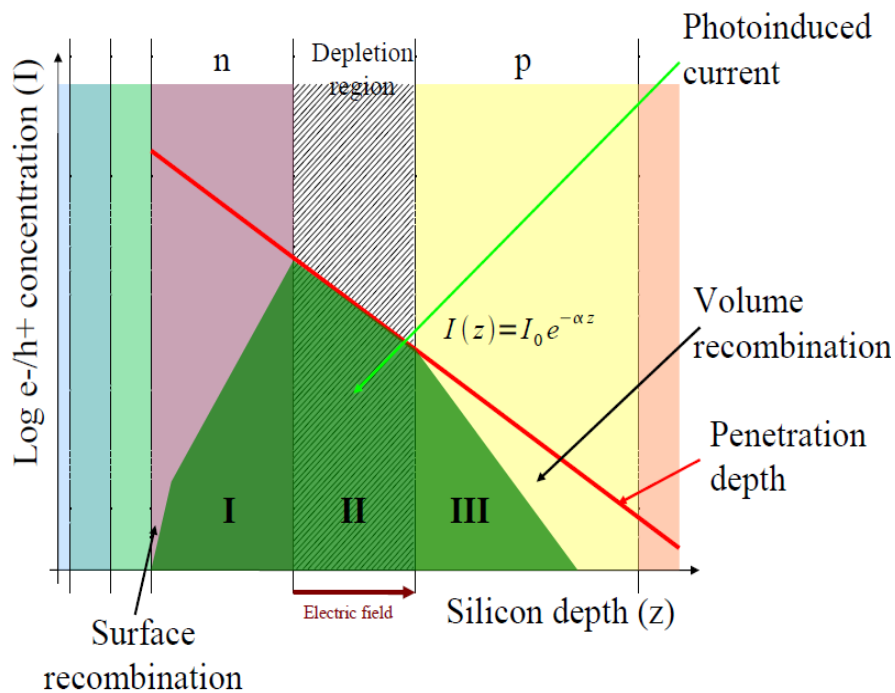


Figure 4: net contribution of each area of an N-well in P-substrate photodiode.

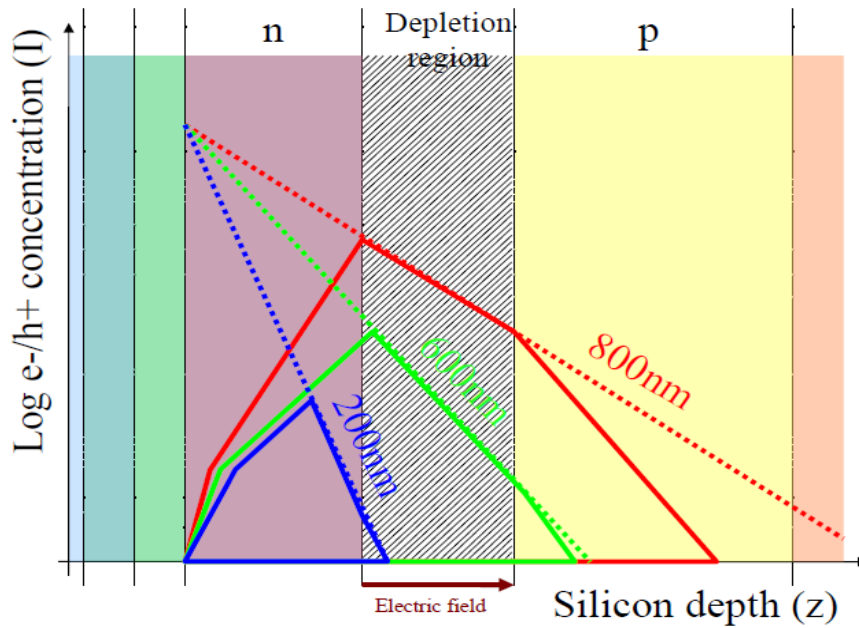


Figure 5: net contribution of each area of an N-well in P-substrate photodiode.

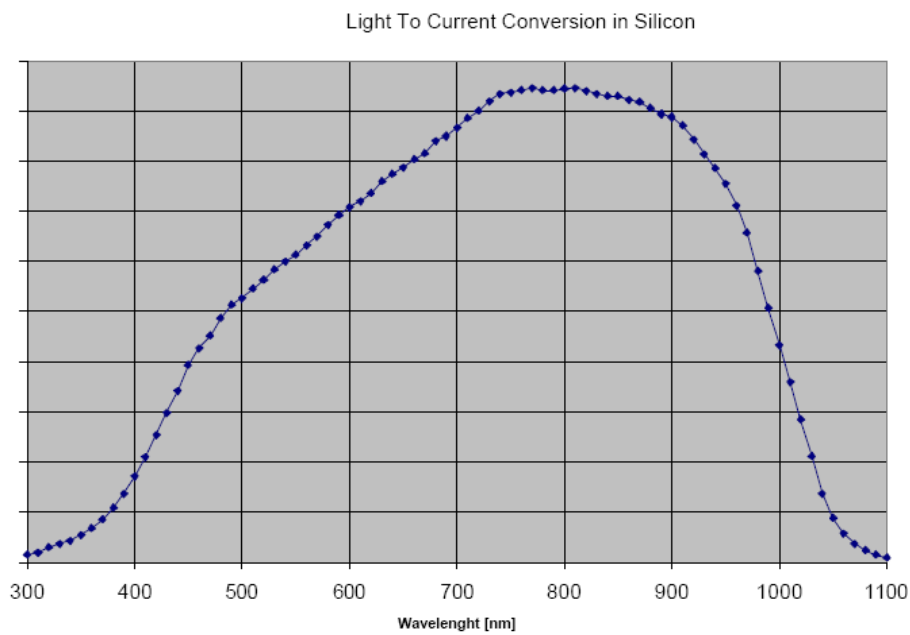


Figure 6: light to current conversion in silicon (example)

Figure 7 summarizes the shape of the fundamental curve. The 100% efficiency line is the theoretical limit if the device is perfect.

Response for small wavelengths increases if the lifetime under the surface is long (small density of surface states).

Response for small wavelengths increases if the junction depth is low (junction close to the surface).

Response for high wavelengths increases if the lifetime of the carriers in the bulk P-substrate is long.

Response for high wavelengths increases if the junction deep (junction close to the surface).

Response for high wavelengths increases if the epitaxial layer is far from the surface.

The line of maximum efficiency is given by the relation

$$R = \left[\frac{\eta \lambda [nm]}{1240} \right]_{\eta=1}$$

where R represents the responsivity and η represents the quantum efficiency.

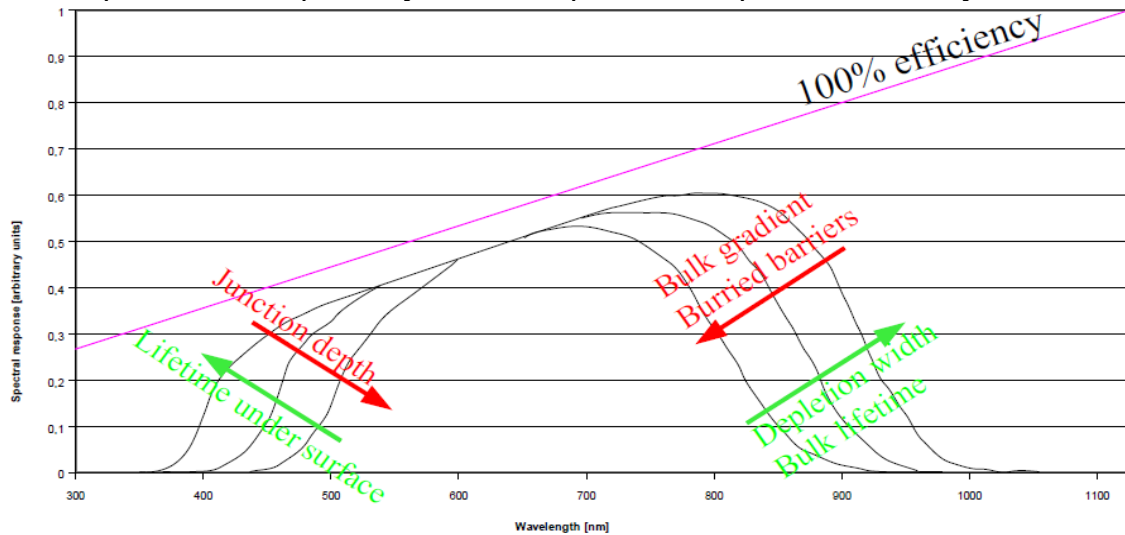


Figure 7: physical effects that affect the shape of the spectral response curve

Quantum efficiency versus wavelength is defined as the ratio between the actual responsivity and the line of maximum efficiency.

Spectral response curves are sometimes plotted versus photon energy instead of light wavelength. In this case the spectral response shows a real plateau in its middle part.

3. Optical effects in material stacks

Light falling on an image sensor reaches the boundary between two mediums: air and silicon. Any electromagnetic wave reaching such a boundary has to undergo changes in amplitude and direction and eventually a split in several rays or a change of polarization.

In the case of a real image sensor there is not only one boundary but a series of boundaries due to the stack of materials used to make the sensor. Figure 8 shows a typical device structure without color filters.

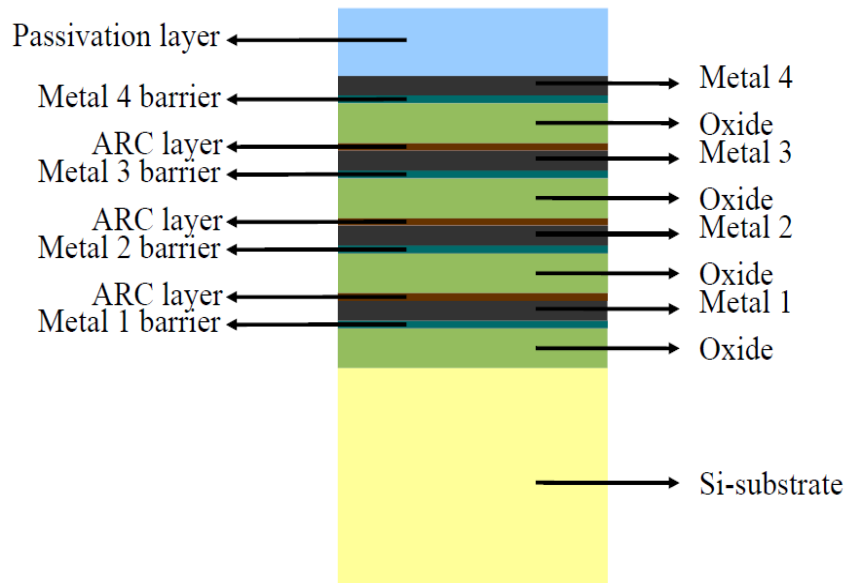
Part of light reaching each boundary will be reflected back (almost specular reflection if the surface is sufficiently flat) and the remainder will be transmitted to the next medium by refraction following Snell's law. The light reflected back will go back to a previous boundary where it will be partly reflected into the sensor and partly refracted outside of the sensor (see figure 9). The light reflected back into the sensor will interfere with the original ray. This structure is similar to the so called Fabry-Pérot interferometer. Several Fabry-Pérot-like interferometers are stacked in a real device according to figure 8. We will consider actual Fabry-Pérot interferometers for the discussion.

As not all light is transmitted into the silicon there is a loss of optical power affecting the spectral response. As the reflectivity of a surface is a function of the wavelength, the spectral response is affected non-uniformly.

As the boundary surfaces usually have low reflectance for best performance the finesse coefficient of the interferometer, defined as

$$F = \frac{4R}{(1-R)^2}$$

is small and the efficiency of the interferometer is low. However, even if this efficiency is low, interference will still be visible in the response as a ripple.



Passivation layer	SiN	1µm
Metal 4	AlCu	800nm
Metal 4 barrier	Ti/Tn	90nm
Oxide	SiO2	950nm
ARC layer	Ti/Tn	120nm
Metal 3	AlCu	400nm
Metal 3 barrier	Ti/Tn	90nm
Oxide	SiO2	950nm
ARC layer	Ti/Tn	120nm
Metal 2	AlCu	400nm
Metal 2 barrier	Ti/Tn	90nm
Oxide	SiO2	950nm
ARC layer	Ti/Tn	120nm
Metal 1	AlCu	400nm
Metal 1 barrier	Ti/Tn	90nm
Oxide	SiO2	940nm
Silicon substrate EPI layer	Si	15µm
Silicon substrate	Si	600µm

Figure 8: typical wafer structure with example materials and thicknesses

The period of the ripple is non-constant. Fabry-Pérot theory tell us that the width of the peaks and the distance between two peaks are linked by the following relations

$$\Delta \lambda = \frac{\lambda_0^2}{2n l \cos(\theta) + \lambda_0}$$

$$\frac{\Delta \lambda}{\delta \lambda} = \frac{\pi}{2 \arcsin((1/\sqrt{F}))}$$

where $\Delta \lambda$ is the distance between two peaks, $\delta \lambda$ is the full-width half-maximum of the closest peak, λ_0 is the central wavelength of the closest peak, θ is the refraction angle of the interfaces and n is the refractive index of the material. Figure 10 shows the shape of the interference peaks in the spectral response. The ratio $\Delta \lambda / \delta \lambda$ is called finesse. Figure 11 plots the transmittance of a Fabry-Pérot interferometer versus the wavelength of the incoming light.

These formula, as well as figure 11, also show that the response curve is a function of the incidence angle of the light on the sensor. The spectral response of a camera (the system) will then depend on the chief ray angle of the lens.

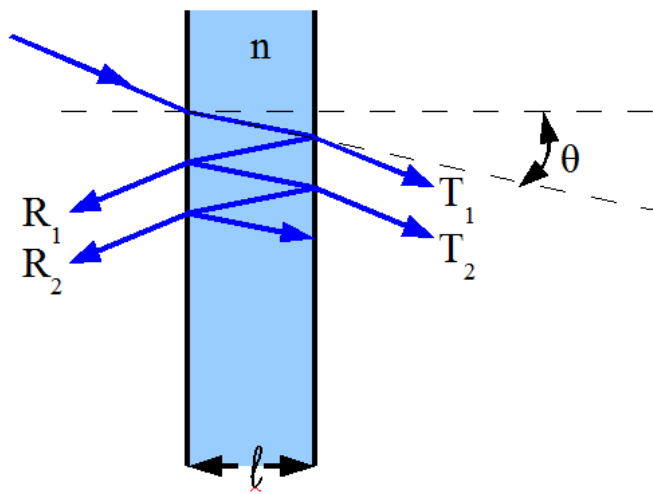


Figure 9: a Fabry-Pérot interferometer

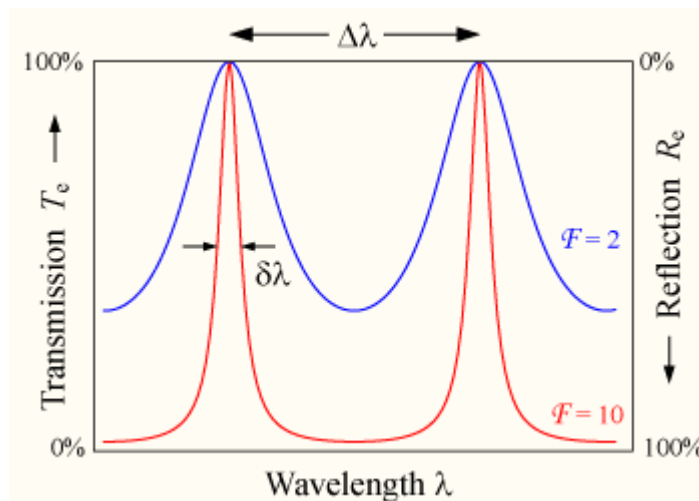


Figure 10: interference peaks in the spectral response for two values of the ratio $\Delta \lambda / \delta \lambda$ (reproduced from [3])

As the distance between two consecutive peaks depends on the central wavelength of these peaks and as the full-width half-height of the peaks depends on the distance between two peaks, it is clear

that the distance between two peaks increases with wavelength.

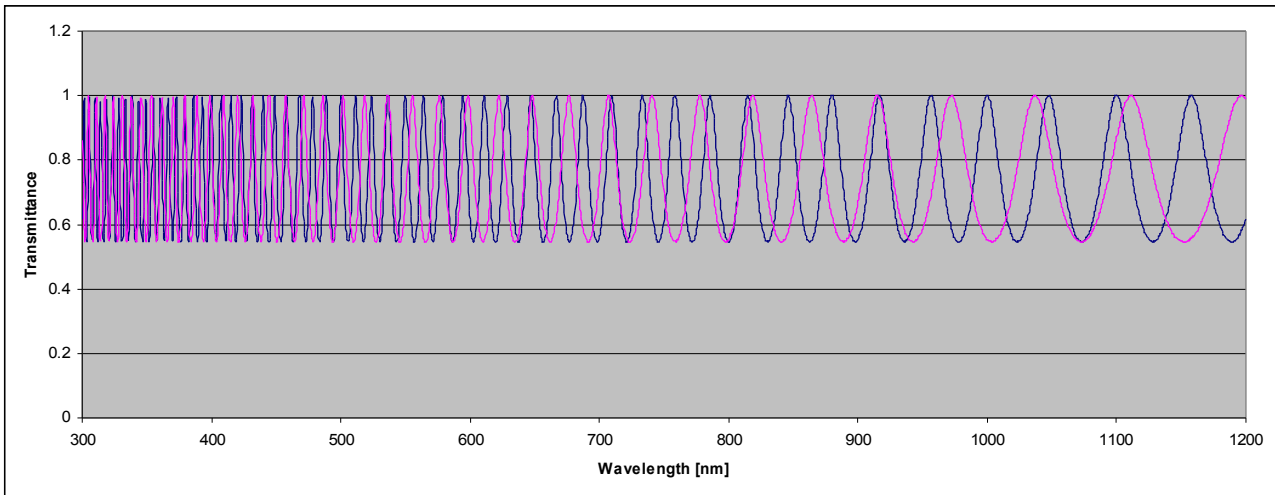


Figure 11: transmittance of a Fabry-Pérot interferometer over wavelength for $n=1.1$, $R=0.15$, $l=10\mu\text{m}$ and $\theta=0$ degrees and 45 degrees.

A series of Fabry-Pérot may generate responses like in figure 12 depending on their properties. Fortunately for Silicon, the layers or oxide and ARC that are present on top of Silicon match very well and the secondary ripple is unnoticed.

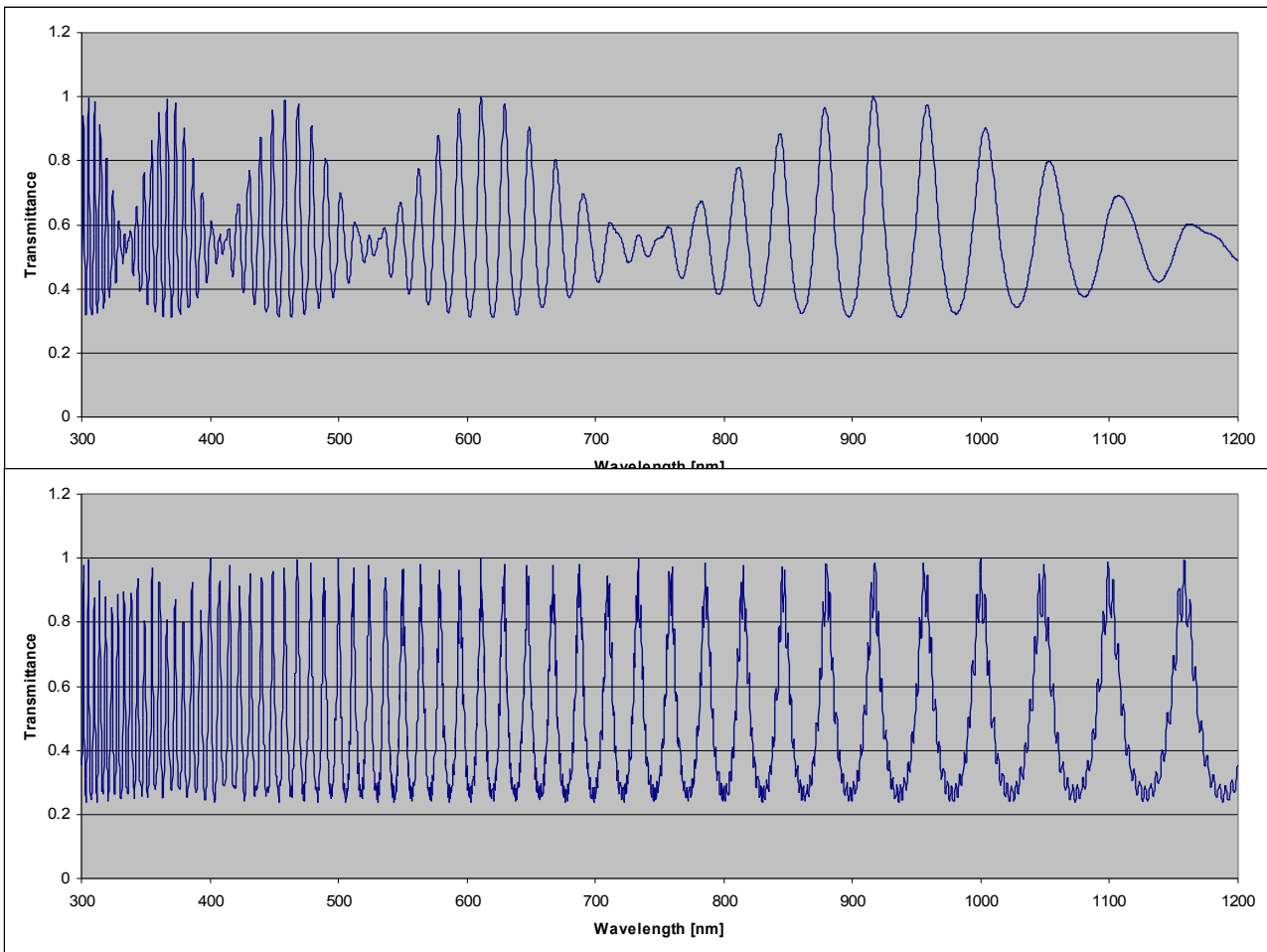


Figure 12: two examples of response plots for series of Fabry-Pérot interferometers.

Real stacks like the one of figure 8 behave similarly to the simple Fabry-Pérot case summarized above.

Putting everything together, one can simulate an approximation of the device response based on semiconductor physics and the theories presented above. A simple simulation gives the example curve of figure 13, real measurement of a photodiode gives the curve of figure 14.

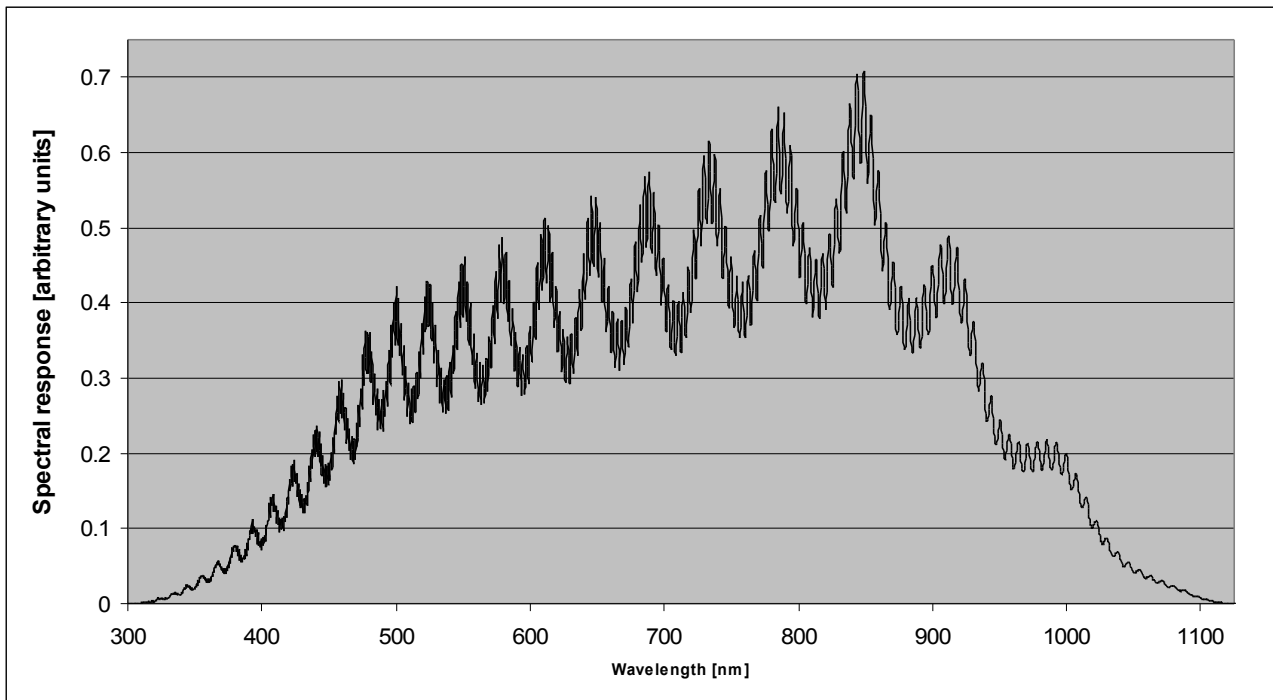


Figure 13: simulated spectral response for an incidence angle of 0 degrees on a silicon photodiode with the top oxide layer and a finish layer.

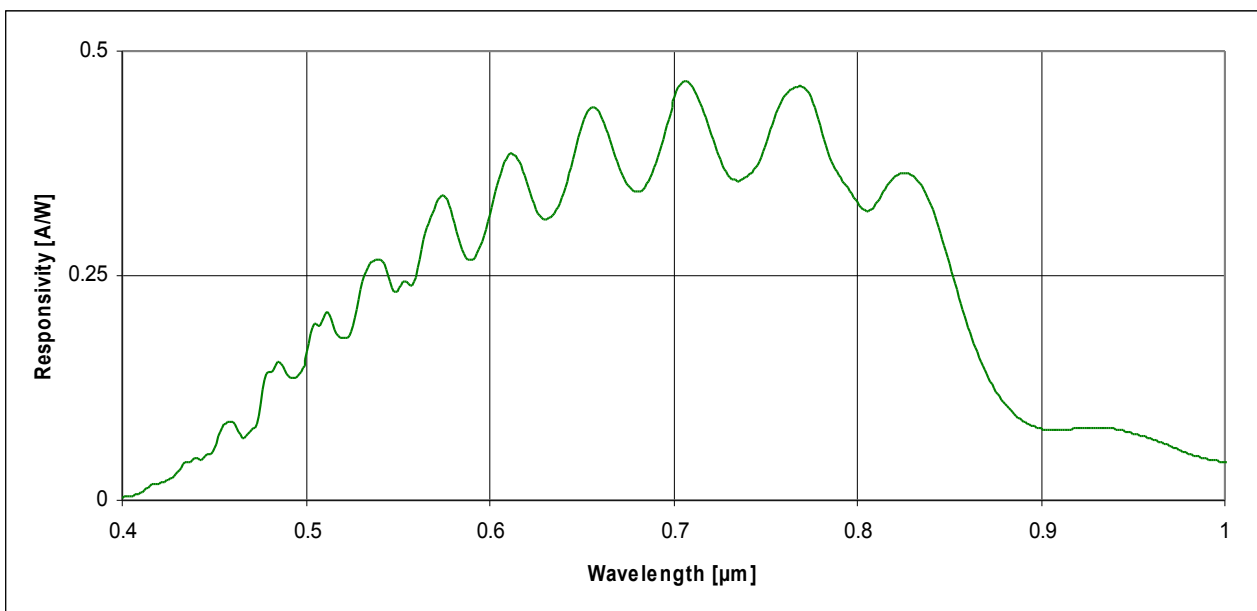


Figure 14: actual measurement of an advanced photodiode

4. Color filters

The most common type of color filters is a patterned layer of organic materials topped by a protective layer. These filters absorb all wavelength except the regions that they are supposed to select.

Usual RGB filters have a passband around the useful frequency and in near infra red, they do not have perfect transmission in the pass band and they don't have perfect rejection in the stop band. They will then affect the amplitude of the response curve at all wavelengths. Additionally, color filters add additional layers to the material stack and these signals may introduce or modify the high frequency ripple of the response curve compared to the monochrome version of the same sensor.

5. Measuring the spectral response

Measuring the response curve is usually made by measuring the output of the sensor for various illumination wavelength. A single wavelength illumination is impossible, there will be in all measurement setups a window of illumination, usually gaussian shaped, with a certain width and a central frequency. The response curve plot is not a plot versus wavelength but a plot versus the central wavelength of the illumination.

The width (full-width at half-height) of the illumination spectrum will depend on the precision of the equipment, usually a monochromator operating on a calibrated wide band light source. If the measurement needs to be precise in wavelength then the power of the light reaching the sensor will be low and the precision of the measurement will be reduced by the presence of any small intensity ambient light and the readout noise of the sensor under test. If the measurement needs to be precise in amplitude then the spectrum of the light source will need to be larger to allow for a higher illumination power and the details of the ripples of the spectral response will disappear. This last effect is depicted in figure 15 where a measurement (blue curve) is compared with the expected sensor response (gray curve). We can see that the high frequency ripples located in the blue and green part of the spectrum have disappeared by averaging.

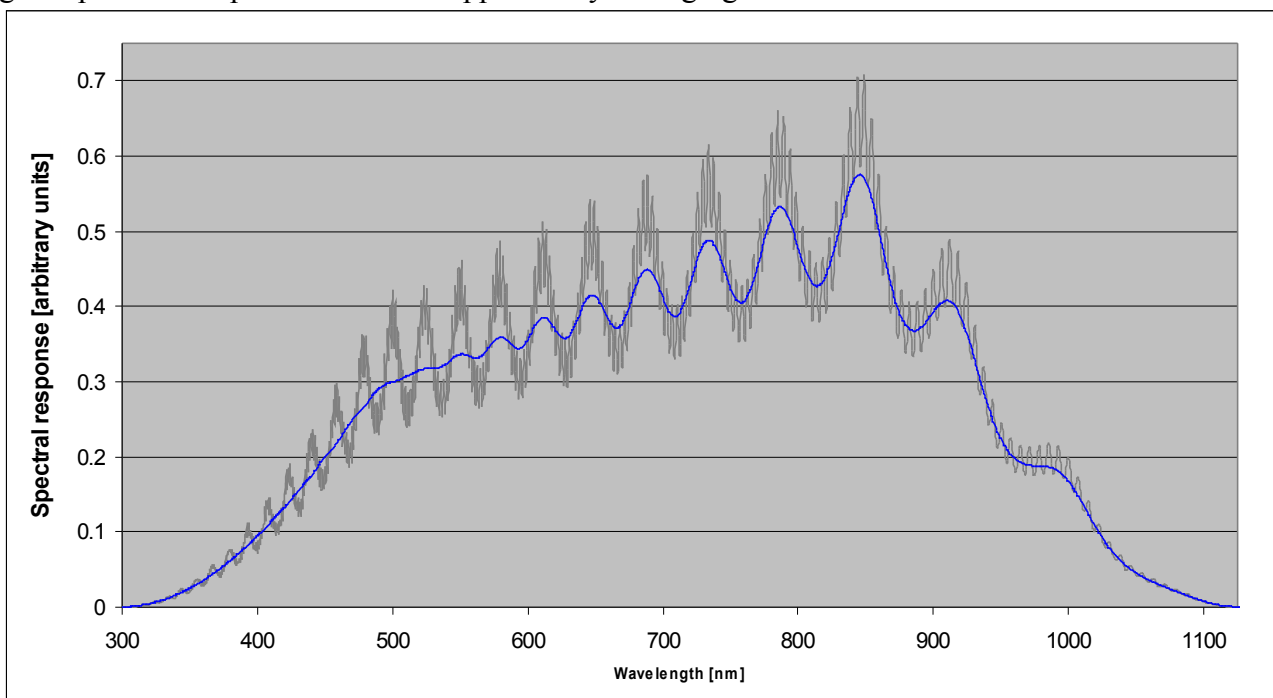


Figure 15: real response curve and the corresponding measurement (simulation)

The selection of the bandwidth of the monochromator is then a trade-off between the required precision in wavelength and the required precision in amplitude of the response. As the ripples are usually of less importance due to their variation from wafer to wafer or their dependency on the chief ray angle, a larger window is usually preferred as the most important parameters to measure are the quantum efficiency in the central part of the spectral response and the limits of this central part (usually 3dB and 10%), as represented on figure 16.

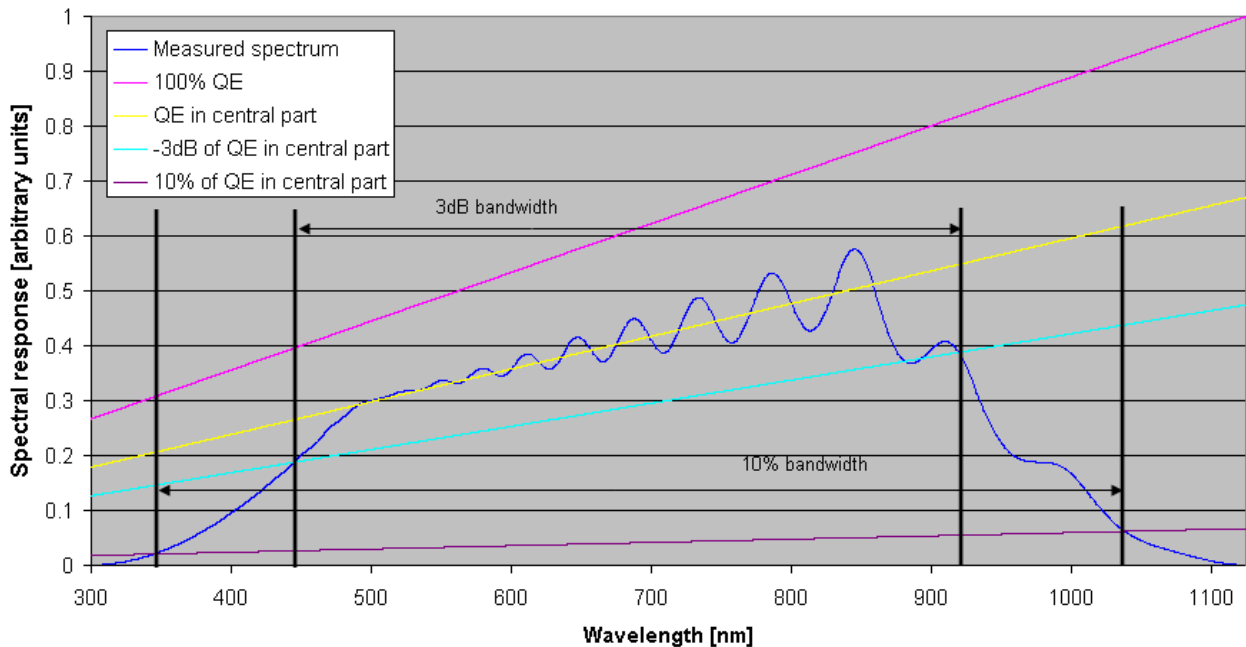


Figure 16: Important measurements of the spectral response

Bibliography

- [1] R. Lange, "3D Time-of-Flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology".*Dissertation Department of Electrical Engineering and Computer Science at University of Siegen*, p54 (2000)
- [2] S.M. Sze, "Semiconductor Devices Physics and Technology" *AT&T Bell Laboratories*, p257, (1985)
- [3] Wikipedia (<http://en.wikipedia.org/wiki/File:Etalon-2.png>)