Universitat de Barcelona

Development and optimization of silicon based light sources for integration into a sensor platform

Joan Juvert Sández

Universitat de Barcelona

Institut de Microelectrònica de Barcelona

Development and optimization of silicon based light sources for integration into a sensor platform

A thesis submitted in partial fulfillment of the requirements for the degree of

PhD in Physics

by

Joan Juvert Sández

Under the research line

Materials electrònics, sensors i micro- i nanosistemes

Director

Carlos Domínguez Horna

Tutor

Blas Garrido Fernández

April 1, 2014

I am the very model
of a scientist Salarian
I've studied species Turian
Asari and Batarian
I'm quite good at genetics
as a subset of biology
because I am an expert
which I know is a tautology
My xenoscience studies
range from urban to agrarian
I am the very model
of a scientist Salarian.

<div align="right">

Mordin Solus performing
Gilbert and Sullivan

</div>

Acknowledgements

I will keep this short because I believe the acknowledgements section of thesis dissertations is overused too often, and it's good to take a stand and make a point every now and then. You will not find my neighbors or the lady from the grocery store listed here.

In all seriousness, I would like to thank Dr. Carlos Domínguez and Dr. Andreu Llobera for giving me the opportunity of working towards my PhD (and with a paycheck at the end of every month no less, which is a sort of luxury these days), as well as to Dr. Blas Garrido for being my tutor at the Universitat de Barcelona.

I would also like to acknowledge the help received from the staff of the clean room whenever it was needed.

The organizers and faculty of the *Advanced Scientific Programming in Python* summer course deserve a spot here too. They work for free to offer a very useful course that I recommend to anyone. And the course is free too. Seriously, this is tier 1 awesomeness.

Special thanks go to my partner in crime, Alfredo González. Frustration and failure divided by two results in only half the original frustration and failure, or so I like to believe. Furthermore, his usually-more-optimist-than-mine views have often been a short (very short) morphine shot to my *realist* views. It's also worth noting his resilience to being carpet-bombed with my nerdrage and *realism*, and his titanic endurance of a desktop environment designed with old ladies in mind.

My parents deserve a spot in this section too, but not because they got me an education (the public school system gives you that for free[1]), but because they strove, and they strove *real hard*, to get me the *best* education they possibly could, which paved the way to where I stand now.

And the most special thanks go to Marta for being so wonderful and for being there for the good and for the bad.

Specially for the good.

But specially for the bad.

No one has endured more of my nerdrage than herself, and that deserves praise.

Last time I checked, thanks and acknowledgements were still tax-free (not sure how long this is going to last), so anyone who thinks he deserves a spot here, feel free to write your name in the margins of this page.

Now let's cut to the chase.

For Tuchanka!

<div align="right">

Joan Juvert

April 1, 2014

</div>

---

[1] Related expenses may apply.

Contents

List of Figures

List of Tables

Introduction

1.1 Motivation and goal

The attention that silicon based photonic materials have drawn in the past two decades has been spurred by the necessity of the microelectronics industry to overcome the bottlenecks in the interconnects between single devices that arise from the extreme integration levels that have been achieved. However, the applications of a silicon based, CMOS compatible light source range from communications to sensing and illumination.

To this day, almost all the photonic devices required for a full silicon based photonic system have been realized successfully, such as optical waveguides, switches and optical modulators, photodetectors and more [1,2]. However, an efficient, CMOS compatible, silicon based light source is still absent. The present work is part of an ongoing effort to achieve a CMOS compatible, silicon based light source that can be integrated monolithically in a sensor platform.

Silicon nanostructures overcome the indirect bandgap of silicon, which is the main limitation as far as the emission efficiency of silicon based light sources is concerned, through quantum confinement [3]. In order to obtain silicon nanoclusters, our group has been using the popular approach of fabricating silicon rich oxide (SRO) or silicon rich nitride (SRN) layers which, after thermal annealing at high temperatures, have been shown to undergo a phase separation that results in pure silicon nanoclusters embedded in the dielectric matrix [4–6]. Other approaches to achieve efficient emission from silicon include porous silicon [7], SiO/SiO_2 superlattices [8], which allow for a very tight control of the size of the silicon nanocrystals, or colloidal silicon quantum dots [9].

As far as the integration goes, no monolithically fabricated, CMOS compatible optical transceiver has been reported yet [10], although integration of light sources based on III-V semiconductors into silicon platforms has been achieved in different ways [10–12].

In our group, the previous work on this subject was primarily focused on the fabrication of the active SRO and SRN layers, the relation between the fabrication

parameters and the structural and photoluminescent properties of those layers, and the demonstration of electroluminescence in full devices using those layers as an active material.

The goal of the present work is twofold. First, after the optical and structural properties of the luminescent material have been studied and the electroluminescence in full devices has been demonstrated, it is necessary to study in detail their optical and electrical properties from the point of view of the device. It is expected that such an analysis will give some clues as to what strategies to follow in order to improve the performance of the devices, namely emitted optical power, efficiency and reliability, as well as to control the emission spectrum.

The second goal of the present work is to propose, design, fabricate and characterize a prototype for an optical transceiver in which all components, namely emitter, waveguide and detector, are monolithically integrated in a CMOS compatible process, and that can be used as a sensing platform.

1.2   Outline of this thesis

The first of the two goals stated in the previous section, that is the characterization of silicon based electroluminescent devices, is covered in chapters 2 and 3.

Chapter 2 describes the fabrication procedure to obtain the active layers and the full devices. The results of the photoluminescent and structural characterization of the active layers are presented in order to relate them to the results of the optical and electrical characterization of the full devices later in the chapter.

During the characterization of the devices, massive differences were found between the photoluminescence of the active layers and the electroluminescence of the devices, which can easily be ascribed, at least qualitatively, to interference effects. In order to provide a more quantitative analysis of this phenomena, a detailed study of the effect of the interference effects in the electroluminescence spectra of the devices is presented in chapter 3.

Chapter 4 covers the second of our goals and describes the design, fabrication and characterization of a prototype for a CMOS compatible, silicon based optical transceiver that can be used as a sensing platform. We will discuss the challenges involved in establishing a design and fabrication process that allows the integration of an emitter, a waveguide and a detector in a single block.

Finally, in chapter 5 we summarize the conclusions extracted from this work.

Fabrication and characterization of silicon based luminescent devices

2.1 Introduction

This chapter will be concerned with the fabrication and characterization of silicon based light sources using CMOS compatible processes. The goal is to study the optical and electrical characteristics of the samples and relate them to the fabrication parameters in order to find an optimal fabrication as far as efficiency, power and reliability go.

The samples that we are going to study consist of an active layer, typically silicon rich oxide (SRO, SiO_x, $x \leq 2$), silicon rich nitride (SRN, Si_3N_x, $x \leq 4$), or both, on top of a silicon substrate. A polysilicon layer acts as the top electrode. We shall divide the following study in two parts. Section 2.2 deals with the fabrication of the samples up to the obtainment of the active layer and its structural and optical characterization. Section 2.3 will be concerned with the remaining steps of the fabrication process until a full device is obtained, and with its optical and electrical characterization.

2.2 Fabrication and characterization of the active layer

This section deals with the fabrication and characterization of the active layer of our devices. In 2.2.1 the steps necessary for the obtainment of the active layer will be discussed in detail. In 2.2.2 the techniques used in the structural and optical characterization of the layers will be presented, and the results of the characterization will be discussed.

2.2.1 Fabrication

The substrates of our samples (figure 2.1(a)) are B doped (p-type), (100) oriented silicon wafers with resistivity between 0.1 and 1.4 Ω cm, which corresponds to a
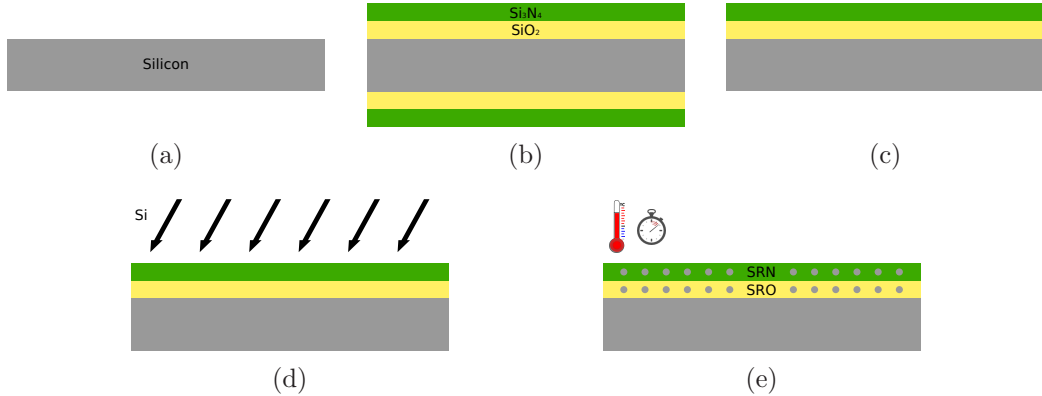
Figure 2.1: Steps for the fabrication of a silicon based luminescent layer.

doping concentration of roughly 1×10^{17} cm^{-3} [13]. If we do not plan on fabricating full devices but instead are only interested in the active layer, then the type and concentration of the dopants is irrelevant.

After RCA cleaning [14] of the wafers, the active layer can be grown or deposited. For ion implanted SRO, a thermal SiO$_2$ layer is grown at 1000 °C in an oxygen atmosphere [15,16]. It is also possible to deposit a SRO by plasma enhanced chemical vapor deposition (PECVD) [17] or low pressure chemical vapor deposition (LPCVD) [17]. Alternatively, a Si$_3$N$_4$ layer can be deposited by LPCVD. This process can be repeated to obtain a multilayered system, for instance an LPCVD Si$_3$N$_4$ layer on top of a thermally grown SiO$_2$ layer, as illustrated in figure 2.1(b).

The PECVD process deposits the layer on one side of the wafer only, because the samples are placed horizontally on the reactor. Thermal oxidation and LPCVD process of SRO or Si$_3$N$_4$, on the other hand, will result in an oxide or nitride layer on both sides of the wafer. Because of the low pressure, LPCVD hot wall reactors allow many wafers to be coated simultaneously in a stand up configuration without detrimental effects to film uniformity. Therefore, for LPCVD and thermal growth, the layers on the back side should be eliminated if we plan on biasing the devices using the silicon substrate as a contact. To do that, the front side is spin coated [18] with a photoresist for protection and the layers on the backside are removed by plasma etching [19] or wet etching [20]. Finally, the photoresist is removed from the front side by plasma etching (figure 2.1(c)).

For stoichiometric SiO$_2$ and Si$_3$N$_4$ layers, the silicon excess is introduced by silicon ion implantation [21] (figure 2.1(d)). Implantation energy and doses depend on the thickness of the layers and the target silicon excess. For typical thicknesses of a few tens of nanometers, the implantation energies are in the range 25–50 KeV, whereas doses are of the order of 10^{16}–10^{17} cm^{-2}, as determined in previous works [22, 23]. Implantation profiles follow a Gaussian distribution. For SRO layers, by combining two implantations with different energies and doses across a thin Si$_3$N$_4$ buffer layer, it is possible to obtain reasonably flat implantation profiles, as found

by simulation with SRIM [24] or Sentaurus Process [25]. After the implantation process, the Si_3N_4 buffer layer can be removed by wet etching [20] or, alternatively, left as part of the active medium. Note that the implantation process will produce a transition region of oxynitride between the Si_3N_4 and SiO_2 layer, as we will see later, at the end of section 2.2.2.3.

The silicon excess will be quantitatively defined as the atomic concentration of silicon not bound to form SiO_2 or Si_3N_4, that is, free to nucleate into silicon nanoclusters. Mathematically:

$$XS_{Si} = [Si] - \frac{1}{2}[O] - \frac{3}{4}[N] \tag{2.1}$$

Here, [Si], [O] and [N] are the atomic concentrations of silicon, oxygen and nitrogen, respectively, that is, the ratio of the number of Si, O and N atoms to the total number of atoms.

After the silicon rich layer (SRO, SRN or both) is in place, a high temperature thermal annealing in N_2 atmosphere induces the formation of silicon nanoclusters in the silicon rich dielectric layer [4–6], figure 2.1(e). Typical temperatures range from 900 ºC to 1250 ºC and times from 1 to 4 hours, depending on the fabrication technique. The selected temperature and time for each technique is based on the best photoluminescence observed in previous works [22, 23].

Table 2.1 lists the annealing temperatures and times for the samples discussed in this chapter. In table 2.1 and all other tables in this chapter, each fabrication is labeled so that the first letter, in capitals, stands for the fabrication technique: I for *ion implantation*, P for *PECVD* and L for *LPCVD*. The following characters, in lowercase, stand for the kind of EL: p for *points*, h for *homogeneous* and pl for the sets that were intended exclusively for photoluminescence measurements. The different kinds of EL will be studied later, in section 2.3.2.1.

This concludes the fabrication of the active layer. In section 2.3.1 the fabrication process will be resumed in order to describe the steps required to achieve a full device from this point in the process.

2.2.2 Structural and optical characterization

In this section a few useful techniques for structural characterization will be reviewed, as well as their application to our samples. Ellipsometry and reflectometry are used to measure the thickness and refractive index of thin layers. X-Ray Photoelectron Spectroscopy provides useful information about the chemical composition of the samples.

2.2.2.1 Ellipsometry and reflectometry

In this work two different techniques have been used to measure the thickness and refractive indices of thin layers, namely ellipsometry and reflectometry. In the paragraphs that follow we will briefly review both techniques.

Fab. id.	Active layer	Temperature (°C)	Time (min)
Ip	SRN/SRO	1100	240
Ip	SRO	1100	240
Ph	SRO	1250	60
Ih	SRO	1100	240
Iph	SRO/SRN	1100	240
Iph	SRO	1100	240
Ipl	SRO	1100	240
Ipl	SRN	1100	60
Lph	SRO	1100	180

Table 2.1: List of the temperatures and annealing times used in the fabrication of the samples discussed in this chapter. For the samples Ipl, a different annealing time was used for the wafers with SRO and those with SRN active layer. See the text for an explanation of the labels used to identify each fabrication.
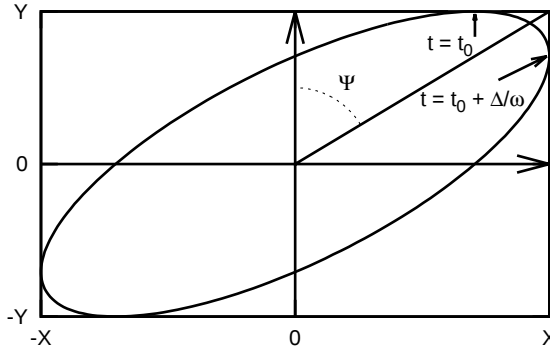
Figure 2.2: Polarization ellipse. The parameters Ψ and Δ completely characterize the state of polarization of the wave.

Ellipsometry In a coordinate system xyz any plane wave traveling along the z axis can be decomposed into two electric field components:

$$E_x(z, t) = X \, e^{i\Delta_x} \, e^{ik_z z} \, e^{-i\omega(t-t_0)} \tag{2.2}$$

$$E_y(z, t) = Y \, e^{i\Delta_y} \, e^{ik_z z} \, e^{-i\omega(t-t_0)} \tag{2.3}$$

The oscillatory parts in the spatial and temporal variables are common to both components, but the field amplitudes X and Y and the global phase of each component, Δ_x and Δ_y, are in general different. For a fixed z, it can be shown that the the vector $\vec{E} = (E_x(t), E_y(t))$ describes an ellipse such as the one depicted in figure 2.2.

In figure 2.2, $\Delta = \Delta_x - \Delta_y$ is the phase difference between the x and y components of the field. For $\Delta = 0$ the ellipse collapses into the diagonal line at an angle Ψ

measured from the y axis, and this corresponds to linear polarization. For increasing values of Δ the ellipticity increases and the major axes of the ellipse rotate until they coincide with the xy axes for $\Delta = \pi/2$

The angle Ψ is defined as $\tan \Psi = X/Y$ and thus expresses the ratio between the amplitudes in the x and y directions.

Therefore, the parameters Ψ and Δ completely characterize the polarization of the wave. In fact, it is clear that the ratio of the complex amplitudes in x and y can straightforwardly be expressed in terms of the ellipsometric angles:

$$\frac{X \, e^{i\Delta_x}}{Y \, e^{i\Delta_y}} = \frac{X}{Y} \, e^{i(\Delta_x - \Delta_y)} = \tan \Psi_{xy} \, e^{i\Delta_{xy}} \tag{2.4}$$

Note that this description of the polarization depends on the chosen reference system. Now, if we have a light beam incident on a surface, using the Fresnel equations [26] for the parallel (p) and perpendicular (s) components of the light, we can write:

$$\tan \Psi_r \, e^{i\Delta_r} = \frac{r_p \, P \, e^{i\Delta_p}}{r_s \, S \, e^{i\Delta_s}} = \frac{-r_p}{r_s} \, \tan \Psi_i \, e^{i\Delta_i} \tag{2.5}$$

where $P \, e^{i\Delta_p}$ and $S \, e^{i\Delta_s}$ are the parallel and perpendicular components of the incident field, respectively, r_p and r_s are the reflection coefficients for the p and s components, respectively, and the r and i subscripts in Ψ and Δ stand for the *reflected* and *incident* waves. Rearranging the last equation:

$$\frac{r_p}{r_s} = \frac{\tan \Psi_r}{\tan \Psi_i} \, e^{i(\Delta_r - \Delta_i)} = \tan \Psi \, e^{i\Delta} \tag{2.6}$$

Therefore, measurement of the change in the polarization state of a wave upon reflection on a surface will provide information regarding the optical properties of the reflecting material through the ratio r_p/r_s.

Several techniques exist to measure the ellipsometric parameters, the most common one being *null ellipsometry*. It consists in preparing an incident beam with a polarization state such that it will become linearly polarized after being reflected by the sample. A linear polarizer followed by a compensator (an optical element that introduces a phase difference between its fast and slow axes) is used to prepare the state of polarization of the incident beam, whereas another linear polarizer (usually called *analyzer*) is used to check whether the reflected light is linearly polarized or not. This configuration is known as the PCSA configuration (Polarizer-Compensator-Sample-Analyzer), and is illustrated in figure 2.3.

Once the ellipsometric angles are known the extraction of the optical parameters from r_p/r_s can be more or less complicated depending on the characteristics of the sample being analyzed.

Detailed information on ellipsometry can be found in any of the numerous publications on the subject, such as [27] and [28].

The ellipsometric measurements in this work have been carried out with a Rudolph Research Autoel IV ellipsometer, which operates at wavelength 632.8 nm and an incident angle of 70 deg. All the measurements are carried out in five points of each wafer.
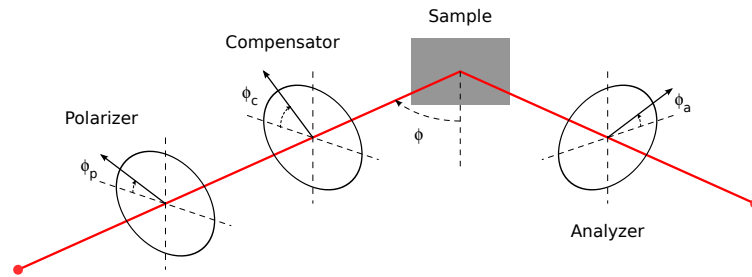
Figure 2.3: PCSA configuration usually found in null ellipsometers.
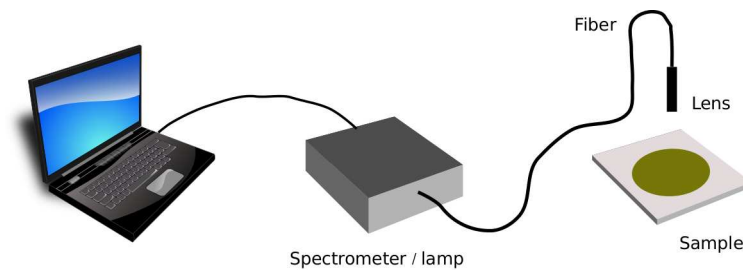
Figure 2.4: A typical reflectometry setup.

Reflectometry   Reflectometry consists in measuring the reflectance R of a system as a function of wavelength. Since $R(\lambda)$ can be calculated for any multilayer structure, for instance using the characteristic matrix formalism (see for instance reference [26] or section 3.2.1.9 in chapter 3 of the present work), the calculated reflectance can be compared to the measured one. The calculation of R is done in terms of parameters such as the refractive index and thickness of the layers. An iterative process can tweak those parameters in order to find the best fit to the measured reflectance.

Figure 2.4 depicts a typical reflectometry setup. An optical fiber coupled to a lens is used to focus the light from a lamp perpendicularly onto the sample, and at the same time collect the reflection from the surface of the sample and bring it to a spectrometer. Under the sample, a mirror deflects the incident beam so as to not collect light reflected on the platform where the sample rests.

As with any other method based on fitting experimental data with a model, it is important to fix the known parameters in order to reduce the parameter space as much as possible, otherwise one can get completely meaningless results.

The reflectometry measurements in this work have been carried out with a Filmetrics F20, which allows the simultaneous measurement of thicknesses between 15 nm and 50 μm and the refractive index in the range 400–1000 nm.

Results   Since the thickness of the layers has an important impact in the electrical and optical behavior of our devices (see section 2.3.3 in the present chapter and
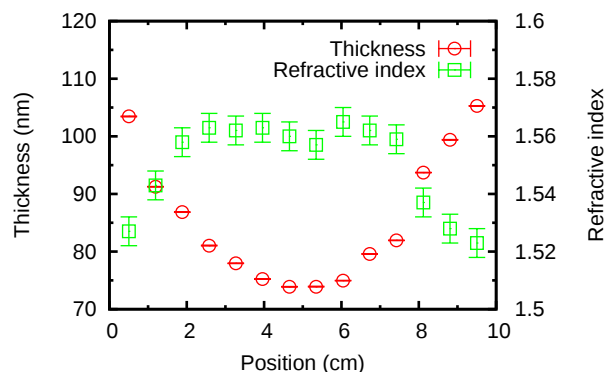
Figure 2.5: Typical non-uniformity in the refractive index and thickness observed in our PECVD samples. The measurements have been taken along a diameter of the wafer. The positions are approximate.

3.4.3.4 in chapter 3), its precise measurement is an important task. Significant non-uniformities or lack of control of the thickness of the active layer or the electrode can lead to low repeatability of the results across different devices of the same wafer. Identifying those non-uniformities can explain apparently odd results in the characterization of a wafer.

Figure 2.5 shows the typical thickness and refractive index profile of a PECVD SRO layer measured along a diameter of the wafer. It is clear that the reactor, in this example an Oxford Instruments Plasmalab80 Plus, deposits faster close to the edges of the wafer and that the composition of the film is richer in oxygen in that area.

Table 2.2 lists the thickness and refractive index of the samples used in this work. The measurements are usually carried out by ellipsometry and reflectometry in order to compare the results for consistency, or lack thereof. Each wafer is probed at five different locations. The values quoted in table 2.2 correspond to the mean value of the five measurements, and the error to the standard deviation. Therefore, each error is given with the typical 68% confidence limit [29].

2.2.2.2 X-ray photoelectron spectroscopy

X-ray photoelectron spectroscopy (XPS) is a technique which provides both quantitative and qualitative information about the atomic and molecular composition of the surface of a sample.

The technique is based on the photoelectric effect. The surface of the sample is irradiated by a highly monochromatic photon source. The photons are absorbed by the electrons of the surface atoms and, if the photon energy is high enough, some electrons can overcome their binding potential and be extracted from the sample with a remainder of kinetic energy, i.e:

$$h\nu = E_K + E_B \tag{2.7}$$

Sample	Thickness (nm)	Refr. index
Fabrication Ip		
1	27 ± 2	1.60 ± 0.01
2	24 ± 4	1.74 ± 0.01
3	22 ± 4	1.83 ± 0.01
4	29 ± 4	1.55 ± 0.03
5	24 ± 4	1.62 ± 0.04
6	21 ± 5	1.5 ± 0.2
Fabrication Ph		
2	59 ± 7	1.81 ± 0.01
3	60 ± 9	1.68 ± 0.01
5	55 ± 6	1.57 ± 0.01
6	53 ± 5	1.53 ± 0.01
Fabrication Ih		
2	45 ± 5	NA
3	48 ± 4	"
4	49 ± 9	"
6	57 ± 1	"
Fabrication Iph		
1	NA	NA
2	NA	NA
3	NA	NA
5	28.6 ± 0.3	1.48 ± 0.01
6	17 ± 1	1.47 ± 0.2
7	26.7 ± 0.6	1.56 ± 0.01
Fabrication IPl		
3	38.8 ± 0.1	1.46 ± 0.01
4	41.7 ± 0.2	1.50 ± 0.01
6	40.4 ± 0.6	2.05 ± 0.02
7	41.3 ± 0.6	2.21 ± 0.02

Table 2.2: Thickness and refractive index, at 632.8 nm, of the samples studied in this chapter. The errors correspond to the standard deviation over the measurement on five different points in each wafer. For samples Iph 1, 2 and 3, we cannot measure the thickness of the SRO reliably because it is buried under the nitride layer and because of the oxynitride layer formed between the oxide and nitride during implantation.
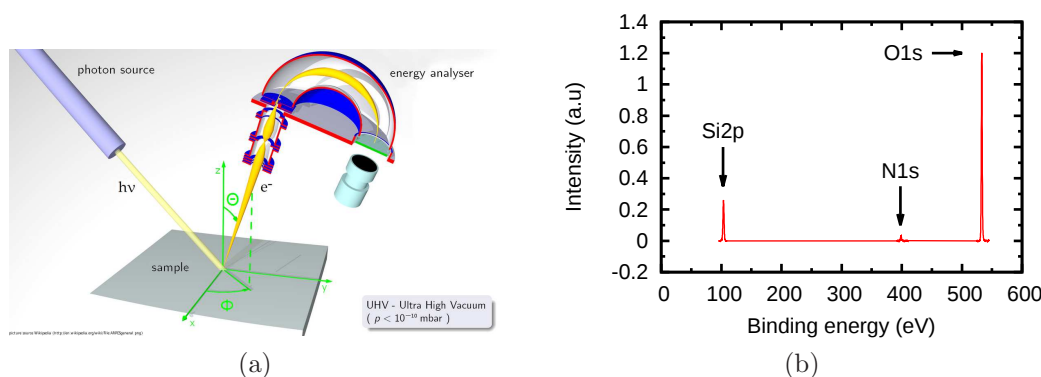
Figure 2.6: (a) Schematic of an XPS setup. (b) An XPS spectrum of an SRO sample. The peaks correspond to the 2p orbital of silicon and the 1s orbital of oxygen and nitrogen.

where $h\nu$ is the energy of the incident photon, E_B is the binding energy of the electron and E_K is the kinetic energy of the same electron after being extracted from the sample. Given the typical binding energies of the electrons in an atom, X-ray photons from AlKα (1486.6 eV), MgKα (1253.6 eV) or TiKα (2040.0 eV) are used in XPS.

For a given energy of the incoming photons, electrons with a different binding energy will be detected with different kinetic energies. Analyzing the energy spectrum of the extracted electrons with an electron spectrometer gives a series of peaks at certain energies which correspond to the different electronic states in the atoms composing the sample. The spectrum obtained for a given atom essentially constitutes a fingerprint of the atom.

Furthermore, the intensity of the peaks in the spectrum is related to the concentration of the elements in the sample, and the position of the peaks depends on the environment of the atom, for example the oxidation state. Therefore, the analysis of the spectrum gives very detailed information regarding the chemical composition of the sample.

This technique can be applied to solids and liquids and is non destructive. However, in order to obtain depth profiles of the composition of the sample, it is usually combined with etching of the surface. The spectral analysis of the ejected electrons requires ultra high vacuum.

Figure 2.6(a)[1] shows a typical XPS setup. Figure 2.6(b) shows an XPS spectrum of a typical SRO sample, with peaks corresponding to the 2p orbital of silicon, the 1s orbital of oxygen, and the 1s orbital of nitrogen.

More details regarding this powerful technique can be found in [30], among others.

---

[1]Image source: http://en.wikipedia.org/wiki/File:ARPESgeneral.png

Sample	Imp. dose $(10^{16}$ cm$^{-2})$	Nominal thick. (nm)	Si. excess (at. %)
Fabrication Ip			
4	1.2	30	1.3 ± 0.3
5	1.5	"	4.6 ± 0.3
6	3.0	"	4.0 ± 0.6
Fabrication Ph			
2	NA	50	16.0 ± 0.6
3	"	"	12.0 ± 0.2
5	"	"	8.0 ± 0.6
6	"	"	6.0 ± 0.2
Fabrication Ih			
2	6.2	50	12.0 ± 0.3
3	7.4	"	13.0 ± 0.4
4	9.0	"	14.0 ± 0.2
6	10.3	"	15.0 ± 0.5
Fabrication Iph			
5	1.2	30	0.46 ± 0.02
6	1.5	30	0 ± 0.09
7	3.0	30	1.3 ± 0.5
Fabrication Lph			
1	NA	40	5.1 ± 0.4
2	NA	80	4.0 ± 0.4
3	NA	40	5.1 ± 0.4
4	NA	80	4.0 ± 0.4

Table 2.3: Values of silicon excess extracted from XPS measurements.

Table 2.3 lists the silicon excesses of the samples that have been measured by XPS. The XPS measurements of samples Ip and Iph were carried out at the Servicios de Análisis y Caracterización de Sólidos y Superficies of the Universidad de Extremadura with a Thermo Scientific K-Alpha XPS, and the silicon excesses were extracted from the raw XPS data by A. González. For the older samples Ph, Ih and Lph, the XPS measurements were carried out in the Centres Científics i Tecnològics of the Universitat de Barcelona.

For ion implanted samples, it is possible to calculate an estimation of the silicon excess from the implantation profile and dose. In the case of SRO, for instance, the density of Si atoms and O atoms in the original SiO_2 matrix (before the implantation), can be easily calculated from the density and molar mass of SiO_2 as follows:

$$[X] = \frac{\rho}{m} \times N \times n \tag{2.8}$$

where [X] stands for the atomic density of each species in the molecule, in this case Si or O, in atoms per cubic centimeter; ρ is the the density of the material, in
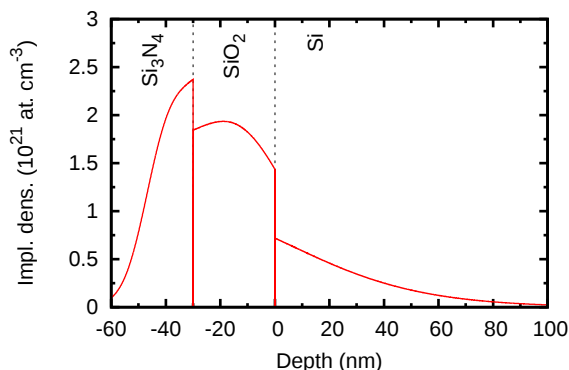
Figure 2.7: Implantation profile of one of our double implantations through a Si_3N_4 layer. In this example, the SiO_2 layer receives only 45% of the implanted ions.

this case SiO_2, in grams per cubic centimeter; m is the molar mass in grams per mol; N is the Avogadro number and n is the stoichiometry of the particular species, in this case 1 for Si and 2 for O.

Furthermore, the atomic density of the implanted Si can be calculated in several ways. A first approximation is to divide the implantation dose by the thickness of the SiO_2 layer, paying attention to the units. However, this calculation assumes that all the implanted atoms end in the SiO_2 layer and that they are uniformly distributed. While the latter is probably an acceptable first approximation to get an average silicon excess for the whole layer, the former is not, since a significant part of the implanted ions will end up past the SiO_2, in the substrate, or, in the case of our typical double implantation through a nitride layer, in the nitride. To illustrate this, figure 2.7 shows the Sentaurus Process simulation of a double implantation (25 KeV, 0.96×10^{16} cm^{-2} and 50 KeV, 0.24×10^{16} cm^{-2}) into a 30 nm SiO_2 layer through a 30 nm Si_3N_4 layer. It is clear that the SiO_2 layer receives only a fraction of the total implantation, 45% in this case.

Therefore, we can extract the mean value of the silicon implanted density in the SiO_2 layer directly from the simulation, or we can calculate the fraction of the total dose in the SiO_2, 0.45 in the case of figure 2.7, and use it as a weight in the calculation of the mean density:

$$[Si]_{Impl} = \frac{\alpha\, d}{t} \qquad (2.9)$$

where $[Si]_{Impl}$ stands for the atomic density of the implanted Si; α is the calculated fraction of the dose in the SiO_2 layer; d is the total implanted dose in atoms per cubic centimeter and t is the thickness.

The fraction α is trivially calculated after numerical integration of the implantation profile between the relevant boundaries.

Once we know the atomic densities of the implanted Si and the Si and O of the original SiO_2, the silicon excess can be calculated according to our previous
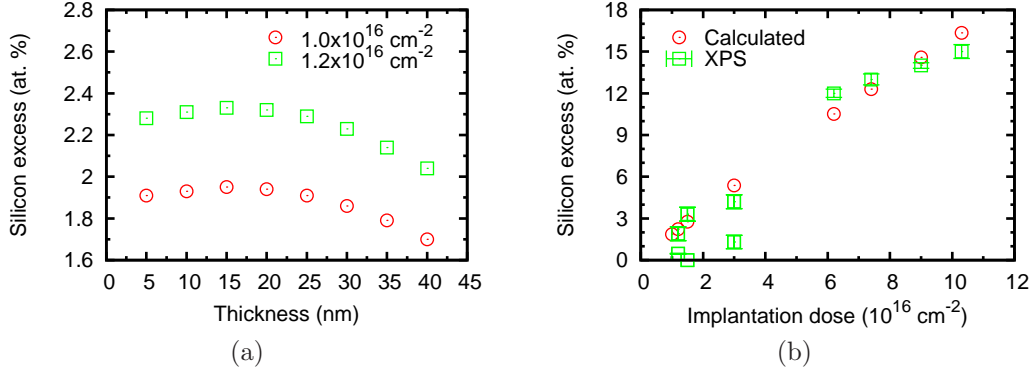
Figure 2.8: Calculated silicon excess for our typical double implantation process through a 30 nm Si_3N_4 layer, as a function of: (a) the SiO_2 thickness, and (b) the implantation dose with the SiO_2 thickness fixed at 30 nm. In (b) the experimental silicon excesses from XPS analysis are also plotted for comparison.

definition, equation 2.1, as:

$$XS_{Si} = \frac{[Si]_{Impl}}{[Si]_{Impl} + [Si]_{SiO_2} + [O]_{SiO_2}} \tag{2.10}$$

In equation 2.10 we are calculating what fraction of the *total* amount of atoms in the SRO layer corresponds to the implanted Si ions.

In figure 2.8(a) we plot the calculated silicon excess as a function of the thickness of the SiO_2 layer of figure 2.7. For our typical range of thicknesses, the change in the silicon excess is not very significant for an estimation of the silicon excess.

Figure 2.8(b) plots the silicon excess as a function of the implantation dose for the same double implantation process, with the thickness of the SiO_2 fixed at 30 nm, and the silicon excesses of samples Ip, Ih and Iph as extracted from XPS measurements. Clearly, the values of silicon excess calculated with our simple model can be used to get an estimation of the silicon excess for our typical double implantation process through a 30 nm nitride layer, and it should be valid for thicknesses of the SRO around 30 nm per our remarks about figure 2.8(a).

2.2.2.3   Optical characterization: photoluminescence

Photoluminescence (PL) consists in the emission of photons from the relaxation of electrons that have been previously photoexcited, i.e light emission from the recombination of electron-hole pairs generated by photon absorption. The study of the PL spectra of the sample provides insight into the emission mechanisms taking place in the sample.

An experimental setup for the measurement of the PL was built from scratch for this work. A diagram of the setup is shown in figure 2.9 and a picture in figure
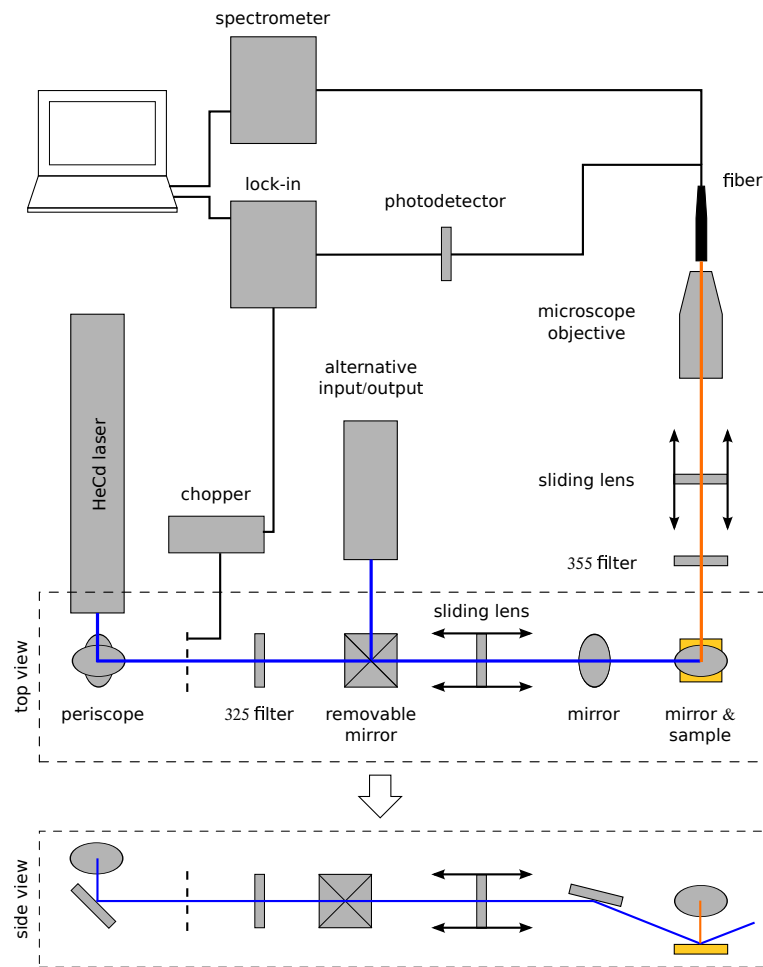
Figure 2.9: Schematic of the PL setup.

2.10. The optical elements of the system (mirrors, filters and lenses) are all aligned in a system of rails that makes it easy to add and remove elements to the setup, as well as change their order, without modifying the existing alignment.

The optical pumping is provided by an HeCd laser from Vm-Tim. Since the output port of the laser is not at the same height as the system of rails holding the optical elements, a set of two mirrors is used to lower the beam and align it to the axis of the system (see *periscope* in the *top view* and *side view* in figure 2.9). Next in line, an optical chopper (Stanford Research Systems SR540) pulses the incoming beam at a certain frequency in order to improve the signal to noise ratio through the use of a lock-in amplifier (Stanford Research Systems SR830 DSP). After the chopper, a laser line clean-up filter selects the 325 nm line of the HeCd laser, discarding all other components of the incoming beam.

Next, a cubic holder allows the easy insertion of a mirror or beam splitter that,
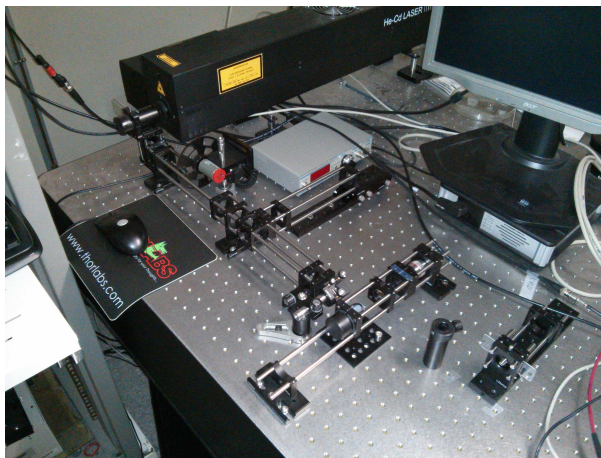
Figure 2.10: The PL setup used in this work.

depending on its orientation, can be used to divert the HeCd laser to a second line or, alternatively, to use a different light source as input.

A lens is then used to focus the beam. The lens is mounted on a holder that can slide along the rails in order to adjust the focal point of the beam. Next, a UV mirror deflects the beam downwards to strike the sample, which rests horizontally on the breadboard, at an angle of approximately 45 deg. (see the *side view* in figure 2.9). The specular reflection of the laser on the surface of the sample is blocked in order to avoid damage to unaware passers-by.

The emission from the sample in the direction normal to the surface is deflected into the axis of the output rails by a mirror positioned above the sample (see the *mirror & sample* in figure 2.9).

The output system consists of a long pass filter that blocks all wavelengths below 355 nm, in particular the scattered emission from the 325 nm line of the pumping laser. Next, a lens collimates the emission into a microscope objective, which in turn couples the light into an optical fiber that brings the emission to whichever device we use to analyze the light.

In order to study the spectra of the emission, the optical fiber is connected to a spectrometer (Ocean Optics QE65000), upper branch in figure 2.9. Alternatively, a monochromator (Newport CS260) and a photodetector can be used for spectral analysis. Although the monochromator is potentially more accurate than the spectrometer, given the low intensity of the PL of our samples, it was necessary to work with both the entrance and exit slits at their maximum aperture, which resulted in a resolution equivalent, or even worse, than that of the spectrometer. In these circumstances, the ability of the spectrometer to capture all wavelengths simultaneously was more convenient.

In order to measure the intensity of the PL without regard to its spectral composition, two different photodetectors were used.

For measurements of the absolute emitted power in Watts, a calibrated silicon photodetector (Newport 918D-UV-OD3 with Newport 1931-C control unit) was used. Alternatively, a photomultiplier tube (PMT, Hamamatsu H9656) connected to a lock-in amplifier was used for measurements that require high signal to noise measurements, for instance measuring small changes in the PL intensity with time.

The lock-in amplifier works as follows: the driving signal, in this case the HeCd laser, is switched on and off with a known frequency (modulation by a square wave). In our setup this modulation is carried out by a mechanical chopper. The frequency of the chopper is controlled by the user and fed into the lock-in. Since the driving photon flux is switched on and off, the measured signal also exhibits a periodic behavior at the same frequency as the driving frequency. The lock-in amplifier filters the measured signal by removing all its Fourier components at frequencies other than the driving frequency. That means that all contributions to the signal that are not due to actual emission of the sample (i.e, noise) are discarded. A good example of that is the characteristic signal at 50 Hz due to the lighting of the laboratory. The output signal of the lock-in is a voltage proportional to the coefficient of the term corresponding to the driving frequency in the Fourier expansion of the input signal, which in turn is proportional to its intensity.

One caveat when using the lock-in is that one must be careful when comparing measurements of samples that respond differently to the shape of the modulated pumping used in the measurements. For instance, consider two samples that show the same emission intensity when pumped by a continuous, constant photon flux. If the first sample shows a very fast response to the optical excitation, then its emission will follow closely the pumping square wave. If the second sample shows a slow response to the optical excitation, then its emission will not follow the pumping square wave, even if the intensity after the transient is the same as that of the first sample. The signals of these two samples will share the same base frequency, but the value of the coefficients of their Fourier expansion, in particular the coefficient of the term at the driving frequency, will be different. Therefore, the output voltage of the lock-in will be slightly different even when the intensity of both samples in the steady state is exactly the same.

For the spectra measurements, it is important to note that all optical elements used in the collection of the light emitted by the sample modify the spectrum in some way because their response typically depends on the wavelength. For example, the reflectance of the mirrors depends on the wavelength, and the optical fibers exhibit a different attenuation at different wavelengths. Furthermore, the detectors themselves have different sensitivities at different wavelengths. The *responsivity* of a photodetector is defined as the ratio between the output current of the detector and the input optical power, and is given in Amperes per Watt:

$$r(\lambda) = \frac{i}{P} \quad [\mathrm{A\,W^{-1}}] \tag{2.11}$$

This can be generalized for the whole detection system, not only the photodetector, by the adimensional ratio between the measured intensity and the real intensity,
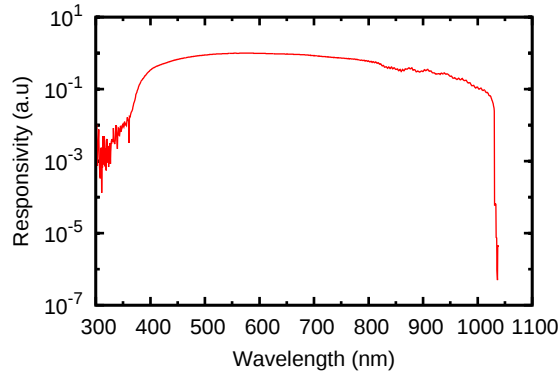
Figure 2.11: Responsivity of the PL setup depicted in figure 2.9 for the spectrometer branch.

both in arbitrary units, taking into account all the elements of the collection system. In order to determine the responsivity of the entire detection system one just needs to measure a light source with a known spectrum and divide the measured spectrum by the real one. Of course, if the actual radiance of the source is known then the responsivity can be measured with the correct units instead of as an adimensional ratio. Once the responsivity $r(\lambda)$ is known, all the measured spectra can be corrected by dividing the recorded intensity value at each wavelength λ by the corresponding value of the responsivity $r(\lambda)$. The responsivity of the spectrometer branch of the collector system depicted in figure 2.9 is plotted in figure 2.11.

As already mentioned, measurements of PL intensity can be carried out with a calibrated photodetector that will display the measured intensity in Watts, rather than a voltage that is proportional to the actual intensity. In order to do that, the photodetector control unit measures the photocurrent generated by the incoming optical signal and then divides it by the responsivity at the wavelength of the incoming light, which is set manually by the user before starting the measurements. The table of responsivities as a function of the wavelength has been loaded by the manufacturer after proper calibration of the detector.

This works well if the measured light is reasonably monochromatic, but is clearly wrong when it is not. To compensate for this, one must record the generated photocurrent and then divide by an *effective responsivity* that factors in the spectrum of the measured light.

This effective responsivity can be calculated as follows: consider the spectrum $s(\lambda)$, in $W\,nm^{-1}$, of the light being measured. The spectrum can clearly be rewritten as $Ps_n(\lambda)$, where P, in W, is the area under the curve of $s(\lambda)$, and $s_n(\lambda)$, in nm^{-1}, is the spectrum $s(\lambda)$ normalized to area 1. We can measure $s(\lambda)$ with a spectrometer and P is precisely the unknown power we want to find. It is clear that each wavelength of the incoming beam contributes $i(\lambda)\,d\lambda = r(\lambda)Ps_n(\lambda)\,d\lambda$ to the total photocurrent in the detector, with $i(\lambda)$ in $A\,nm^{-1}$. The total photocurrent can

be found by integration of $i(\lambda)$, i.e:

$$I = P \int_0^\infty r(\lambda) s_n(\lambda) \, \mathrm{d}\lambda \tag{2.12}$$

It is clear that the integral in equation 2.12 can be interpreted as an effective responsivity:

$$r_e = \int_0^\infty r(\lambda) s_n(\lambda) \, \mathrm{d}\lambda \tag{2.13}$$

Since I is measured by the photodetector, $r(\lambda)$ is provided by the manufacturer and $s_n(\lambda)$ can be measured with a spectrometer, we can calculate the effective responsivity r_e and find the optical power incident on the photodetector as:

$$P = I r_e^{-1} \tag{2.14}$$

Before finishing this section, it is worth noting that the PL setup presented here can be easily tweaked in order to measure reflectance or transmittance. It is also worth commenting on possible ways in which the setup could be expanded to allow for a more complete set of characterizations.

A high power lamp and a monochromator can be used for PL excitation measurements (PLE). In this kind of measurement, the wavelength of the pumping light is varied while the PL is monitored. In order to get the most out of the technique, the temperature of the system is usually lowered with a cryogenic station.

The addition of a photon counter or a highly sensitive cooled photodetector coupled to an amplifier and an oscilloscope could be used to measure time resolved PL, which would yield information about excitation cross sections and typical lifetime of the radiative centers in the sample.

Results In order to investigate the PL of SRO and SRN layers, several samples were fabricated. Table 2.4 lists their fabrication characteristics.

Fabrication Ipl consists in two wafers with an LPCVD Si_3N_4 layer, and an additional two with a thermal SiO_2 layer, all of them with a nominal thickness of 40 nm (real thickness between 38 and 42 nm according to ellipsometry measurements, see table 2.2). For each material, one of the samples was implanted with Si ions with a dose of $1.0 \times 10^{16} \, \mathrm{cm}^{-2}$ and the other with a dose of $3.0 \times 10^{16} \, \mathrm{cm}^{-2}$. The results of the PL measurements are shown in figure 2.12.

It is clear from figure 2.12 that increasing the silicon content produces a redshift in the PL peak. This redshift of the peak with silicon content has been observed consistently in accordance with the literature [6,31,32]. For another example of this behavior, see figure 2.13, corresponding to fabrication Ip, where the redshift of the peak with increasing implantation dose is very clear. PL measurements of PECVD samples can be found in [33].

However, what we want to note here is that the same silicon excess produces two different bands in the nitride and oxide layers, the nitride band appearing in the high energy end of the visible spectrum whereas the oxide band in the lower

Sample	Material	Dose $(10^{16} \text{ cm}^{-2})$	Si exc. (at. %)	Energy (KeV)	Nominal thick. (nm)
		Fabrication IPl			
3	SRO	1	2.7	15	40
4	SRO	3	7.8	15	"
6	SRN	1	2.5	20	"
7	SRN	3	7.2	20	"
		Fabrication Ip			
4	SRO	1.2	1.3 ± 0.3	25/50	30
5	"	1.5	4.6 ± 0.3	"	"
6	"	3.0	4.0 ± 0.6	"	"
		Fabrication Iph			
1	SRO/SRN	1.2	0.46 ± 0.02	25/50	30/30
3	SRO/SRN	3.0	1.3 ± 0.5	"	30/30
5	SRO	1.2	0.46 ± 0.02	"	30
7	SRO	3.0	1.3 ± 0.5	"	30

Table 2.4: Fabrication parameters of the samples used for PL measurements. In all of them the excess silicon is achieved by implantation of silicon ions. For the samples that underwent a double implantation process, the energies of the first and second implantations are given separated by a slash, and the given dose corresponds to the total implantation. In the bilayers, the SRN is on top of the SRO. The values of silicon excess of the samples IPl have been calculated using the procedure outlined in section 2.2.2.2
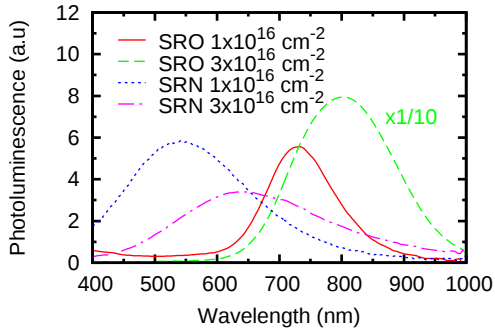
Figure 2.12: PL of SRO and SRN with different implantation doses (samples IPl). The curve of the SRO layer with dose $3 \times 10^{16} \text{ cm}^{-2}$ has been scaled by 1/10 to fit in with the others.
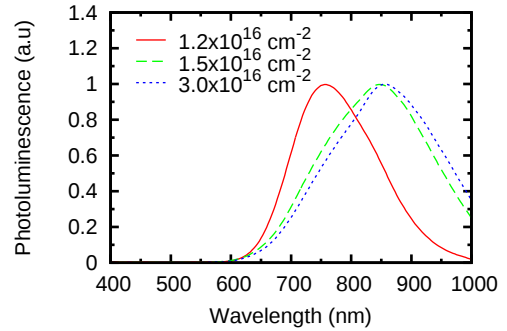
Figure 2.13: Normalized PL spectra of implanted SRO samples Ip 4, 5 and 6. Note the redshift with increasing implantation dose.

(a) Dose 1.2×10^{16} cm^{-2}, samples Iph-5 and Iph-1

(b) Dose 3×10^{16} cm^{-2}, samples Iph-7 and Iph-3.
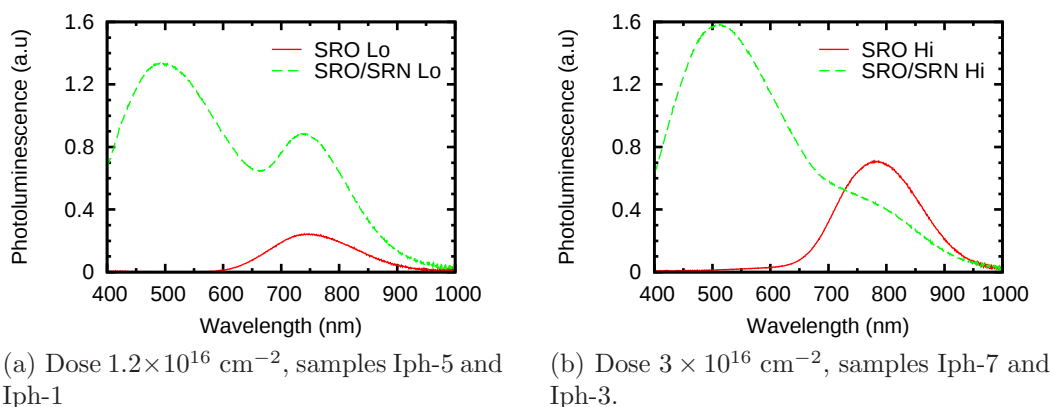
Figure 2.14: PL of a single implanted SRO layer and that of a SRO/SRN bilayer with different implantation doses.

energy end. Even if we compare the nitride layer with high implantation dose to the oxide layer with low implantation dose, their bands are still well separated by at least 100 nm. That will be important in justifying the identification of each of the two bands observed in SRN/SRO bilayer systems.

Fabrication Iph consisted in sets of two ion implanted samples, one with an implanted oxide layer only, and another with an implanted oxide/nitride bilayer. The implantation dose was different in each set (see table 2.4). Figures 2.14(a) and 2.14(b) show the PL of the sample with only the SRO layer compared to that with the SRO/SRN bilayer, for low (Lo) and high (Hi) implantation doses. In both cases, the addition of the nitride layer adds a new band to the emission of the single SRO layer.

XPS evidence, see figure 2.15, seems to suggest that, even though the implantation process, which is targeted to the middle of the oxide layer, leaves an implantation tail in the silicon nitride according to the simulations (see figure 2.7), the amount of silicon excess in the nitride layer is actually very low. That indicates that the emission due to the nitride is probably located at the interface between the oxide and the nitride, where the material is rather an oxynitride.

Note that the situation was different in the nitride samples shown in figure 2.12. In that case, the implantation in the samples with a nitride layer was of course targeted at the nitride. Nevertheless, observation of the results presented in figures 2.12 to 2.14(b) indicates that the emission due to the nitride/oxynitride and the oxide are clearly discernible regardless of the implantation dose.

Section 3.5.2 in chapter 3 discusses in detail the modification of the emission of bilayer systems due to optical interference. Our ability to decompose the total emission in two bands that can be ascribed to emission in the nitride and the oxide will be of utmost importance.
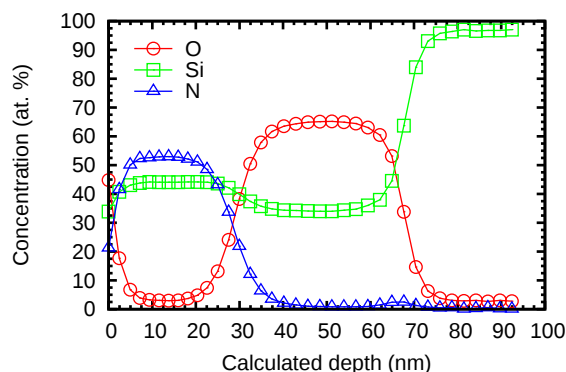
Figure 2.15: XPS profile of a sample with the same bilayer as sample Iph-5. The silicon excess in the nitride is not significant and there is a clear oxynitride region between the nitride and the SRO.

2.3   Fabrication and characterization of the devices

This section will be concerned with the fabrication and characterization of finished devices. In 2.3.1 we will pick up the fabrication process where we left off in section 2.2.1. In sections 2.3.2 and 2.3.3 we will discuss the optical and electrical characterization of the finished devices, respectively.

2.3.1   Fabrication

In addition to the steps described in section 2.2.1, further processing of the samples is required in order to obtain MIS-like (metal–insulator–semiconductor) devices that can be electrically biased.

Since the back side of the wafer will be used as a contact, a boron ion implantation with dose of $5 \times 10^{14}\,\mathrm{cm^{-2}}$ at 80 KeV, followed by a thermal annealing at 800 °C during 30 minutes for dopant activation, is carried out on the back side in order to achieve an ohmic contact [34, 35] (figure 2.16(a)).

The next step consists in the LPCVD of a 350 nm thick polysilicon layer. Since the LPCVD process will deposit on both sides of the wafer, it is convenient to deposit a thin PECVD SiO_2 layer on the back of the wafer previous to the deposition of the polysilicon, so that the latter will be easy to remove from the back because it will be resting on SiO_2 rather than on the silicon substrate (figure 2.16(b)).

After the polysilicon has been deposited, it is doped until degeneracy with phosphorus by diffusion of $POCl_3$ [36]. Next, the phosphosilicate glass generated during the diffusion is eliminated by wet etching (figure 2.16(c)).

Photolitography [37] is used to define an array of polysilicon gates with different geometries and sizes on the front side of the wafer. The polysilicon is etched from the front and the back by reactive ion etching [19] and the photoresist used as an etching mask is removed by plasma etching (figure 2.16(d)).
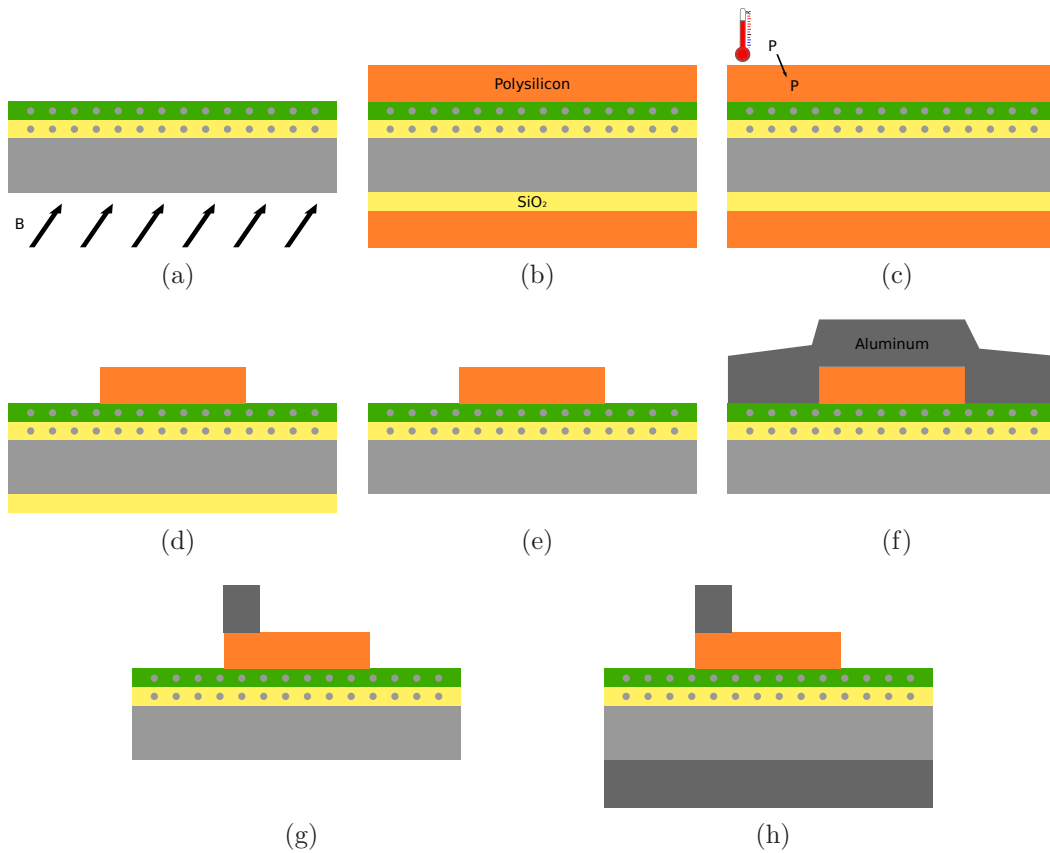
Figure 2.16: Steps for the fabrication of a silicon based luminescent device. This is a continuation of the steps described in section 2.2.1.

In order to eliminate the PECVD SiO_2 layer from the back of the wafer, a protective resist is deposited on the front side, the SiO_2 is removed from the back by plasma etching and again the photoresist is eliminated from the front side by plasma etching (figure 2.16(e)).

Once the polysilicon gates are in place, the aluminum contacts can be deposited. A 1 μm thick layer of aluminum is sputtered [38] onto the wafer followed by a photolitographic process (figure 2.16(f)). The aluminum is etched by wet etching to define the contacts (figure 2.16(g)). A similar aluminum layer is then sputtered onto the back of the wafer and finally the samples are annealed in forming gas (N_2 + 5% H_2) at 350 ºC in order to sinter both the front and back aluminum contacts (figure 2.16(h)).

Figure 2.17 shows the layout of a complete chip in samples Ip and Iph. Red corresponds to the polysilicon level and blue to the aluminum level. The smallest square capacitor, in the upper left corner of the chip, is 0.5 mm in size on its side. Each larger capacitor increases the length of its size in 0.5 mm respect to the
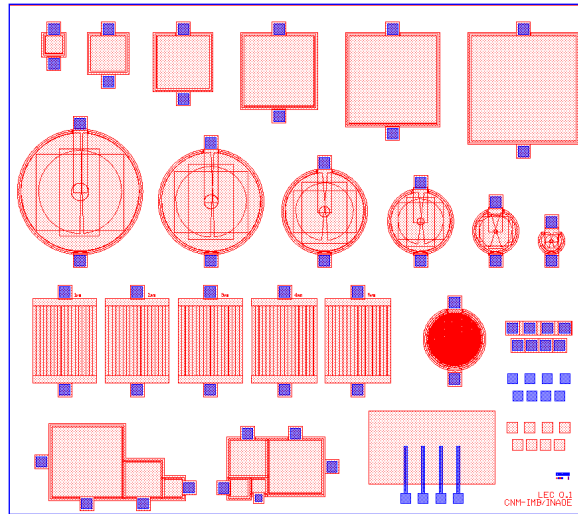
Figure 2.17: Layout of the mask used in our devices. Red corresponds to the polysilicon level and blue to the aluminum level.

previous one, up to 3 mm for the largest capacitor. For each square capacitor there is a circular capacitor with the same area. There are other structures such as Van der Pauw and interdigitated structures, and a structure (lower left) that replicates the layout of the chips in the mask set that was used for the fabrication of the older samples Ph, Ih and Lph.

At this point the devices are fully fabricated and can be biased between the top aluminum contact and the substrate.

2.3.2 Optical characterization: electroluminescence

Electroluminescence (EL) consists in the emission of photons due to the flow of electrical current through a system. For example, injection of an electrical current through the active material can result in the excitation and relaxation of electrons in the material due to impact with the injected electrons, or electrons injected through one end of the material may recombine with holes injected through the other end.

An experimental setup for the measurement of the EL was built from scratch for this work. A diagram of the setup is shown in figure 2.18 and a picture in 2.19. The wafer is placed in an aluminum chuck (built in-house) and held in place through vacuum. The chuck itself rests on a platform that can move in any direction through the action of three step-motors (Micos VT-80), one for each direction in a typical XYZ configuration. One or two tungsten probes, the position of which can be tightly controlled too, are used to bias the samples. A Keithley 2430 source-meter is used to establish a voltage difference between the probe and the chuck, or possibly between
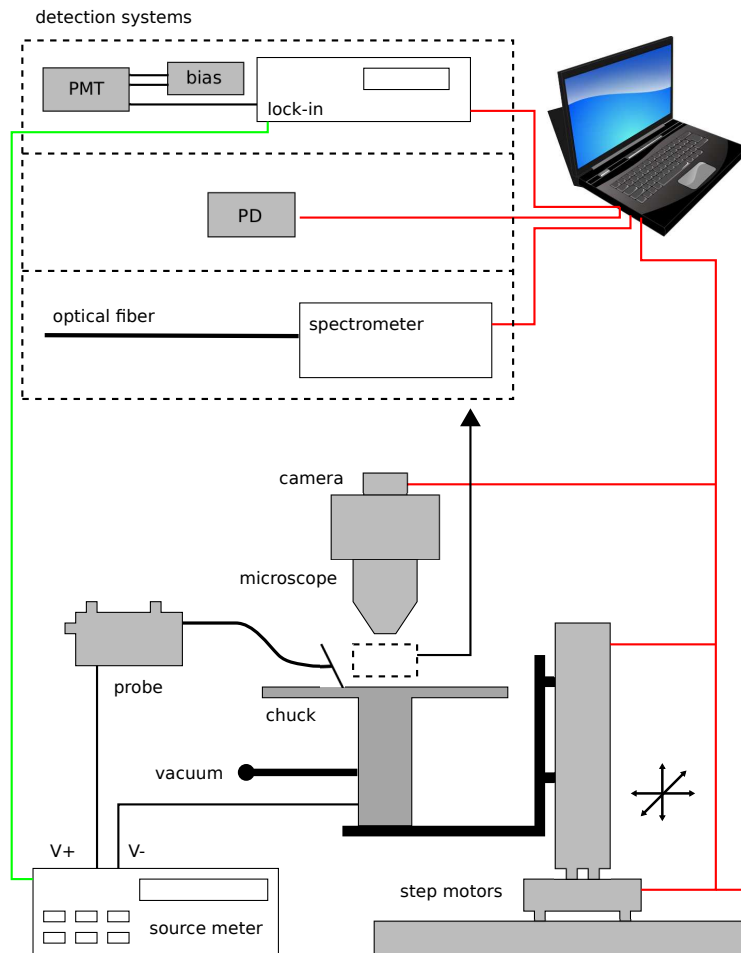
Figure 2.18: Schematic of the EL measurement setup.

two probes, and measure the current at the same time. It can also be used to fix the current and measure the voltage, and it admits a simple pulsed operation.

Above the wafer, a microscope is used to check the placement of the probe in the sample as well as to examine it. Optionally, a camera can be attached to the microscope in order to capture images of the sample.

In order to capture the light emission from the sample, three different detection systems have been used.

For spectrum measurements, an optical fiber is placed on top of the sample with a probe similar to the one used for biasing. The fiber collects the light emitted by the sample and brings it to a spectrometer (Ocean Optics QE65000) for analysis.

For intensity measurements there are two options. A calibrated photodetector (Newport 918D-UV-OD3 with Newport 1931-C control unit) can be used to measure absolute power after the corrections discussed in section 2.2.2.3. For measurements in arbitrary units, a more sensitive photomultiplier tube (PMT, Hamamatsu H9656)
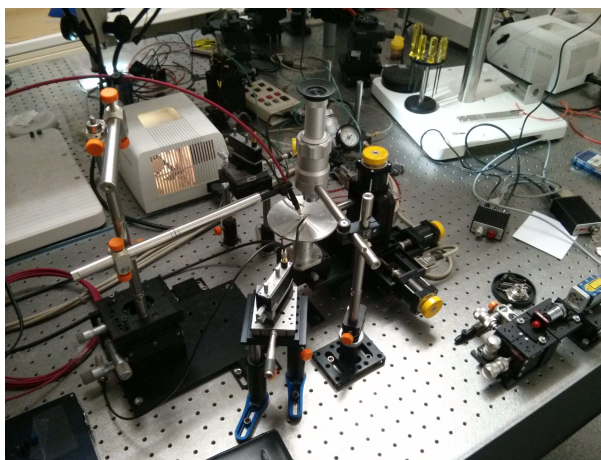
Figure 2.19: The EL setup used in this work.

can be used. The PMT is biased by a voltage source and, in order to improve the signal to noise ratio, its output is connected to a lock-in amplifier. The reference signal is received from the trigger output of the source-meter, which should be operated in pulsed mode.

The step-motors, source-meter and detection system of choice are all connected to a computer that controls all of them by means of a custom built LabView program. The program allows to automatically perform current-voltage (or voltage-current) sweeps while at the same time measuring the EL intensity or spectrum. That allows for the correlation of the EL signal to the electrical behavior of the sample. The measurements can be carried out automatically on an array of devices in a wafer without any intervention of the experimenter.

The corrections to the measurements discussed in section 2.2.2.3 also apply here. In addition, since the photodetector only collects the light emitted by the sample in a limited solid angle, we have to take that into account in case we want to find the total power emitted by the sample. In order to find the total power from the measurement, we must know, at least approximately, the emission profile of the sample. It is reasonable to expect the profile to follow closely Lambert's cosine law, and indeed our samples have shown such an emission profile [39].

According to Lambert's cosine law, which expresses the amount of power emitted per unit solid angle and unit area of the emitter, the power density as a function of the observation angle is:

$$I(\theta, \phi) = I_{\max} \cos(\theta) \quad \left[\mathrm{W\,sr^{-1}\,m^{-2}}\right] \tag{2.15}$$

Where θ lies in the range $[0, \pi/2]$ and is the azimuthal angle measured from the direction normal to the emitter, and ϕ is the angle of revolution and lies in the range $[0, 2\pi]$. Note that although equation 2.15 does not depend on ϕ (the emission is symmetric across a whole revolution) a dependence on ϕ has been indicated to make it clear that the emission takes place in volume, not just in a plane.
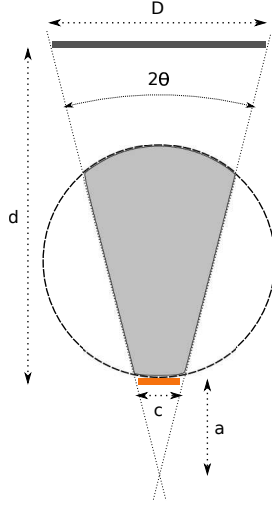
Figure 2.20: Schematic view of the limited solid angle subtended by the detector. The circle represents the emission profile according to Lambert's cosine law, and its shaded area corresponds to the fraction of that emission that is actually collected by the detector. The remaining emission is lost.

The power measured by the detector is the integral of the emission across the solid angle subtended by the detector (see figure 2.20) multiplied by the area of the emitter, i.e:

$$P_{\mathrm{m}} = c^2 \int_{\Omega} I_{\max} \cos(\theta) \, \mathrm{d}\Omega = c^2 I_{\max} \int_0^{2\pi} \mathrm{d}\phi \int_0^{\Theta} \cos(\theta) \sin(\theta) \, \mathrm{d}\theta =$$
$$c^2 I_{\max} \frac{\pi}{2} [1 - \cos(2\Theta)] \quad (2.16)$$

Where c^2 is the area of the emitter and Θ is the semi-angle (plane angle) subtended by the detector, as depicted in figure 2.20. We consider the typical size c of the emitter to be much smaller than the typical size D of the detector so that the integral along θ can be carried out from 0 to the same Θ regardless of the specific point in the area of the device.

Note that in equation 2.16 we know P_{m}, which is the magnitude measured by the detector, and also c and Θ, since they can be measured directly. In fact, Θ can be expressed in terms of the sizes c and D of the emitter and detector, respectively, as well as the distance d that separates them, as:

$$\Theta = \arctan\left[\frac{c}{2d}\left(\frac{D}{c} - 1\right)\right] \quad (2.17)$$

where $d \cdot (D/c - 1)^{-1} = a$ (see figure 2.20). Therefore, $I_{\max}$ can be calculated as:

$$I_{\max} = \frac{2P_{\mathrm{m}}}{c^2 \pi [1 - \cos(2\Theta)]} \quad (2.18)$$

27

<table>
<tr><td>(a) 10 mA, 41.3 V</td><td>(b) 200 mA, 56.8 V</td></tr>
</table>
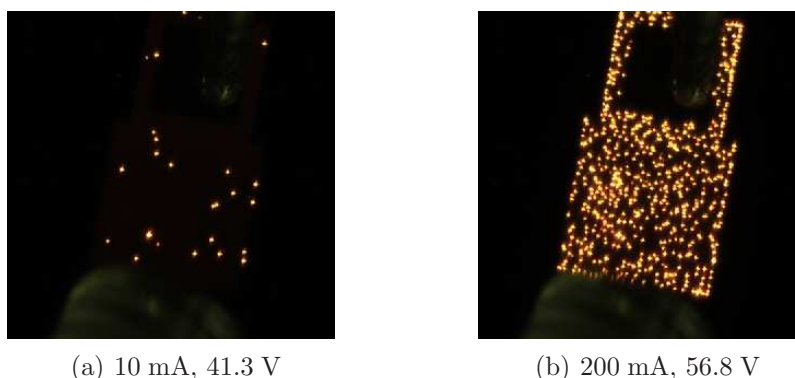
Figure 2.21: Emission through points observed at different currents in a device from sample Ip-3.

The total emitted power is the integral of Lambert's law across the whole hemisphere, which we can now express in terms of known quantities:

$$P_{\mathrm{t}} = c^2 \pi I_{\max} = \frac{2P_{\mathrm{m}}}{1 - \cos(2\Theta)} \tag{2.19}$$

That means that the total power P_{t} can be found by multiplying the measured power P_{m} by a factor $2/c^2\pi[1 - \cos(2\Theta)]$

2.3.2.1   Results

One of the first results that were obtained from the EL characterization is that the luminescence comes in three different ways. One of them is through multiple discrete luminescent points scattered across the area of the device, as shown in figure 2.21. Another way, much less common, is through emission along the edge of the device, as shown in figure 2.22. Finally, the emission can be observed to be homogeneously distributed in the active area, as illustrated in figure 2.23(a). In some cases, a coexistence between different kinds of emission is observed, like the point observed in an otherwise edge emission in figure 2.22 or the superposition of emission points to an underlying homogeneous emission as shown in figure 2.23(b). In fact, the superposition of points and uniform emission is also present in figure 2.21, although the homogeneous component is much fainter. It is entirely possible that there is always a homogeneous component in the emission but it is too faint to be detected in most cases where points are present. We shall now comment on each of these types of emission.

The emission through points comes with high injection currents, from a few mA to several hundreds of mA at voltages of several tens of volts for the typical thicknesses of our samples, typically between 20 and 60 nm. Moreover, it is associated with a capacitance-voltage curve that does not conform to the typical behavior of a MOS system, as we will see later in the chapter. That points to a device operating

(a) 100 mA, 30.3 V

(b) 400 mA, 37.2 V
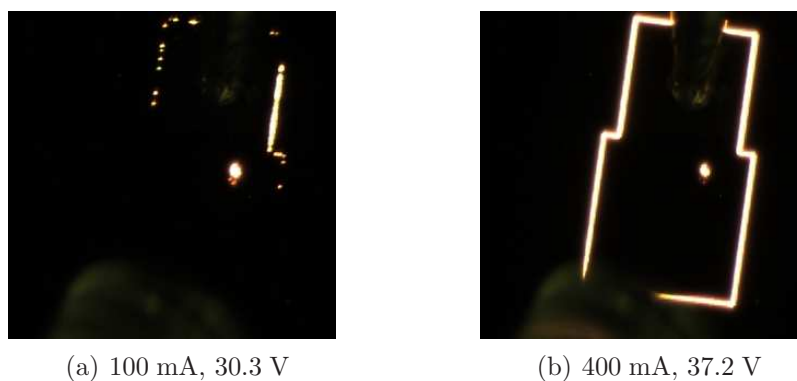
Figure 2.22: Emission along the edge of the device at different currents in a device from sample Ip-6.

(a) 100 μA, 53.0 V, Iph-3

(b) 60 mA, 42 V, Iph-7
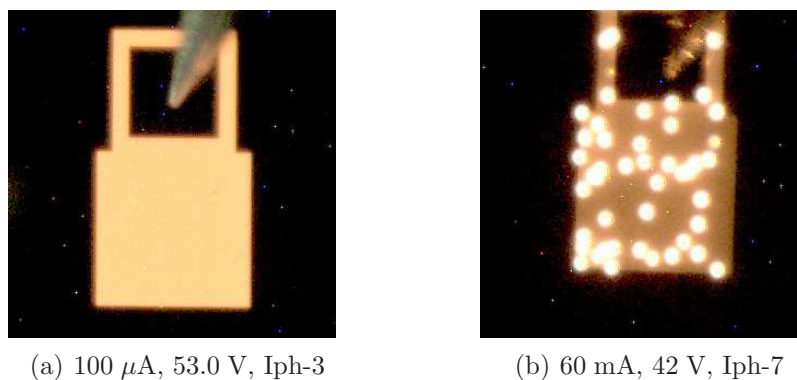
Figure 2.23: Homogeneous and hybrid emission in different samples of fabrication Iph.

under breakdown conditions, either hard or soft breakdown [40], where a luminescent point appears on the regions where a low resistivity path has formed between the top and bottom electrodes. The higher the voltage, the more breakdown points appear in the area, as illustrated in figures 2.21(a) and 2.21(b).

This behavior has been observed in SRO samples fabricated with different techniques (PECVD, LPCVD and ion implantation), with different silicon contents and variation in the vertical layout of the samples (for instance adding a thin nitride layer between the substrate and the SRO). We have not been able to restore the breakdown paths consistently by reversing the bias, as reported by other authors [41]. This is probably due to the fact that we are always working with SRO layers where phase separation has occurred during annealing whereas, apparently, the kind of resistive switching observed in [41] occurs in samples where no phase separation has taken place.

Whereas the emission through points has been observed in multiple samples, the

edge emission has only been observed in a couple of samples. For that reason, we believe that this is the result of some anomaly in the fabrication, such as a reduction in thickness near the edges of the active area, that favors conduction in that region.

The edge emission takes place at similar currents and voltages as the emission through points. Note that, as the voltage is increased, the distance from the probe at which luminescence is observed increases. For example, in figure 2.22(a), corresponding to 100 mA, only the rim of the contact pad is partially illuminated near the probe. It is not until the current hits the 400 mA mark that the flow seems to reach far enough so as to illuminate the rim of the full device, as depicted in figure 2.22(b).

Finally, figure 2.23(a) shows the homogeneous emission of an ion implanted sample corresponding to a device in sample Iph-3. In the samples with homogeneous emission, the current needed to obtain the luminescence is much lower than in the samples with emission through points at similar voltages, with hundreds of μA being a typical figure, i.e three orders of magnitude less. The efficiency of the different emission types will be discussed later in the section.

In order to discuss in more detail the properties of the different kinds of emission, a set of representative samples will be studied. In the following paragraphs the results of the characterization of each fabrication will be discussed. After that, the results will be compared and some conclusions will be extracted.

Fabrication Ip (points/edge emission)   It consists of two groups of three samples. A 10 nm Si_3N_4 layer and a 40 nm stoichiometric PECVD SiO_2 are deposited on top of the silicon substrate of the first group of samples, whereas a 30 nm thermal SiO_2 is grown on the second group of samples. A 30 nm Si_3N_4 layer is then deposited on all six samples. Two implantations of silicon ions are then carried out through the Si_3N_4 layer in order to achieve a flat profile. The first implantation is carried out at 25 KeV and the second at 50 KeV. The individual doses are adjusted so as to obtain three different total doses in each group. The samples are then annealed at 1100 °C during 240 minutes. The Si_3N_4 layer on top of the SRO is then removed and the polysilicon gate is deposited. Table 2.5 lists the fabrication parameters of the samples. See sections 2.2.1 and 2.3.1 for more details regarding the fabrication process.

Figure 2.24 shows the normalized spectra for all samples of the fabrication Ip. Except for the sample with the highest implantation dose of the group without the nitride layer (cyan short-dashed line), all other curves are remarkably similar. The peaks observed at around 600, 700 and 900 nm are due to interference effects, as will be shown in chapter 3. Other than small differences in the position and relative intensity of the peaks, which could be attributed to small differences in the layer structure or the expected redshift of the intrinsic emission with increasing implantation dose, from the device perspective the spectrum is largely the same in all samples.

The outlier of the group (cyan short-dashed line) in figure 2.24 corresponds to

	Fabrication Ip					
Sample	Material	Dose $(10^{16}$ cm$^{-2})$	Si exc. (at. %)	Energy (KeV)	Nom. thick. (nm)	Lum.
1	SRN/SRO	1.2	1.3 ± 0.3	25/50	10/30	P
2	"	1.5	4.6 ± 0.3	"	"	P
3	"	3.0	4.0 ± 0.6	"	"	P
4	SRO	1.2	1.3 ± 0.3	"	30	P
5	"	1.5	4.6 ± 0.3	"	"	E/P
6	"	3.0	4.0 ± 0.6	"	"	E

Table 2.5: Fabrication parameters of the samples Ip. In the bilayer systems, the SRN is below the SRO. In all the samples the SRO is obtained through ion implantation of silicon. However, in samples 1, 2 and 3 the base oxide is obtained from a stoichiometric PECVD, whereas in samples 4, 5, and 6 the oxide is thermally grown. In the luminescence column, P stands for *points* and E for *edge*.
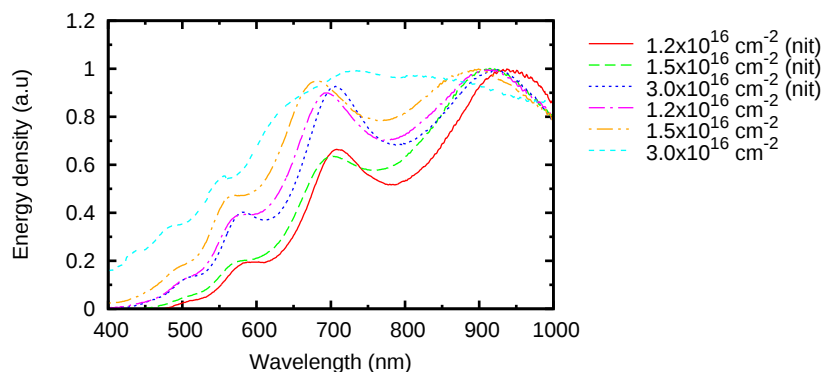
Figure 2.24: Normalized EL spectra of the samples of fabrication Ip. The sample with edge emission (cyan short-dashed line) presents a significantly smoother spectrum.

the sample that shows edge emission. Its spectrum follows the same trend as the other samples but in this case the interference peaks are less noticeable. Since the emission is taking place at the edge of the active area, it seems reasonable to expect a lesser effect of the interference phenomena, since a significant part of the captured light has not even crossed the polysilicon gate.

Figure 2.25 shows the typical optical power and power conversion efficiencies of the samples Ip as a function of the current density. The optical power increases with the driving current in a slightly sublinear relation and reaches values of several tens of nW. As for the efficiency, it exhibits a clear maximum value at a certain current and it tends to zero rapidly for smaller currents. For higher currents it decreases slowly. At low currents the error in the efficiency measurement is large. That is because the power conversion efficiency is calculated as the quotient between the
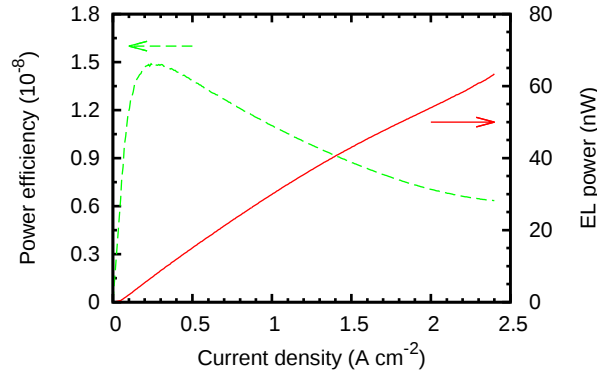
Figure 2.25: Typical EL power and power efficiency as a function of the current density for the samples of fabrication Ip, exhibiting emission through points and/or edge emission.

Figure 2.26: (a) Maximum power efficiency and (b) current density at the maximum efficiency, both as a function of the implantation dose for the samples Ip.

optical power and the electrical power. The relative error in the quotient is the sum of the relative errors in the divisor and the dividend. As we approach low currents, and thus low emission, the relative error in the measurement of the optical power increases rapidly and dominates the error in the efficiency.

Although these errors are not shown in figure 2.25, they have been taken into account for the determination of the maximum efficiency of each wafer shown in figure 2.26(a). However, it was found that the intrinsic variability in the maximum efficiency of different devices of the same wafer was considerably larger than the error in each individual measurement. Therefore, the error bars shown in figure 2.26(a) correspond to the standard deviation of the maximum power conversion efficiency over its measurement in five to ten devices in the same wafer, and therefore are given with the typical 68% confidence limit.

| | | Fabrication Ph | | |
Sample	Material	Si exc. (at. %)	Nominal thick. (nm)	Lum.
2	SRO	16.0 ± 0.6	50	H
3	"	12.0 ± 0.2	"	"
5	"	8.0 ± 0.6	"	"
6	"	6.0 ± 0.2	"	"

Table 2.6: Fabrication parameters of the samples Ph. The SRO layer is obtained by PECVD. In the luminescence column, H stands for *homogeneous* emission. The silicon excess was extracted from XPS measurements.

According to figure 2.26(a), all the efficiencies of the samples Ip are in the order of 10^{-8}. The SRO samples seem to show essentially the same efficiency within the error, regardless of the implantation dose. As for the samples with the thin nitride layer below the SRO, the sample with the lowest implantation can be considered to have the same efficiency as its counterpart without the nitride, within the error bounds. The other two samples present a higher efficiency. Note that these two samples have very similar silicon excesses even though one has twice as high an implantation dose (see table 2.5). In general, except for the sample with a nitride layer and the lowest implantation, the implantation dose does not seem to play a major role in the efficiency in this range of implantation doses.

Figure 2.26(b) plots the current density at which the maximum efficiency is observed. Even though there is a very large statistical variation across different devices of the same wafer, it seems like we could expect more efficient samples to achieve their best efficiency at lower current densities. This will be observed in the next set of samples too.

Fabrication Ph (homogeneous emission)   It is a set of four samples. A 50 nm, non-stoichiometric silicon oxide layer is deposited by PECVD on all samples. The ratio between the precursor gases, SiH_4 and N_2O, is adjusted so as to yield a different silicon excess in each sample, namely 6%, 8%, 12% and 16%. During the PECVD process, the temperature of the substrate was 300 ºC and the radio frequency power density was $0.07 \ W \ cm^{-2}$. The thermal annealing was 1250 ºC for 60 minutes. The polysilicon gate is then deposited and patterned. Table 2.6 lists the fabrication parameters for this set of samples. See sections 2.2.1 and 2.3.1 for a discussion of the fabrication process.

Before moving forward to the characterization of the samples, it is worth noting that the breakdown of these samples comes at lower currents/voltages for higher silicon excesses. In particular, there are no results for the sample with 16% silicon excess because the breakdown occurred so soon after the start of the measurements that it was not even possible to get any luminescence reading.

Figure 2.27 shows the normalized spectra of the samples with 6%, 8% and 12%
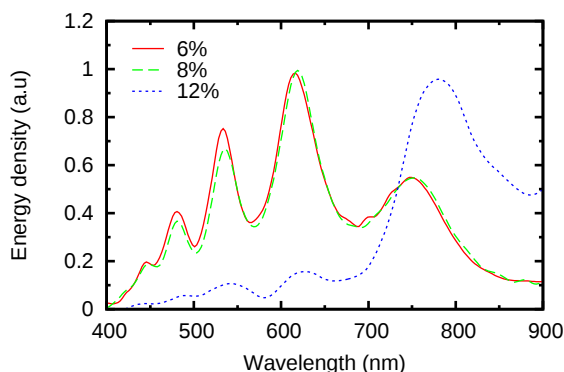
Figure 2.27: Normalized EL spectra for the PECVD samples Ph.
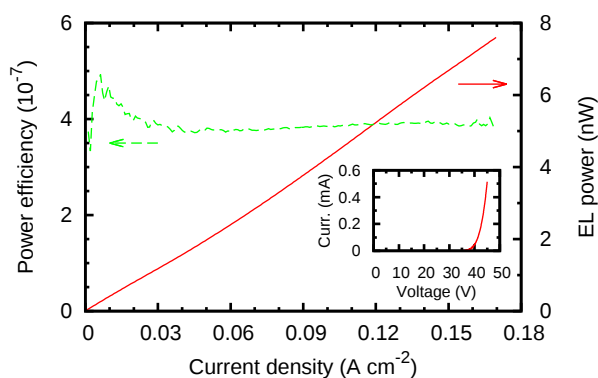
Figure 2.28: Typical EL power and power efficiency as a function of the current density for the samples Ph, exhibiting homogeneous emission. The inset shows a typical IV curve for these samples.

silicon excess. It is clear that there is no significant difference between the spectra of the samples with 6% and 8% silicon excess, whereas the sample with 12% silicon excess shows a comparatively larger contribution in the red region.

Figure 2.28 shows the typical EL power and power efficiency as a function of the current density. The behavior is not very different from what we already showed for the samples Ip. The power increases with the current density in an almost linear way, slightly superlinear in this case, and the efficiency shows a maximum at low currents. For higher currents the efficiency drops to a flat value, or decreases very slowly when the current density is increased, depending on the particular device. Typical powers are in the nanowatt range, roughly one order of magnitude less than samples Ip. However, the efficiency lies in the order of 10^{-7}, roughly one order of magnitude more than that of samples Ip.

From the inset of figure 2.28 it is clear that the voltage does not change much in the relevant range of currents (roughly from 40 to 45 V). That means that dividing
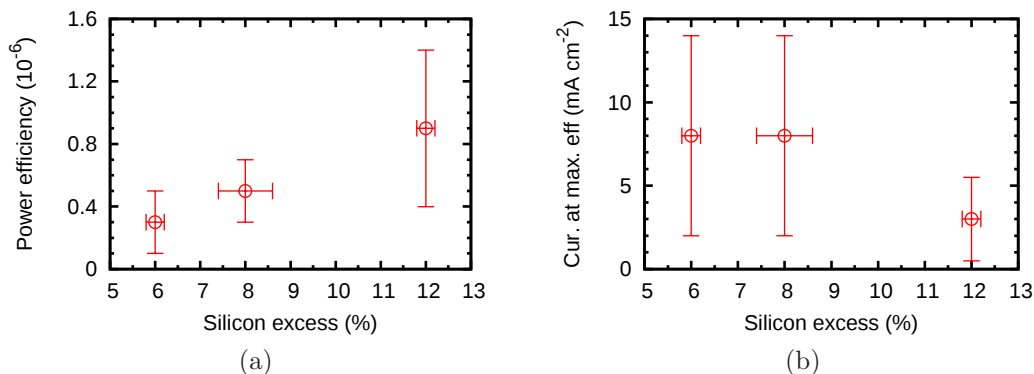
Figure 2.29: (a) Maximum power efficiency and (b) current density at the maximum efficiency, both as a function of the silicon excess for the samples Ph.

the EL between the current or the electrical power (voltage × current) does not make much of a difference as far as the functional form of the result is concerned. That explains why a plot of the quotient between the EL and the current, which is related to the external quantum efficiency, shows essentially the same functional form as the power efficiency plotted in figure 2.28. This is valid in general for all our samples. The linear regions of the EL correspond to flat regions in the quantum efficiency and, per the previous argument, the power efficiency. Superlinear regions in the EL, on the other hand, correspond to increasing efficiency, whereas sublinear regions correspond to decreasing efficiency.

The maximum efficiency of the samples seems to increase with the silicon excess as shown in figure 2.29(a), although the statistical fluctuation across devices of the same wafer is still large (see the discussion about errors for the samples Ip in page 32). It is worth noting that the actual power emitted at maximum efficiency is very low. Therefore, an actual use of the devices at maximum efficiency is not really feasible.

At the same time, it seems that the current at which the maximum efficiency is observed is lower for the sample with the highest efficiency, as shown in figure 2.29(b), although it is difficult to establish a trend because the samples with 6% and 8% show the same current at maximum efficiency and the errors are large. A decrease of the current at maximum efficiency with the silicon excess will be more clearly observed in the next batch of samples.

An important subject, as far as the potential application of the devices is concerned, is their durability. Figure 2.30(a) shows the evolution of the EL with time when the current density is fixed at $6\ \mathrm{mA\ cm^{-2}}$. It is clear, as already noted, that the higher the silicon excess, the shorter the lifetime of the devices before breakdown. Towards the end of their life, the EL signal fluctuates wildly. That is probably related to a blinking of the EL [42] at frequencies above the driving frequency. Since the output signal is filtered by a lock-in amplifier, a change in the frequency spec-
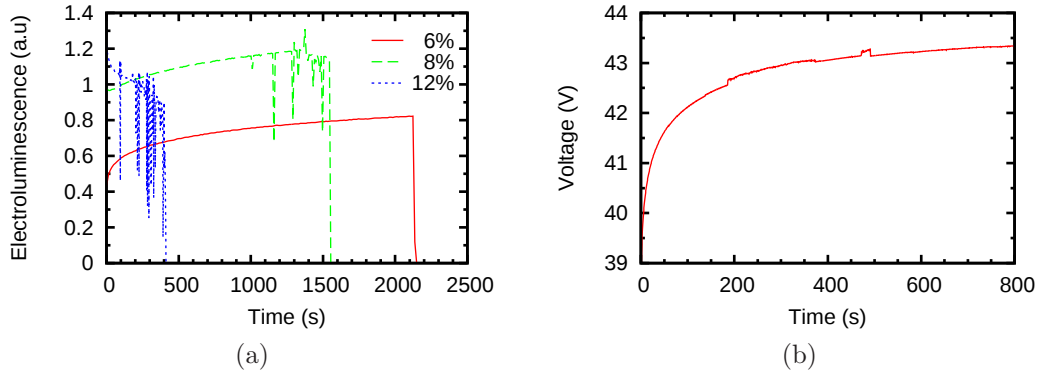
Figure 2.30: Evolution of (a) the EL and (b) the voltage with time for the samples Ph. The current density is fixed at $6 \, \mathrm{mA \, cm^{-2}}$.

trum of the optical signal between measurements will affect the output of the lock-in in an unpredictable way. In this circumstances the measurements become unreliable.

When the EL stops, simultaneously the voltage needed to keep the $6 \, \mathrm{mA \, cm^{-2}}$ drops abruptly to about 1 V, which indicates that either soft or hard breakdown has occurred and the current does not flow uniformly across the device anymore, but instead flows through some preferential pathways [43].

It is known that the degradation of the oxide in MOS systems is related to the absolute amount of charge that has been injected into the oxide [44]. In order to find a relation between the time to breakdown and the injected charge, different devices of the same wafer were operated under different current regimes and their time to breakdown was measured. Due to the long times required to perform this kind of study, the number of data points is small and the estimation of the errors is rough. For the set Ph it was not possible to find a clear relation between the fixed current and the time to breakdown, although such a trend was found in implanted samples, as we shall see in the next batch of samples.

Since the current density driving the device is kept constant at $6 \, \mathrm{mA \, cm^{-2}}$, one would expect the EL to be constant too. However, that is not the case. In the samples with 6% and 8% silicon excess the EL increases with time, whereas it decreases in the sample with 12% silicon excess. That means that the quantum efficiency changes during operation, since we observe a change in the EL intensity while the amount of electrons injected into the system is kept constant.

In figure 2.30(b) the evolution of the voltage during the duration measurement is plotted. The increase in the voltage needed to keep up the $6 \, \mathrm{mA \, cm^{-2}}$ driving the EL indicates that the electrical properties of the material are changing during the operation of the device, and therefore the material itself changes, at least locally. Since the light emission is the result of a competition between the radiative and non-radiative recombinations in the material, a change in the relative oscillator strengths

| Fabrication Ih | | | | | |
Sample	Material	Dose $(10^{16}\ \mathrm{cm}^{-2})$	Si exc. (at. %)	Energy (KeV)	Nom. thick. (nm)	Lum.
2	SRO	6.2	12.0 ± 0.3	25/50	50	H
3	"	7.4	13.0 ± 0.4	"	"	"
4	"	9.0	14.0 ± 0.2	"	"	"
6	"	10.3	15.0 ± 0.5	"	"	"

Table 2.7: Fabrication parameters of the samples Ih. In all the samples the SRO is obtained through a double ion implantation of silicon. The dose corresponds to the sum of the doses during each implantation. In the luminescence column, H stands for *homogeneous* emission.

of the radiative and non-radiative transitions will affect the efficiency of device. The change in the electrical properties of the material during operation strongly hints towards that being what is actually happening.Unfortunately, it is not possible to get a more intimate picture of the phenomenon from our limited set of measurements.

Fabrication Ih (homogeneous emission)   In this batch of samples, the implantation of silicon ions in 50 nm thick thermally grown SiO_2 provides the SRO layer. The implantation takes place through a 30 nm thick LPCVD Si_3N_4 layer, and the process is split into two implantations in order to achieve a flat implantation profile, as described in section 2.2.1. The first implantation is carried out at 25 KeV and the second one at 50 KeV. The thermal annealing takes place at 1100 °C during 4 hours. After the annealing, the Si_3N_4 layer is removed and the polysilicon gate is deposited by the same process used in all other samples and discussed in section 2.2.1. Table 2.7 lists the fabrication parameters of the samples.

Figure 2.31 shows the normalized spectra of all the samples in the set. Clearly the spectrum is the same regardless of the implantation dose. It is worth noting that the spectra are remarkably similar to that of the sample with 12% silicon excess of the fabrication Ph shown in figure 2.27.

The optical power and the power efficiency, plotted in figure 2.32, follow the same trends observed in all other samples so far. The evolution of the optical power with the current density is sublinear and reaches typical values of the order of 1 nW. The power efficiency exhibits a maximum like the other samples and decreases for higher currents.

The values of the maximum efficiency as a function of the implantation dose are represented in figure 2.33(a). It is clear that the efficiency increases with the implantation dose in this batch of samples.

At the same time, figure 2.33(b) seems to indicate that the current at which the maximum occurs decreases with the implantation dose, even though the errors are still significant. Given that the difference in power between the samples is small, we conclude that the increase in the power efficiency is mostly a result of the
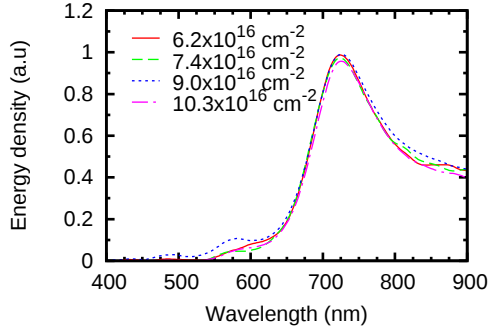
Figure 2.31: Normalized EL spectra of the homogeneous emission in samples Ih.
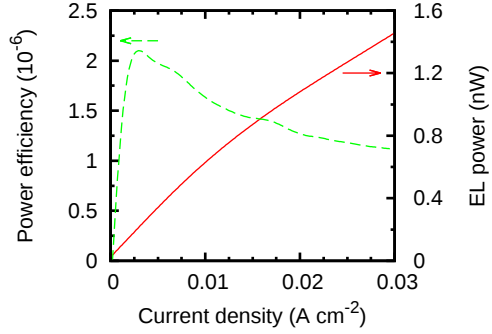
Figure 2.32: Typical EL power and power efficiency as a function of the current density for the samples Ih, exhibiting homogeneous emission.
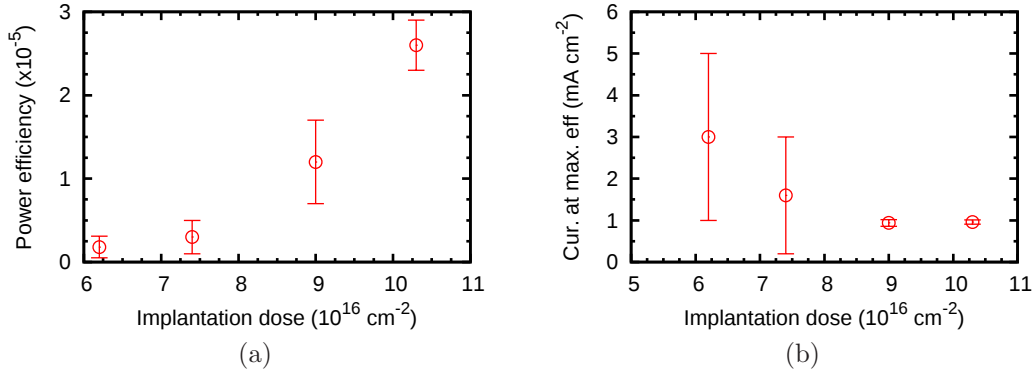
Figure 2.33: (a) Maximum power efficiency and (b) current density at the maximum efficiency, both as a function of the implantation dose for samples Ih.

reduced electrical power needed to reach the maximum efficiency. Since the energy distribution of the emitted photons is the same as per the spectra of figure 2.31, we can also conclude that the external quantum efficiency is indeed higher in the samples with a higher implantation dose.

The evolution of the EL with time is plotted in figure 2.34. Contrary to what was observed in fabrication Ph (PECVD), the breakdown of the ion implanted samples comes first in samples with lower implantation dose. The duration of the implanted devices is decidedly better than those of their PECVD counterparts. The implanted sample with the shortest lifetime is still more durable than the most durable of the PECVD samples of the set Ph.

Contrary to what happened in the set Ph, it was possible to observe a clear trend between the time to breakdown and a fixed current density for the implanted samples, as shown in figure 2.35. The higher the current density, the shorter the
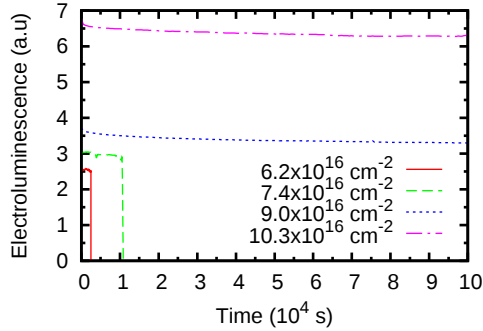
Figure 2.34: Evolution of the EL with time for samples Ih. The current density is fixed at 6 mA cm^{-2}.
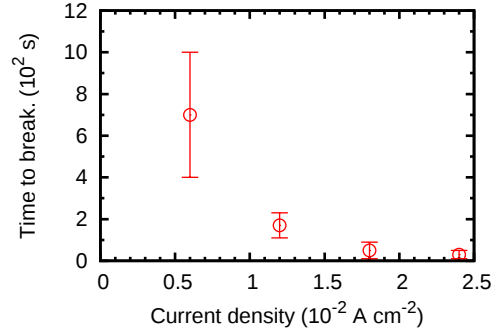
Figure 2.35: Time to breakdown as a function of the fixed current density for the sample Ih-2.

Fabrication Iph						
Sample	Material	Dose (10^{16} cm^{-2})	Si exc. (at. %)	Energy (KeV)	N. thick. (nm)	Lum.
1	SRO/SRN	1.2	0.46 ± 0.02	25/50	30/30	H
2	"	1.5	0 ± 0.09	"	"	"
3	"	3.0	1.3 ± 0.5	"	"	"
5	SRO	1.2	0.46 ± 0.02	"	30	H/P
6	"	1.5	0 ± 0.09	"	"	"
7	"	3.0	1.3 ± 0.5	"	"	"

Table 2.8: Fabrication parameters of the samples Iph. The SRO is obtained through a double implantation. The dose corresponds to the sum of the individual doses. In the samples with double layer, the SRN is on top of the SRO. In the luminescence column, H stands for *homogeneous* emission and P for *points*.

lifetime.

Fabrication Iph (points/homogeneous emission) This batch of samples is similar to Ih in the fabrication process, although the implantation doses are different and the target thickness of the SRO layer is 30 nm instead of 50 nm. The samples were divided in two groups. One of them follows the same fabrication process as Ih. The other one differs in that the 30 nm nitride layer used during the implantation is not removed. As a result we have three samples with an SRO layer and three more samples with SRO/SRN. It is worth noting here that, according to the simulations of the implantation process, there should be some significant implantation tail in the nitride layer, even though the implantation is targeted at the oxide layer. However, XPS analysis showed no significant silicon excess in the nitride layer (see figure 2.15), only a smooth transformation from Si$_3$N$_4$ to SRO through a continuous gradation of silicon oxynitride. Table 2.8 lists the fabrication parameters of this set of samples.

(a) With nitride layer.

(b) Without nitride layer
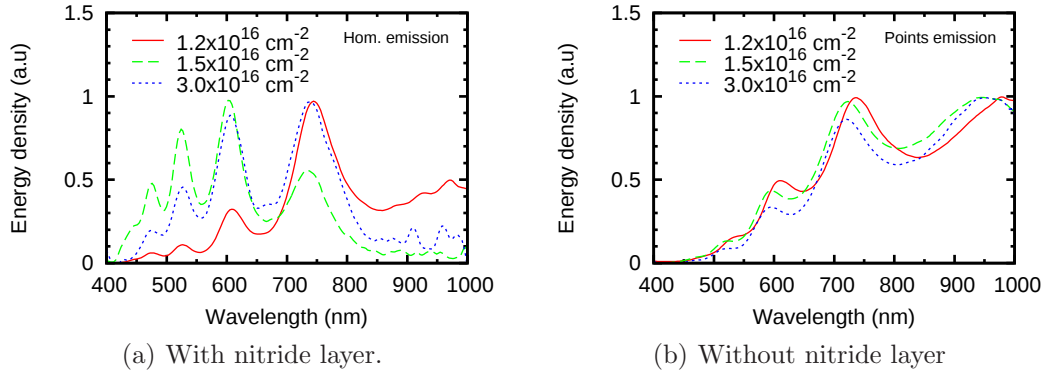
Figure 2.36: Normalized EL spectra of the samples Iph with and without nitride layer. Pictures 2.37(a) to 2.37(c) correspond to the spectra in (a).

(a) Dose 1.2×10^{16} cm^{-2}

(b) Dose 1.5×10^{16} cm^{-2}

(c) Dose 3.0×10^{16} cm^{-2}
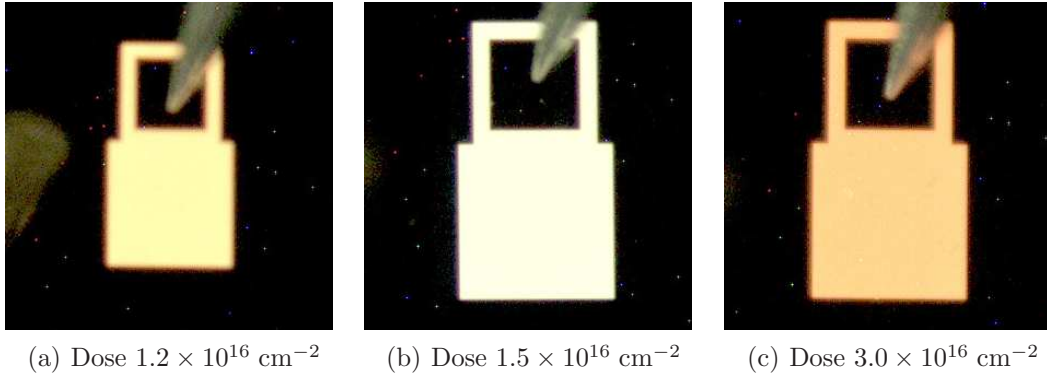
Figure 2.37: EL of the samples Iph 1, 2 and 3, showing homogeneous emission with clearly different color.

Figures 2.36(a) and 2.36(b) show the EL spectra of the samples with the nitride layer (homogeneous emission) and without it (emission through points), respectively.

Figure 2.37 shows pictures of the devices whose spectrum is represented in figure 2.36(a). There is a clear difference in the color of each implantation dose.

Figures 2.38(a) and 2.38(b) show the typical EL power and power efficiency for the samples Iph with and without the nitride layer, respectively, as a function of the current density. The functional form follows the same trend in both groups of samples, which is also similar to what we have seen in all other samples. The optical power is of the order of 10 nW in both groups, but the operation currents are very different: a few mA per square centimeter for the samples with the nitride and homogeneous emission, and a few tenths of A per square centimeter for the samples without the nitride and emission through points. As far as the efficiency goes, the

(a) With nitride, homogeneous emission

(b) Without nitride, emission through points
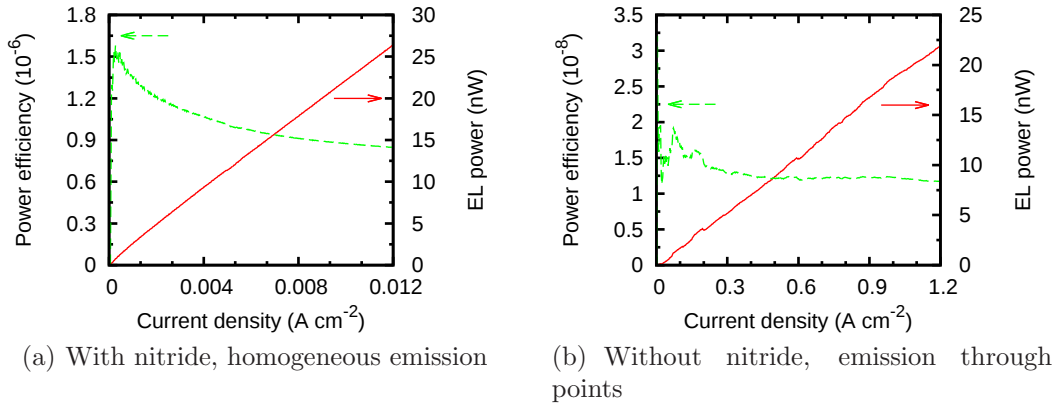
Figure 2.38: Typical EL power and power efficiency as a function of the current density for the samples Iph, with and without the nitride layer.
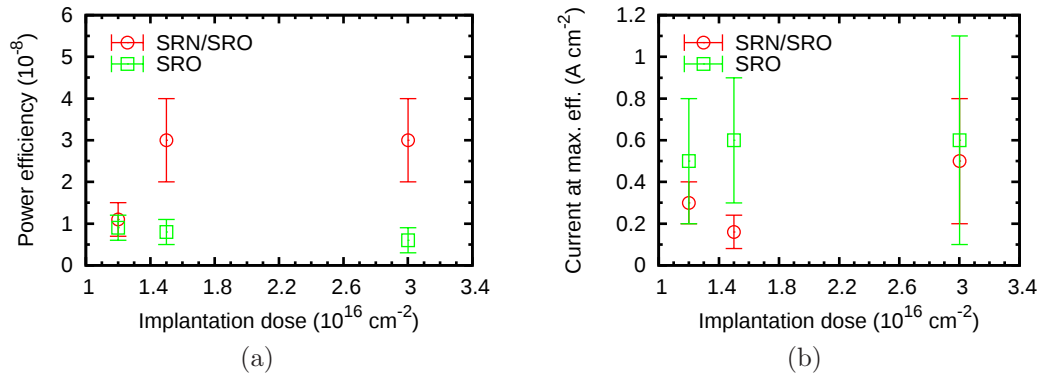

(a)

(b)
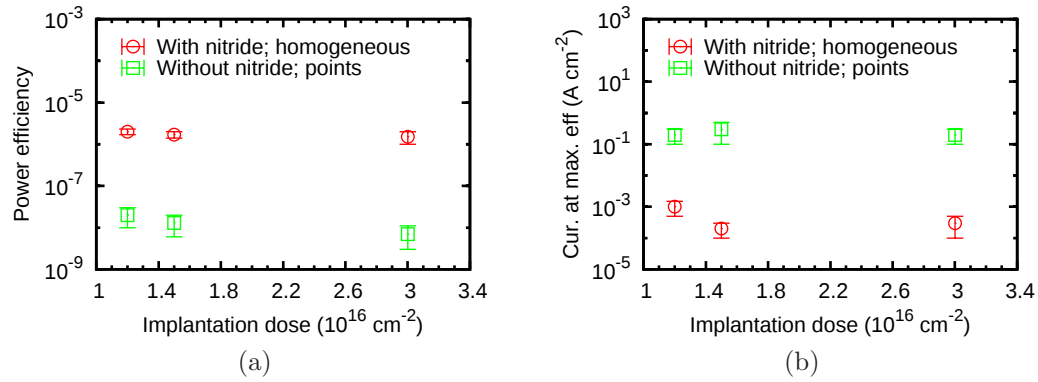
Figure 2.39: (a) Maximum power efficiency and (b) current density at the maximum efficiency, both as a function of the implantation dose for the samples Iph.

samples with homogeneous emission reach the 10^{-6} mark whereas the samples with emission through points stay at 10^{-8}.

The values of maximum efficiency as a function of the implantation dose are plotted in figure 2.39(a). It is readily apparent that the samples with homogeneous emission are two orders of magnitude more efficient than those with emission through points. That is why we can get the same emission intensity at much lower currents.

Although it is not clearly seen in the logarithmic representation of figure 2.39(a), the most efficient sample with homogeneous emission (dose 1.2×10^{16} cm^{-2}) records a mean efficiency value of 2.0×10^{-6}, whereas the less efficient one (dose 3.0×10^{16} cm^{-2}) reaches a mean value of 1.5×10^{-6}. However, when the error bars are considered, it is not clear that the efficiency decreases with the implantation dose,

(a) With nitride, current density fixed to $9 \, \text{mA} \, \text{cm}^{-2}$.

(b) Without nitride, current density fixed to $2 \, \text{A} \, \text{cm}^{-2}$.
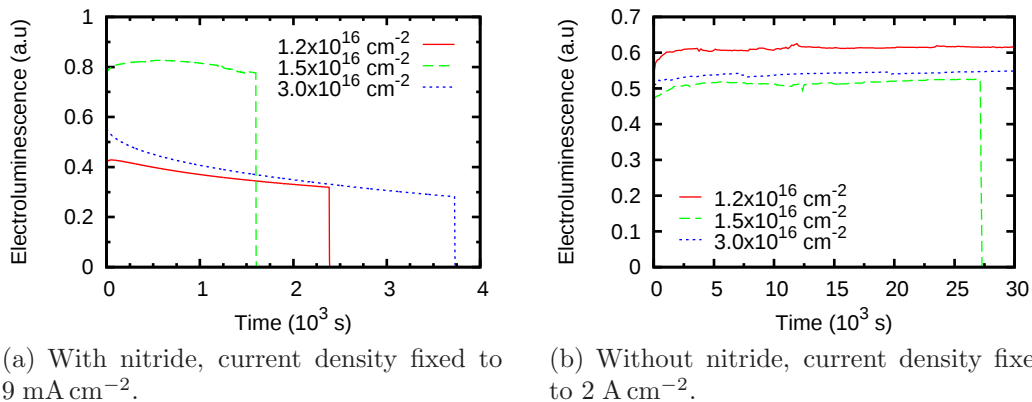
Figure 2.40: Evolution of the EL with time for the samples Iph with and without the nitride layer.

as the mean values of the efficiency indicate. In fact, all three samples could be considered to show the same efficiency within the statistical error. The same can be said of the samples showing emission through points, since the maximum and minimum mean efficiency values, 2×10^{-8} and 0.7×10^{-8} for implantation doses 1.2×10^{16} and $3.0 \times 10^{16} \, \text{cm}^{-2}$, respectively, can also be considered to be the same once the error bars are factored in.

It is worth noting that the samples with emission through points show a more unstable emission than their counterparts with homogeneous emission. See for instance the efficiency curve in figure 2.38(b).

In figure 2.39(b) it can be confirmed that the samples exhibiting homogeneous emission consistently reach their maximum efficiency at much lower currents than their counterparts with emission through points. However, no definite trend can be identified with the implantation dose.

Figures 2.40(a) and 2.40(b) show the evolution of the EL with time for the samples with and without nitride, respectively. The current density was fixed to $9 \, \text{mA} \, \text{cm}^{-2}$ for the samples with nitride and to $2 \, \text{A} \, \text{cm}^{-2}$ for the samples without nitride.

Even though the samples without the nitride show a more unstable EL, over long periods of time their emission is more steady than for the samples with the nitride layer. Moreover, their time to breakdown is roughly 10 times longer.

2.3.2.2 Discussion

Table 2.9 summarizes the sample sets that have been studied in this section, with the basic fabrication parameters of each sample, their maximum efficiency and the current density at which the maximum efficiency is observed. Note that every fabrication usually consists of more wafers than listed in table 2.9, such as samples intended for PL or X-ray photoelectron spectroscopy studies. Other fabrications

not listed in table 2.9 were carried out for other purposes, such as measuring the complex index of refraction of SRO and SRN layers or PL studies. Only the samples that feature complete devices are listed in table 2.9.

As far as the emission spectrum is concerned, it is clear, after looking at the figures of the previous section, especially 2.24, 2.31 and 2.36(b), that the clean, clear cut differences that can be observed in the PL (see figure 2.13) are washed out by the interference effects. Since the differences in the EL spectra are due to differences in both the intrinsic emission of the layer and the interference patterns, mostly due to differences in the thickness of the layers, the interpretation of the spectra in terms of the origin of the EL is not feasible.

However, from the perspective of a final device, we do not need to concern ourselves with the origin of the luminescence, and we can just focus on the features of the EL spectra as it is. From this point of view, it is clear that single SRO layers provide a limited modulation in color by using the silicon content as a parameter.

Previous works have indicated the range of silicon contents that lead to optimal PL emission in both implanted and PECVD SRO layers [45]. For implanted samples, the variation of the implantation dose around the previously determined optimal range, from roughly 10^{16} cm^{-2} to 10^{17} cm^{-2}, does not provide a useful change in the spectrum, which is to be expected since the PL measurements tell us that we can expect shifts in the peak emission of as much as 100 nm. In order to bring significant color tunability to the table it is necessary to add a second emission band that can be combined with the SRO band to generate the different spectra shown in figure 2.36(a), which results in a significant color tunability, as exemplified in figures 2.37(a) to 2.37(c). That is one reason why the SRO/SRN bilayers are more interesting than single SRO or single SRN layers as far as a the color tunability is concerned.

From the spectra of fabrication Iph we gather that the samples with the nitride layer (figure 2.36(a)) present different spectra depending on the implantation dose. The difference is significant enough so as to translate to a clearly different color of the emission, as seen in figures 2.37(a) to 2.37(c). From the discussion in section 2.2.2.3 and figures 2.14(a) and 2.14(b), corresponding to the samples Iph with the lowest and highest implantation doses, respectively, we know that these spectra have contributions from two bands, one due to the oxide and the other due to the nitride. It is the relative intensity between both bands that determines the final spectrum and color. Even though the EL spectra shown in figure 2.36(a) appears quite different from the PL spectra of figures 2.14(a) and 2.14(b), it will be shown in chapter 3 that the differences are exclusively due to interference effects.

Regarding the spectra of the samples Iph without the Si$_3$N$_4$ layer (figure 2.36(b)), the differences between their spectra are much less significant from the color perspective. Nevertheless, there is a small difference in the position of the peaks, the sample with lowest implantation dose being redshifted with respect to the sample with the highest dose. This is opposite to what we observed in the PL spectra: the PL peak of the SRO layer with the lowest implantation dose appears blueshifted with respect to the peak of the SRO layer with the highest implantation dose (see

Sample #	Dose (10^{16} cm^{-2})	Si exc. (at. %)	Material	Thickness (nm)	EL	Efficiency	Current (mA cm^{-2})
Fabrication Ip (implantation)
1	1.2	1.3 ± 0.3	SRN/SRO	10/30	P	$(1.1 \pm 0.4) \times 10^{-8}$	300 ± 100
2	1.5	4.6 ± 0.3	"	"	P	$(3 \pm 1) \times 10^{-8}$	160 ± 80
3	3.0	4.0 ± 0.6	"	"	P	$(3 \pm 1) \times 10^{-8}$	500 ± 300
4	1.2	1.3 ± 0.3	SRO	30	P	$(9 \pm 3) \times 10^{-9}$	500 ± 300
5	1.5	4.6 ± 0.3	"	"	E/P	$(8 \pm 3) \times 10^{-9}$	600 ± 300
6	3.0	4.0 ± 0.6	"	"	E	$(6 \pm 3) \times 10^{-9}$	600 ± 500
Fabrication Ph (PECVD)
2	NA	16.0 ± 0.6	SRO	50	None	NA	NA
3	"	12.0 ± 0.2	"	"	H	$(9 \pm 5) \times 10^{-7}$	8 ± 6
5	"	8.0 ± 0.6	"	"	H	$(5 \pm 2) \times 10^{-7}$	8 ± 6
6	"	6.0 ± 0.2	"	"	H	$(3 \pm 2) \times 10^{-7}$	3 ± 2
Fabrication Ih (implantation)
2	6.2	12.0 ± 0.3	SRO	50	H	$(1.8 \pm 1.3) \times 10^{-6}$	3 ± 2
3	7.4	13.0 ± 0.4	"	"	H	$(3 \pm 2) \times 10^{-6}$	1.6 ± 1.4
4	9.0	14.0 ± 0.2	"	"	H	$(1.2 \pm 0.5) \times 10^{-5}$	0.94 ± 0.08
6	10.3	15.0 ± 0.5	"	"	H	$(2.6 \pm 0.3) \times 10^{-5}$	0.96 ± 0.05
Fabrication Iph (implantation)
1	1.2	0.46 ± 0.02	SRO/SRN	30/30	H	$(2.0 \pm 0.3) \times 10^{-6}$	1 ± 0.5
2	1.5	0 ± 0.009	"	"	H	$(1.7 \pm 0.3) \times 10^{-6}$	0.2 ± 0.1
3	3.0	1.3 ± 0.5	"	"	H	$(1.5 \pm 0.5) \times 10^{-6}$	0.3 ± 0.2
5	1.2	0.46 ± 0.02	SRO	30	H/P	$(2 \pm 1) \times 10^{-8}$	200 ± 100
6	1.5	0 ± 0.009	"	"	H/P	$(1.3 \pm 0.7) \times 10^{-8}$	300 ± 200
7	3.0	1.3 ± 0.5	"	"	H/P	$(0.7 \pm 0.4) \times 10^{-8}$	200 ± 100

Table 2.9: Summary of the samples with complete devices that have been studied in section 2.3.2.1.

figures 2.14(a) and 2.14(b)). It is entirely possible that a difference in the thickness of the polysilicon layer is responsible for the shift observed in the EL spectra.

A sample with a broad emission such as the one represented in figure 2.37(c) would allow us to select a narrow emission band with a reasonable intensity in a broad range of the visible spectrum, for instance with an optical cavity such as the one presented in [46]. More work should be devoted to such bilayer systems in order to control the relative intensity of each band.

Regarding the optical power and the power efficiency as a function of the driving current, the trends are very similar in all the studied samples. However, there are important differences in the typical values of power and efficiency depending on the kind of emission, points or homogeneous, as we stated at the beginning of the section.

The samples with emission through points exhibit typical optical powers around 10 nW, whereas the samples with homogeneous emission vary between around 1 and 10 nW, depending on the characteristics. These are, overall, comparable intensities. However, the current density needed to drive the emission is significantly lower in the samples with homogeneous emission: around 1 $A\,cm^{-2}$ for emission through points and around 0.01 $A\,cm^{-2}$ for homogeneous emission. Similar emission powers at currents roughly two orders of magnitude lower imply better efficiencies for the samples with homogeneous emission. It is worth noting here that the voltages vary depending on the sample, and are in fact lower in the samples with emission through points. However, they are always of the same order. As an example, the voltage corresponding to the 20 nW emission power in figures 2.38(a) and 2.38(b), corresponding to homogeneous and point emission, respectively, is 61 V and 38 V. In the efficiency plots of section 2.3.2 we see that the efficiency of the samples with emission through points hardly hits the 10^{-8} mark, whereas that of samples with homogeneous emission varies in the range $10^{-7} \sim 10^{-5}$.

In this regard, the fabrication Iph is particularly interesting because it allows us to compare both emission regimes on closely related samples. The only difference between the subset exhibiting emission through points and the subset exhibiting homogeneous emission is the top nitride layer, or lack thereof, all other parameters being the same. Figures 2.38(a) and 2.38(b) clearly show that at very similar powers, around 20 nW, the current is two orders of magnitude lower in the sample with homogeneous emission (with nitride layer), with voltages in the same order, as we stated in the previous paragraph, which results in an efficiency improved by two full orders of magnitude.

Regarding the resistive component of the active layer, we could consider that the current flow under the homogeneous emission regime takes place in an array of parallel resistances, each one corresponding to a small vertical section of the layer, and all of them with similarly high values. Under these circumstances, the current flow will be the same in every resistance and therefore a uniform current will flow through the area of the device, resulting in a uniform emission too.

If a few preferential, low resistivity paths are found in the device, then each of this paths contributes a low resistance in parallel with the existing array of high
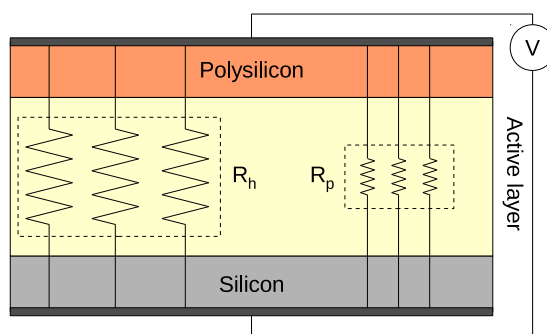
Figure 2.41: The resistive component of the active layer split into two parallel resistances. The current giving rise to homogeneous emission flows through a large R_h, whereas the current giving rise to points flows through a smaller R_p.

resistances. Then we can picture the resistive part of the active layer as two resistances in parallel, R_h and R_p (see figure 2.41). R_h represents an array of high resistances in parallel, whereas R_p represents an array of low resistances in parallel, one for each conductive path. At a fixed voltage V, the current flowing uniformly will be $I_h = V/R_h$, whereas the current flowing through the preferential paths will be $I_p = V/R_p$. The ratio $I_p/I_h = R_h/R_p$ tells us that for R_h a few orders of magnitude larger than R_p, the current flowing uniformly through the layer will be only a small fraction of the current flowing through the preferential paths. That may lead to a situation where the current through the preferential paths is high enough so as to generate detectable emission points whereas the uniform component of the current is not high enough to generate the homogeneous component of the emission. If we can increase the total current without breaking down the device, we will eventually reach a point where the uniform component of the emission is high enough so as to make the homogeneous emission detectable, as observed in the sample pictured in figure 2.23(b). However, if the ratio R_h/R_p is too large, it may not be possible to observe both components at the same time.

As we have seen, the uniform conduction is more desirable than the conduction through preferential paths, and the fabrication Iph indicates that the nitride layer allows for a more controlled injection into the stack, making the appearance of preferential paths more unlikely.

As far as the durability of the devices is concerned, it is difficult to draw any clear conclusion. Fabrication Ph (PECVD) and the subset of Iph with the nitride layer (implanted), present similar times to breakdown. The subset of Iph without the nitride layer presents 10 times longer durations, and the samples with higher implantation dose of the fabrication Ih are 10 times more durable still. Only the samples in Ih were consistent enough so as to show a definite relation between the fixed current density and the time to breakdown. In all other fabrications, the statistical variation between individual devices is too large to allow for any conclusion

to be reached.

2.3.3 Electrical characterization

In this section we will briefly review the electrical characterization techniques used in this work, namely current-voltage and capacitance-voltage characterization, and discuss the results of the characterization of our samples. Our main goal is to study in detail the injection mechanisms in the active layer.

2.3.3.1 Current-voltage characterization

Current-voltage (I-V) characterization is the most basic electrical characterization technique. In order to carry out I-V characterization, the setup presented in section 2.3.2 (figure 2.18) has been used, although for measurements at low voltages an HP4155B semiconductor parameter analyzer was also available instead of the Keithley 2430.

One important piece of information we want to extract from the I-V characterization is the injection mechanism taking place in the device. In the following paragraphs the mechanisms considered for our devices will be presented. A Matlab script was developed[2] to fit the models to the data and extract their relevant parameters in order to assess the goodness of the fit.

The equations shown below have been extracted from [47], [48] and [49]. The meaning of the constants used in the equations is shown in table 2.10[3].

Direct tunneling   An electron moves through a potential barrier in a single hop. It depends strongly on the electric field and is independent of the temperature.

$$J = \frac{q^2}{2\pi h d^2} \left[\left(\phi_B - \frac{E_i\, d}{2} \right) \exp\left(-\frac{4\pi d\sqrt{2m^*q}}{h} \sqrt{\phi_B - \frac{E_i\, d}{2}} \right) \right.$$
$$\left. - \left(\phi_B + \frac{E_i\, d}{2} \right) \exp\left(-\frac{4\pi d\sqrt{2m^*q}}{h} \sqrt{\phi_B + \frac{E_i\, d}{2}} \right) \right] \quad (2.20)$$

Thermoionic emission   Due to the temperature, some electrons gain enough energy to climb the potential barrier. It depends strongly on the temperature.

$$J = R\,T^2 \exp\left(\frac{-q\left(\phi_B - \sqrt{qE_i\,/\,4\pi\epsilon_i}\right)}{k\,T} \right) \quad (2.21)$$

---

[2]The script was originally written by M. Baselga and later extended by J. Juvert.

[3]Note that, for the formulas to be correct, the barrier height or equivalent must be in eV, but the energy part of Boltzmann's constant and Planck's constant is in Joules, not in eV. The reason is that for each barrier height in the formulas there is always a multiplying elementary charge q that transforms it into Joules. Therefore, barrier heights can (and must) be written in eV, which is much more convenient.

R	Effective Richardson constant	
T	Temperature	K
k	Boltzmann constant	J/K
q	Elemental charge	C
ϕ_B	Barrier height seen by an injected electron	V or eV
E_i	Electric field in the insulator	V/m
ϵ_i	Absolute permittivity of the insulator	F/m
ΔE_{ac}	Activation energy of electrons	J
ΔE_{ai}	Activation energy of ions	J
μ	Electron mobility	$\mathrm{m}^2/(\mathrm{V{\cdot}s})$
h	Planck's constant	J·s
m^*	Effective mass of an electron	Kg
m_e	Mass on an electron	Kg
d	Thickness of the dielectric layer	m

Table 2.10: Symbols used in the equations of injection mechanisms along with their meaning and units.

Poole-Frenkel emission   It consists on emission of trapped electrons into the conduction band due to thermal excitation. Here, the barrier height is the depth of the trap potential well.

$$ J \propto E_i \, \exp\left(\frac{-q\left(\phi_B - \sqrt{qE_i\,/\,\pi\epsilon_i}\right)}{k\,T} \right) \tag{2.22} $$

A plot of $\ln\left(\frac{J}{E_i}\right)$ vs. $\sqrt{E_i}$ will give a straight line that we can fit with a polynomial $p(x) = p_1\,x + p_2$. From p_1 we can extract:

$$ \epsilon_i = \frac{q^3}{p_1^2\,\pi k^2 T^2} \tag{2.23} $$

Note that we cannot extract anything from p_2 because of the $\propto$ relation in equation 2.22, which inserts an unknown constant in p_2.

Ohmic conduction   Typical conduction in a metal.

$$ J \propto E_i \, \exp\left(\frac{-\Delta E_{\mathrm{ac}}}{k\,T} \right) \tag{2.24} $$

A plot of J vs. E_i will give a straight line crossing the point $(0, 0)$. We can fit a polynomial $p(x) = p_1\,x + p_2$. The independent term should be $p_2 = 0$. Since the proportionality constant in equation 2.24 is unknown we cannot extract any information from the fit.

Ionic conduction   This is similar to a diffusion process.

$$J \propto \frac{E_i}{T} \exp\left(\frac{-\Delta E_{\text{ai}}}{k\,T}\right) \tag{2.25}$$

Space-charge limited injection   Arises when carriers are injected into a material where very little compensating charge exists, such as insulators or lightly doped semiconductors. For a *trap-free* insulator, space charge limited injection follows the law:

$$J = \frac{9\epsilon_i \mu E_i^2}{8d^3} \tag{2.26}$$

A plot of J vs. E_i^2 will give a straight line crossing the point $(0, 0)$. We can fit a polynomial $p(x) = p_1\,x + p_2$. The independent term should be $p_2 = 0$ and from the value of p_1 we can extract:

$$\mu = \frac{8dp_1}{9\epsilon_i} \tag{2.27}$$

For a description of the corrections involved when taking into account shallow traps in the insulator, see section 9.1.2 in [50]

Fowler-Nordheim tunneling   This is conceptually equivalent to direct tunneling, with the difference that, in this case, the electron tunnels through a partial width of the potential barrier instead of the full width.

$$J = \frac{q^2}{8\pi h \phi_B} \frac{m}{m^*} E_i^2 \exp\left(-\frac{8\pi\sqrt{2qm^*}\,\phi_B^{\frac{3}{2}}}{3hE_i}\right) \tag{2.28}$$

A plot of $\ln\left(\frac{J}{E_i^2}\right)$ vs. $\frac{1}{E_i}$ will give a straight line that we can fit with a polynomial $p(x) = p_1\,x + p_2$. From the parameters p_1 and p_2 we can extract:

$$\phi_B^3 \frac{m^*}{m_e} = \left(\frac{3\,h\,p_1}{8\pi\,\sqrt{2\,q\,m_e}}\right)^2 \tag{2.29}$$

$$\phi_B \frac{m^*}{m_e} = \frac{q^2}{8\pi\,h\,\exp p_2} \tag{2.30}$$

Therefore, both the slope and the independent term of the polynomial contain information on the two relevant parameters. Either of them can be calculated provided the other one is known.

Schottky emission   Takes place in the conduction through metal-semiconductor contacts.

$$J \propto T^2 \exp\left(-\frac{q(\phi_B - \sqrt{q\,E_i\,/\,4\pi\epsilon_i})}{k\,T}\right) \tag{2.31}$$

Trap Assisted Tunneling Corresponds to a tunneling process across the dielectric layer with two or more steps involved.

$$J \propto \exp\left(-\frac{8\pi\sqrt{2qm^*}}{3hE_i}\phi_B^{\frac{3}{2}}\right) \tag{2.32}$$

A plot of $\ln J$ vs. $\frac{1}{E_i}$ will give a straight line that we can fit with a polynomial $p(x) = p_1\,x + p_2$. From the parameter p_1 we can extract either of the following two variables if we fix the value of the other:

$$\phi_B = \left(\frac{3hp_1}{8\pi\sqrt{2m^*q}}\right)^{\frac{2}{3}} \tag{2.33}$$

$$\frac{m^*}{m_e} = \frac{1}{2m_e q\phi_B^3}\left(\frac{3hp_1}{8\pi}\right)^2 \tag{2.34}$$

Note that p_2 does not yield any information since the proportionality constant in equation 2.32 is unknown. Also, ϕ_B in equation 2.32 refers to trap depth.

2.3.3.2 Capacitance-voltage characterization

The capacitance-voltage measurement is a powerful technique that allows the extraction of several parameters of a MIS system. The differential capacitance of a system is defined as:

$$C(V) = \frac{\mathrm{d}Q(V)}{\mathrm{d}V} \tag{2.35}$$

That is the amount of charge added to the system when the voltage is increased by a small amount. It is measured by superposing a small AC voltage to the DC voltage at which we want to measure C. The AC voltage frequency is typically 10 KHz to 1 MHz and 10 to 20 mV in amplitude. Figure 2.42 depicts a typical connection scheme for CV measurements. In our experiments, the DC and AC sources as well as the AC volt and ammeters are integrated in an HP4192A LF impedance analyzer.

Figure 2.43 shows a typical high frequency CV curve for one of our capacitors, which closely resembles that of a typical MOS capacitor with p type substrate. See any classic books like [51] for insight into the CV behavior of a typical MOS capacitor.

In the following paragraphs we will discuss the parameters that can be extracted from capacitance-voltage measurements.

Constants and preliminary considerations Table 2.11 shows the meaning and value[4] of the symbols used in this section.

---

[4]Source: http://physics.nist.gov/constants
http://www.ioffe.rssi.ru/SVA/NSM/Semicond/Si/bandstr.html
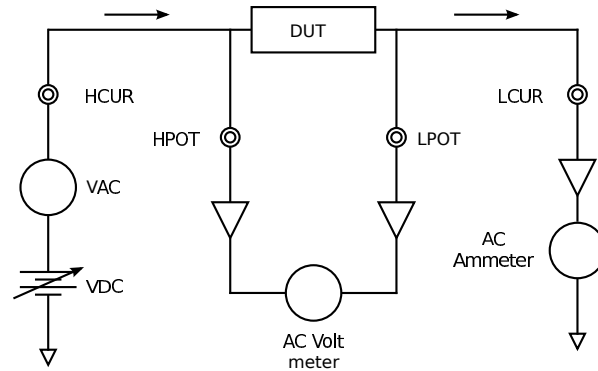
Figure 2.42: Connection scheme for typical CV measurements. The arrows indicate the current flow through the device under test (DUT).
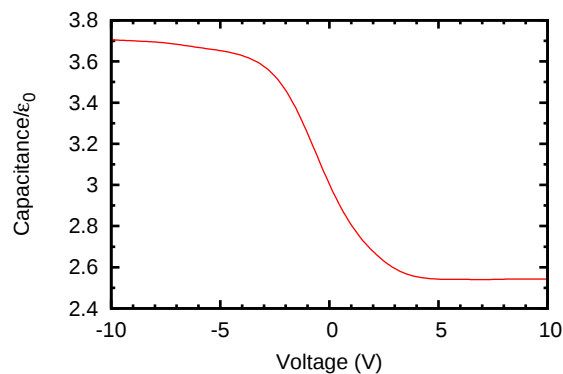
Figure 2.43: Typical high frequency CV curve for a MOS capacitor with p substrate.

ϵ_0	Permittivity of vacuum	$8.8542 \cdot 10^{-12}$	F/m
K_s	Dielectric constant of Silicon	11.68	
k	Boltzmann's constant	$1.38065 \cdot 10^{-23}$	J/K
q	Magnitude of electronic charge	$1.60218 \cdot 10^{-19}$	C
E_g	Silicon bandgap (300 K)	1.12	eV

Table 2.11: Value of the constants used in the expressions of section 2.3.3.2.

In the equations that follow, we will always consider the capacitance in Farads, not Farads per unit area, since this is what is usually measured. Also, p-type silicon will be assumed.

Extended information about the equations used hereafter can be found in [51] and [52]. In particular, approximations used, domain of applicability and other methods to calculate the same magnitude can be found in those references.

Oxide capacitance C_{ox}   The oxide capacitance of a MOS structure or, more generally, the insulator capacitance, can be extracted from the capacitance-voltage characterization. There are several methods, all of them covered in section 6.4 of [52]. The proposed method involves the following equation:

$$\frac{1}{C_{hf,acc}} = \frac{1}{C_{ox}} + \left(\frac{1}{2\beta} \left| \frac{d(1/C_{hf,acc}^2)}{dV} \right| \right)^{\frac{1}{2}} \tag{2.36}$$

where $C_{hf,acc}$ is the high frequency capacitance in the accumulation region. For this method, we plot $1/C_{hf}$ versus the square root of the modulus of the derivative of $1/C_{hf}^2$ with respect to the gate voltage. The plot should show a linear region in the strong accumulation range of the measurement. The intercept with the y axis of a linear fit in this region yields $1/C_{ox}$.

The value calculated with this method is more accurate than that given by the maximum capacitance in the accumulation region of the CV curve. It is worth noting here that the CV curve may saturate at a higher capacitance than C_{ox} because of a large density of interface trap levels. Also, a large oxide fixed charge density may shift the CV curve to such extreme gate biases that the oxide breaks down before accumulation is reached.

Oxide thickness t_{ox}   Once C_{ox} is known, oxide thickness is easily calculated from the equation (see section 3.2.1 of [51] or section 6.4.1 of [52])

$$t_{ox} = \frac{K_{ox}\,\epsilon_0\,A}{C_{ox}} \tag{2.37}$$

where K_{ox} is the dielectric constant of the insulator (also known as relative permittivity) and A stands for the area of the device. Note that the uncertainty in t_{ox} is proportional to that in K_{ox}

Maximum depletion layer width W_{max}   According to section 2.2.1 in [52] the following relation holds

$$W = K_s\,\epsilon_0\,A\left(\frac{1}{C} - \frac{1}{C_{ox}} \right) \tag{2.38}$$

Therefore, the maximum depletion layer width can be calculated from the high frequency CV characteristics by letting C be C_{min} in equation 2.38. Regarding the value of C_{min}, read the discussion at the end of section 6.2.1 in [52].

A relation between $W_{\max}$ and the doping density N_A can be found in section 2.3.4.c of [51]

$$W_{\max} = \left(\frac{2\,K_s\,\epsilon_0\,\psi_L}{q\,N_A} \right)^{\frac{1}{2}} \tag{2.39}$$

where

$$\psi_L = 2.1\,\frac{q\,\phi_B}{k\,T} + 2.08 \tag{2.40}$$

and

$$q\,\phi_B = E_i - E_F = k\,T\,\ln\frac{N_A}{n_i} \tag{2.41}$$

E_i and E_F stand for the intrinsic Fermi level and the actual Fermi level respectively. An accurate value of n_i (intrinsic carrier density) at different temperatures can be found in [53].

Thus, once W is known from equation 2.38, a value of N_A can be worked out through equations 2.39, 2.40 and 2.41. However, a much more convenient way of calculating N_A is discussed in the next section.

Substrate doping density N_A The high frequency CV curve can be used to extract the substrate doping profile (see section 2.2.1 of [52])

$$N_A(W) = \frac{2}{q\,K_s\,A^2\,\frac{d(1/C^2)}{dV}} \tag{2.42}$$

The depth at which the doping density is evaluated is obtained from equation 2.38. That is, the derivative in equation 2.42 is evaluated at a voltage V_0, which corresponds to a capacity C_0 according to our measured CV curve. This C_0 yields a depletion layer width W_0 according to equation 2.38, and that is the depth at which N_A is being evaluated.

However, if our substrate can be assumed to be uniformly doped, the following equation is more convenient (see section 2.2.3 of [52])

$$N_A = \frac{4\,\phi_B}{q\,K_s\,\epsilon_0\,A^2}\,\frac{R^2\,C_{\mathrm{ox}}^2}{(1-R)^2} \tag{2.43}$$

where ϕ_B has been defined in equation 2.41 and $R = C_{\mathrm{inv}}/C_{\mathrm{ox}}$. C_{inv} is the high frequency capacitance in strong inversion (see the discussion at the end of section 6.2.1 of [52]). Note that substitution of equation 2.41 into equation 2.43 leads to a transcendental equation.

Yet another way to calculate the uniform doping density arises from the following empirical relation between C_{inv} and N_A for silicon at room temperature (see chapter 14 of [54])

$$\log(N_A) = 30.38759 + 1.68278\,\log(C_1) - 0.03177\,(\log(C_1))^2 \tag{2.44}$$

where log is the logarithm to base 10, $C_1 = R\,C_{\mathrm{ox}}/[A\,(1-R)]$ and R has been defined in the previous paragraph.

Flatband voltage V_{FB} A graphical method for the determination of V_{FB} can be found in exercise 6.2 of [52]. However, the second derivative introduces too much noise. Smoothing of data helps, but it also distorts the measurement.

Another method is to calculate the capacitance at the flatband voltage, C_{FB}, and from that and the experimental CV curve find V_{FB}.

According to section 6.2.1 of [52] and section 3.2.1 of [51], the total low frequency capacitance is

$$C = \frac{C_{\text{ox}} C_S}{C_{\text{ox}} + C_S} \tag{2.45}$$

In section 6.2.2 of [52] and section 3.2.3 of [51] we find C_S in the flatband condition to be

$$C_{S,FB} = \frac{A K_s \epsilon_0}{\lambda_D} \tag{2.46}$$

where

$$\lambda_D = \left(\frac{k T K_s \epsilon_0}{q^2 N_A} \right)^{\frac{1}{2}} \tag{2.47}$$

is the Debye length. Therefore, substituting equation 2.47 into 2.46 and that into 2.45 we can find C_{FB} provided we know C_{ox}, which we can extract from either the low frequency or the high frequency CV curve as seen in section 2.3.3.2.

Oxide charge   For a discussion of the different types of charges in MOS structures see section 6.1 of [52]. Before considering the equations in the following subsections, please note that there are many contributions to the flatband voltage. According to equation 10.51 of [51] the flatband voltage can be written as:

$$V_{FB} = -\frac{Q_f + Q_{it}}{q} + \phi_{MS} \tag{2.48}$$

This means that both the work function difference between the gate and the silicon and all kinds of oxide charges contribute to the flatband voltage. A particular kind of oxide charge can be derived from a measurement of the flatband voltage if all remaining contributions are known or can be safely assumed. See the references to find out more about the assumptions made in the following equations.

According to section 6.2.4 of [52], the fixed charge, Q_f and N_f, near the oxide–silicon interface can be calculated as follows:

$$Q_f = (\phi_{MS} - V_{FB}) C_{\text{ox}} \tag{2.49}$$

For a MOS structure with n^+ poly-Si gate and p-Si substrate, ϕ_{MS} can be approximated by (see section 6.2.5 of [52]).

$$\phi_{MS} = -\frac{E_g}{2 q} - \frac{k T}{q} \ln \frac{N_A}{n_i} \tag{2.50}$$

For n-Si substrate, change the sign of both terms and change N_A for N_D in equation 2.50.

Fabrication Lph				
Sample	Material	Si excess (%)	Nominal thick. (nm)	Lum.
1	LPCVD SRO	5.1 ± 0.4	40	NA
2	"	4.0 ± 0.4	80	H/P
3	"	5.1 ± 0.4	40	NA
4	"	4.0 ± 0.4	80	NA

Table 2.12: Fabrication parameters of the samples Lph. The SRO layer is obtained by LPCVD. The silicon excesses are known from previous XPS measurements. In the luminescence column, H and P stand for *homogeneous* and *points*, respectively.

The number of states per unit area can be calculated approximately if we consider all the charge to be near the interface:

$$N_f = \frac{Q_f}{q\,A} \tag{2.51}$$

In section 6.2.6 of [52] we find the oxide trapped charge, Q_{ox} to be:

$$Q_{\text{ox}} = -C_{\text{ox}}\,\Delta V_{FB} \tag{2.52}$$

2.3.3.3 Results

In section 2.3.2 we saw that the emission through points is related to high current densities, whereas the homogeneous emission is related to low current densities. We will now show that the high current regime is associated with irregular CV curves, while the low current regime is associated with a CV curve resembling that of a typical MOS system. That reinforces the notion that the emission through points essentially comes with a broken device, which is difficult to control, while the homogeneous emission takes place in a regular MOS system whose electrical properties are well known. In particular, for samples with a well behaved CV curve, the wealth of information described in section 2.3.3.2 can be obtained. This can't be done in broken devices.

To show that, we will use the results of the fabrication Lph, which consists of SRO layers deposited by LPCVD with different silicon excesses, reported in [43]. The fabrication parameters of the samples is quoted in table 2.12.

This set of samples showed a switching behavior. The fresh devices exhibit high conduction. At some point during an IV stress, the current drops to a low conduction state. Further IV sweeps follow the low conduction state. The state cannot be reversed to high conduction by inverting the bias, as described by other authors [41]. This is probably related to the fact that our samples undergo a phase separation, whereas the samples reported in [41] exhibit a reversible behavior consistently as long as there is no phase separation in the SRO.
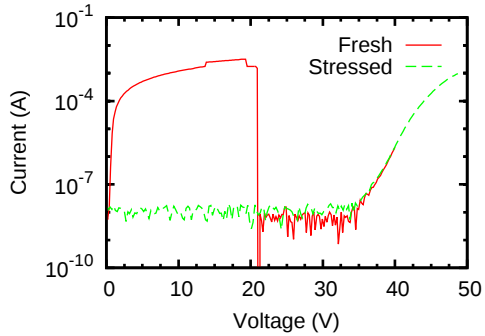
Figure 2.44: Typical IV of a fresh device in sample 2 of fabrication Lph (LPCVD SRO), and the IV of the same device after the first measurement drops it to a low conduction regime.
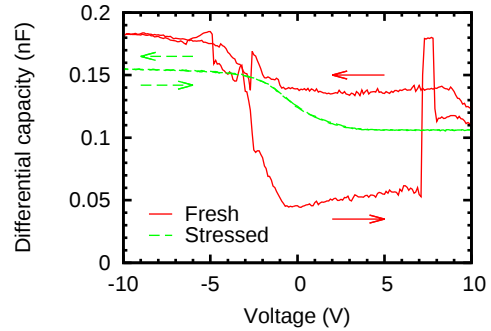
Figure 2.45: Typical CV hysteresis cycle of a fresh device and the CV of the same device once it is in a low conduction regime, for the same LPCVD sample of figure 2.44.

In figure 2.44 it can be seen how the IV characterization on a fresh device starts with a high conduction state that reaches the mA mark at about 10 V. At 20 V it abruptly falls to noise level, and it does not reach the μA mark until around 40 V. Once in this low conduction state, repeating the measurement shows how the device remains in the same low conduction state. This switching behavior has been attributed to the elimination of the conductive paths [43].

Performing a CV characterization on a device in the high conduction state yields a very noisy result from which very little information can be extracted reliably. On the other hand, performing the CV characterization on the same device after an IV stress has dropped it to the low conduction state reveals a well behaved curve from which a wealth of information can be reliably extracted. This is shown in figure 2.45.

All the other samples studied in this work, with the SRO obtained either by PECVD or silicon ion implantation, have not shown this switching behavior. They either show a high conduction state, in which case the CV characteristics are similar to those depicted for the fresh device in figure 2.45, or a low conduction state, in which case they show well behaved CV curves, as depicted in figure 2.46, corresponding to samples of fabrication Ph and Ih (PECVD and implanted, respectively). If the applied voltages are high enough, all of them suffer an intrinsic breakdown.

Figure 2.47 plots the JE characteristics of samples with homogeneous emission and low conduction, in particular fabrications Ph and Ih (see the fabrication parameters in tables 2.6 and 2.7, respectively). It is clear that increasing the silicon excess also increases the conductivity, but note that the implanted samples show, globally, a higher conductivity than the PECVD samples. For instance, the implanted sample with 12% silicon excess shows a higher conductivity than the PECVD sample with 16% silicon excess. The X axis represents the electric field instead of the voltage to account for the difference in thickness between samples. Note that even accounting
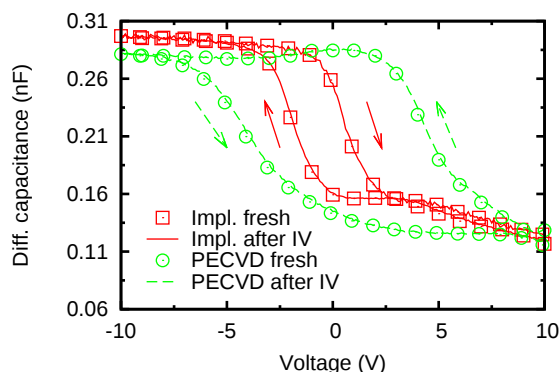
Figure 2.46: Typical hysteresis cycle in CV measurements for PECVD and implanted samples, all exhibiting low current (fabrications Ph and Ih).
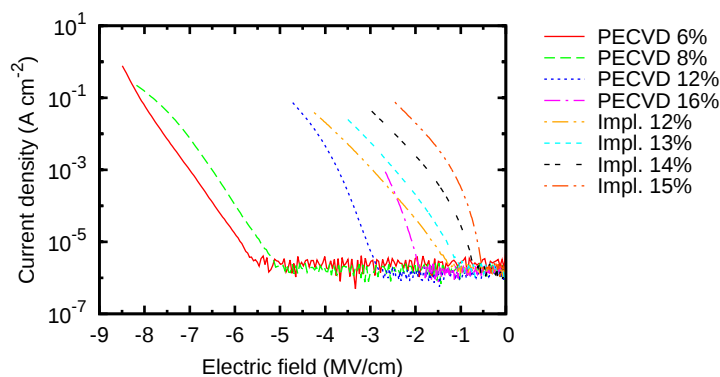
Figure 2.47: IV characteristics of PECVD and implanted samples with different silicon excess and low conduction (fabrications Ph and Ih).

for the error in the thickness, the relation of order between the conductivity of the samples is maintained.

However, what we are interested in is the injection mechanisms taking place in the active layer. Many models were tested (see section 2.3.3.1, but only three of them yielded promising results, namely Fowler-Nordheim tunneling, trap assisted tunneling and Poole-Frenkel emission.

Figure 2.48 shows a Poole-Frenkel plot for samples Ih and Ph. It is clear that linear regions can be identified in all of them, but it is still necessary to check if the parameters that result from a Poole-Frenkel fit make sense.

The relative permittivity of the active layer, as extracted from the CV curves and from a Poole-Frenkel fit over the linear regions of figure 2.48 is shown in figures 2.49 and 2.50 for the implanted samples Ih and the PECVD samples Ph, respectively. The error bars arise from the uncertainty in the fit and in the silicon excess.

The PECVD samples get values of relative permittivity one order of magnitude
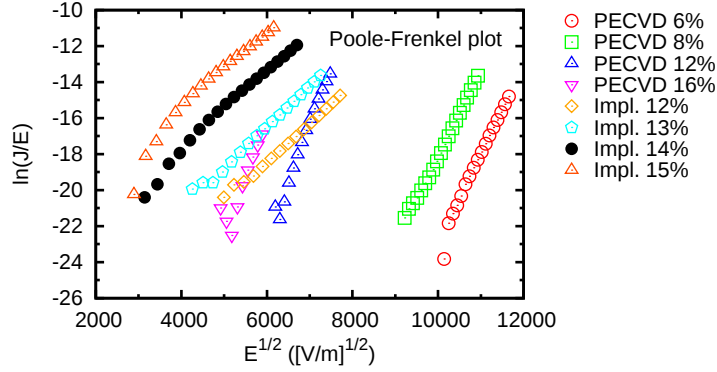
Figure 2.48: Poole-Frenkel plot of the PECVD and implanted samples of fabrications Ph and Ih, respectively, showing low conduction.
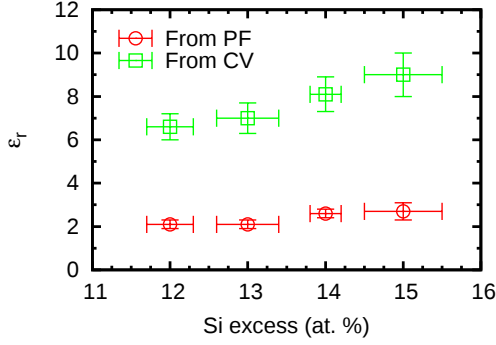
Figure 2.49: Relative permittivity of the implanted samples Ih, as extracted from the Poole-Frenkel fits and the CV measurements.
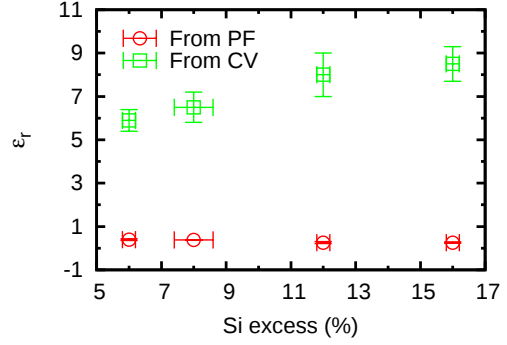
Figure 2.50: Relative permittivity of the PECVD samples Ph, as extracted from the Poole-Frenkel fits and the CV measurements.

lower than what we expect based on the results of CV measurements. The implanted samples also get values lower than expected, but the order of magnitude is correct. That leads us to believe that Poole-Frenkel does not apply to PECVD samples, but there could definitely be a Poole-Frenkel component of the conduction in implanted samples.

Figures 2.51 and 2.52 present a Fowler-Nordheim and trap assisted tunneling plot, respectively, of the same samples. All of the PECVD samples show linear regions in both models. Regarding the implanted samples, only those with 14% and 15% silicon excess show a clear linear region at low electric fields. It is possible that the samples with 12% and 13% silicon excess also present that linear region but it is below our measurement range. In any case, we could not identify a clearly linear region for those two samples. Fitting the data allows us to calculate pairs (m_r, ϕ_b) according to equations 2.29 and 2.33.
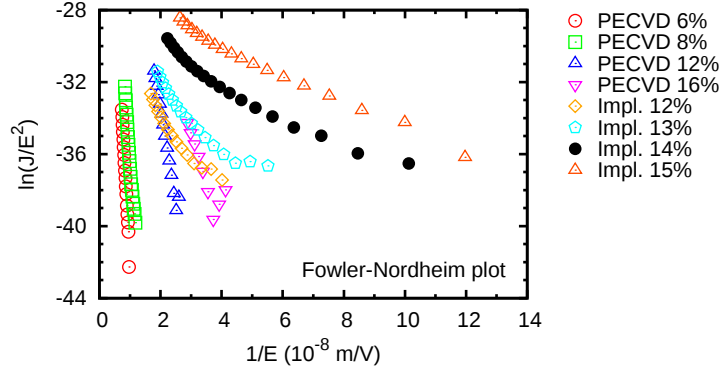
Figure 2.51: Fowler-Nordheim plot of the PECVD and implanted samples, fabrications Ph and Ih, respectively, showing low conduction.
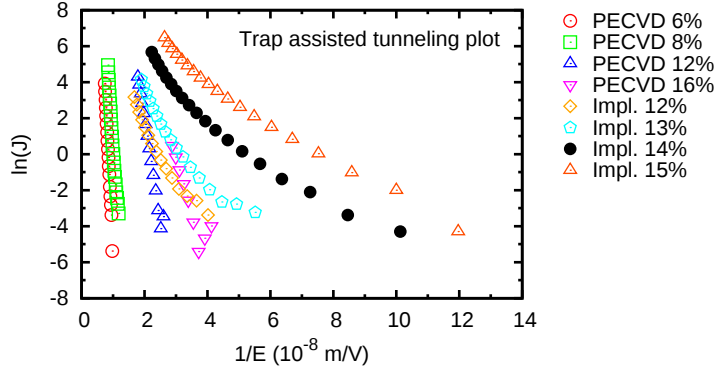
Figure 2.52: Trap assisted tunneling plot of the PECVD and implanted samples, fabrications Ph and Ih, showing low conduction.

In figures 2.53 and 2.54 we plot, for each sample, the functions that give (m_r, ϕ_b) pairs that fit the data. Of course there are infinite such pairs, so we need to somehow define bounds to the accepted values. According to the literature, the effective masses of an electron in silicon are [55] $m_r = 0.98$ (longitudinal relative effective mass) and $m_r = 0.19$ (transverse relative effective mass). For an electron in thermal SiO_2, $m_r = 0.42$ [56]. Regarding holes in silicon, the effective masses are $m_r = 0.16$ (light holes) and $m_r = 0.49$ (heavy holes) [55], whereas in SiO_2 $m_r = 0.58$ [57]. Since the studied material is composed by silicon nanoclusters embedded in SiO_2, it seems reasonable to only accept values of m_r between 0.98 and 0.16. This range of accepted values of m_r is indicated in figures 2.53 and 2.54 by two horizontal dashed lines, and it translates to a range of accepted values of ϕ_b through the plotted functions.

Table 2.13 shows the range of values of ϕ_b that fall within the accepted range of m_r, with the error in the fit factored in. Considering that $\phi_b = 2.8$ eV in thermal SiO_2 [58], the ranges for the PECVD seem appropriate, but those for the implanted
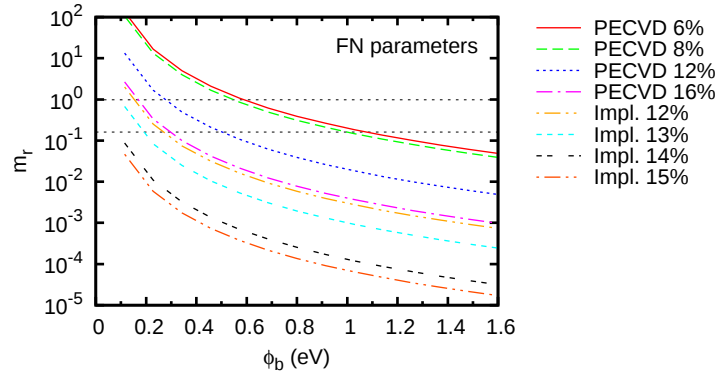
Figure 2.53: Each point in the curves represents a pair ϕ_b, m_r that fits the corresponding FN plot. m_r has been restricted to the range [0.16, 0.98], marked in the graph with the horizontal dashed lines.
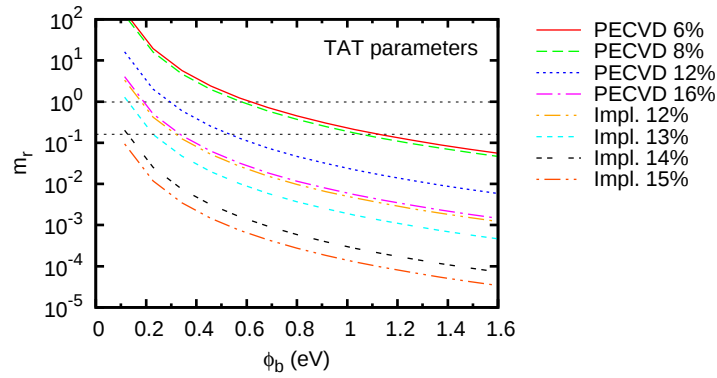
Figure 2.54: Each point in the curves represents a pair ϕ_b, m_r that fits the corresponding TAT plot. m_r has been restricted to the range [0.16, 0.98], marked in the graph with the horizontal dashed lines.

samples are too low.

It is worth noting that the nonlinear region in the Fowler-Nordheim and trap assisted tunneling plots for the implanted samples are precisely the linear regions in the Poole-Frenkel plot. Therefore, it seems that the low field region in the implanted samples has some contribution from trap assisted tunneling or Fowler-Nordheim, which transitions into Poole-Frenkel emission at higher fields.

In conclusion, it is not possible to unequivocally assign a precise injection mechanism to these samples, and a combination of several methods is most likely. Other authors have arrived at similar conclusions [59]. However, it seems safe to assume that Fowler-Nordheim tunneling or trap assisted tunneling play a significant role in the PECVD samples. As for the implanted samples, Fowler-Nordheim or trap assisted tunneling could be playing a role at low fields, while Poole-Frenkel seems

Type	Si exc	ϕ_b^{FN}	ϕ_b^{TAT}
	(%)	(eV)	(ev)
Impl.	14	[0.03, 0.11]	[0.04, 0.15]
Impl.	15	[0.02, 0.09]	[0.04, 0.11]
PECVD	6 ± 0.2	[0.52, 1.22]	[0, 56, 1.22]
PECVD	8 ± 0.6	[0.45, 1.12]	[0.48, 1.19]
PECVD	12 ± 0.2	[0.23, 0.55]	[0.25, 0.58]
PECVD	16 ± 0.6	[0.14, 0.34]	[0.15, 0.37]

Table 2.13: Range of possible values of ϕ_b for the PECVD and ion-implanted samples, fabrications Ph and Ih, respectively, extracted from FN and TAT models after limiting the range of m_r to [0.16, 0.98]. The missing implanted samples did not adjust to the model.

most significant at high fields.

2.4 Conclusion

In this chapter we have studied silicon based, CMOS compatible light sources. Luminescent layers of SRO, SRN and SRO/SRN bilayers have been fabricated. PL and structural characterization has been carried out. Complete electroluminescent devices have been fabricated with those layers, and they have been optically and electrically characterized.

PL and EL characterizations have been carried out in two setups built from scratch.

Different kinds of emission, namely emission through points, edge emission and homogeneous emission, have been identified and correlated with the electrical and optical properties of the active layers, in particular with their efficiency and conductivity, and it has been concluded that the homogeneous emission is related to better efficiencies. A simple model has been proposed to explain the coexistence of different kinds of emission.

Furthermore, the effect of a nitride layer on the luminescence spectrum has been discussed. It has been concluded that the inclusion of a nitride layer helps in achieving a wider emission spectrum that is potentially easier to modulate, although more work is needed in order to find possible ways to control the contribution of each material to the total luminescence.

Finally, the high and low conduction states, associated with the emission through points and homogeneous emission, respectively, have been related to the CV behavior. The injection mechanisms in PECVD and ion implanted samples have been compared and it has been concluded that, even though no single injection model can fit all the data, Fowler-Nordheim tunneling or trap assisted tunneling play a significant role in PECVD devices, whereas in implanted samples Fowler-Nordheim

or trap assisted tunneling are likely to be playing a role at low fields, while Poole-Frenkel takes over at high fields.

During this chapter, many features of the studied spectra have been attributed to interference effects in the multilayer stack. In the following chapter we will look into this in detail and will present a model to quantitatively calculate those effects.

Optical effects

3.1 Introduction

When studying the PL and EL of systems with several stacked layers, the measured spectra will differ from the intrinsically emitted one due to coherent interference and absorption. This is specially noticeable when one of the layers is of a metallic nature, i.e has nonzero extinction coefficient. The effects of the stack on the measured spectrum can affect the position and relative intensities of the observed emission peaks, thus limiting our ability to draw conclusions regarding the mechanisms that originate the luminescence.

As an example, consider two hypothetical emission spectra such as the ones represented in figure 3.1(a). Both spectra are identical Gaussian functions except for a small but clearly detectable shift. If that was the intrinsic emission spectra of two different SRO samples, we could relate that shift to emission due to different traps, or maybe to emission from silicon nanoclusters of different sizes, which could in turn be related to different silicon excesses in the active layer. Furthermore, perhaps a small shoulder could be identified, hinting at another secondary emission mechanism.

Now suppose that the actual emission is modified by the coherent interference due to multiple reflections at the interfaces of the multilayer stack, with the interference being something similar to the dotted line of figure 3.1(a). In that case, what we would observe is not the spectra of figure 3.1(a), but that of figure 3.1(b) instead, that is, the original spectra modified by the interference pattern. Even though the two spectra are still clearly different, the relation of their differences to different emission centers is completely obscured now, not to mention that any hypothetical shoulder in the original spectra would have been completely washed out. Even though we could easily convince ourselves that the peaks observed in the modified spectra are due to interference effects, without a quantitative analysis of those effects it is not possible to draw any credible conclusion as far as the origin of the luminescence is concerned.

In this chapter we present a study of the effect that the multilayered systems
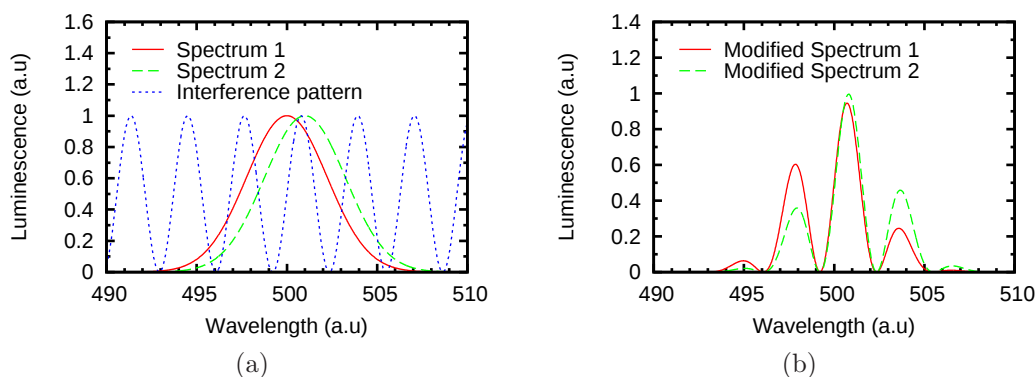
Figure 3.1: (a) Two identical hypothetical spectra with a small but easily detectable shift, and an also hypothetical interference pattern. (b) The same spectra after the interference pattern has been applied.

used in our experiments have on the PL and EL of the active layer. In order to gain quantitative insight into the interference effects in those multilayered systems, we will implement a method proposed by O. H. Crawford in 1988 [60], and we will extend that method in order to account for a continuous distribution of radiative centers in the system.

In section 3.2 we will discuss the theory related to the method proposed by Crawford. For completeness, section 3.2.1 briefly reviews the basic concepts of the electromagnetic theory that are fundamental to the understanding of the method, which will be discussed later in section 3.2.2. The reader well acquainted with the basic equations of the electromagnetic field, the polarization of light, the laws of reflection and refraction and the formalism of the characteristic matrix to deal with the propagation of light in multilayered systems, can safely skip that part and jump right into section 3.2.2. In 3.2.2.1 the basic layout of the method will be briefly presented. In 3.2.2.2 we will discuss our extension of the method in order to apply it successfully to our samples.

Section 3.3 briefly discusses our Python implementation of the Crawford method.

Section 3.4 presents some general results for multilayer systems such as the ones used in our work. The effect of different variables is discussed.

Finally, section 3.5 shows how we can extract information about the intrinsic emission of a sample from the experimentally measured spectrum.

A tutorial on how to use the program used for the simulations can be found in appendix B.

3.2   A bit of theory

The first part of this section presents the fundamental concepts of the electromagnetic theory of light. They are required in order to understand the Crawford method,

which will be explained in the second part of the section.

3.2.1 Electromagnetic theory

A detailed treatment of the principles of the electromagnetic theory of light can be found in [61].

3.2.1.1 The electromagnetic field

The presence of electrical charges in a system induces an *electromagnetic field* that will be represented by the vectors $\vec{E}$ and $\vec{B}$, the *electric vector field* and *magnetic induction vector field*, respectively. Additionally, the *electric current density* $\vec{j}$, the *electric displacement* $\vec{D}$ and the *magnetic vector* $\vec{H}$ are introduced to account for the effect of the electromagnetic field on matter.

Maxwell's equations describe the relations between those five vectors. In the Gaussian system of units, they take the form:

$$\vec{\nabla} \times \vec{H} - \frac{1}{c}\frac{\partial \vec{D}}{\partial t} = \frac{4\pi}{c}\vec{j} \tag{3.1}$$

$$\vec{\nabla} \times \vec{E} + \frac{1}{c}\frac{\partial \vec{B}}{\partial t} = 0 \tag{3.2}$$

$$\vec{\nabla} \cdot \vec{D} = 4\pi\rho \tag{3.3}$$

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{3.4}$$

Equations (3.1) and (3.2) are vector equations and they relate the time and space derivatives of the vector fields. Equations (3.3) and (3.4) are scalar equations. (3.3) defines the *electric charge density* ρ and (3.4) states that no free magnetic monopoles exist. Maxwell's equations hold in any region of space whose physical properties are continuous.

Given a distribution of currents and charges, an additional set of equations are needed in order to uniquely determine the field vectors with equation (3.1)–(3.4). These are known as the *material equations* or *constitutive relations* and they describe the behavior of substances under the influence of the field. In general they are complicated, but for time-harmonic fields[1], bodies at rest or in slow motion relative to each other, and isotropic materials, they take the form:

$$\vec{j} = \sigma\vec{E} \tag{3.5}$$

$$\vec{D} = \varepsilon\vec{E} \tag{3.6}$$

$$\vec{B} = \mu\vec{H} \tag{3.7}$$

---

[1]A general time-harmonic real vector wave is a solution of the wave equation whose components take the form

$$V_i(\vec{r},\ t) = v_i(\vec{r})\cos[\omega t - g_i(\vec{r})], \quad i = x,\ y,\ z$$

Equation (3.5) is the differential form of Ohm's law and it defines the *conductivity* σ. Materials with values of σ negligibly small are called *insulators* or *dielectrics*. Materials with σ not negligibly small are called *conductors*.

Similarly, equation (3.6) defines the *dielectric constant* or *permittivity* ε and equation (3.7) defines the *magnetic permeability* μ. For most materials, including the materials that will be studied in the present work, $\mu \approx 1$. If μ is significantly different than unity then the material is said to be *magnetic*. More precisely, in *paramagnetic* materials $\mu > 1$ whereas $\mu < 1$ for *diamagnetic* materials.

From now on μ will be considered 1 for all the materials of interest in this work. Therefore, dielectric materials will be completely characterized by their permittivity ε, whereas conductive materials will be characterized by their permittivity ε and their conductivity σ.

It is worth noting that, while equations (3.5)–(3.7) are valid in many situations, there are limitations. Very strong fields would require powers of the fields higher than 1 in the right hand side of equations (3.5)–(3.7). The parameters σ, ε and μ may depend on the field strengths, and in some substances the values of the fields at any time depend on the previous values of the fields (materials that exhibit *hysteresis*).

3.2.1.2   The boundary conditions

In section 3.2.1.1 Maxwell's equations were stated for regions of space whose physical properties are continuous. However, many practical situations involve regions of space where physical properties change abruptly. This is the case of the layered systems that will be studied in the present work.

Let T be the surface of discontinuity between two mediums. At a given point P in T, let $\vec{n_{12}}$ be the unit normal pointing from the first into the second medium. $\vec{B^{(1)}}$ and $\vec{B^{(2)}}$ are the values of $\vec{B}$ right next to P in the first and second mediums respectively, and let $\vec{D^{(1)}}$ and $\vec{D^{(2)}}$ be the values of $\vec{D}$ right next to P in the first and second mediums respectively. In the very close vicinity of P all four quantities $\vec{B^{(1)}}$, $\vec{B^{(2)}}$, $\vec{D^{(1)}}$ and $\vec{D^{(2)}}$ can be considered constant. Under these conditions it can be easily shown [61] that the following *boundary conditions* are satisfied:

$$\vec{n_{12}} \cdot (\vec{B^{(2)}} - \vec{B^{(1)}}) = 0 \tag{3.8}$$

$$\vec{n_{12}} \times (\vec{H^{(2)}} - \vec{H^{(1)}}) = \frac{4\pi}{c}\hat{\vec{j}} \tag{3.9}$$

$$\vec{n_{12}} \cdot (\vec{D^{(2)}} - \vec{D^{(1)}}) = 4\pi\hat{\rho} \tag{3.10}$$

$$\vec{n_{12}} \times (\vec{E^{(2)}} - \vec{E^{(1)}}) = 0 \tag{3.11}$$

Equation (3.8) states that the normal component of the magnetic induction is continuous across the surface, while equation (3.9) states that the tangential component of the magnetic vector is discontinuous across the surface, its discontinuity

being equal to $4\pi\hat{\vec{j}}/c$, where $\hat{\vec{j}}$ is the surface current density present in the surface T between the two mediums.

Similarly, equation (3.10) states that the normal component of the electric displacement presents a discontinuity across the surface, its value being $4\pi\hat{\rho}$, where $\hat{\rho}$ is the surface charge density in the surface. Equation (3.11) states that the tangential component of the electric vector is continuous across the surface.

3.2.1.3 The energy of the electromagnetic field

In this section it will be shown how the light intensity can be interpreted as the energy flux of the electromagnetic field.

Using equations (3.1) and (3.2) and the material equations (3.5)–(3.7), it can be shown that the following equation holds:

$$\frac{\mathrm{d}W}{\mathrm{d}t} = -\frac{\delta A}{\delta t} - Q - \int_S \vec{S} \cdot \vec{n} \,\mathrm{d}S \tag{3.12}$$

Here, δA is the work done by the field on the electric charges of the system and we have used the following definitions:

$$W = \int_V (w_e + w_m) \,\mathrm{d}V \tag{3.13}$$

$$w_e = \frac{1}{8\pi}\vec{E} \cdot \vec{D} \tag{3.14}$$

$$w_m = \frac{1}{8\pi}\vec{H} \cdot \vec{B} \tag{3.15}$$

$$Q = \int_V \vec{j_c} \cdot \vec{E} \,\mathrm{d}V = \int_V \sigma \vec{E}^2 \,\mathrm{d}V \tag{3.16}$$

$$\vec{S} = \frac{c}{4\pi}(\vec{E} \times H) \tag{3.17}$$

In the previous equations, the integration volume V is enclosed by the surface S.

If the material is a nonconductor ($\sigma = 0$), then $Q = 0$. Furthermore, if we consider the boundary of the volume to be so far away that the field can be considered zero, then we have $\mathrm{d}W/\mathrm{d}t = \delta A/\delta t$, which justifies the interpretation of W as the electromagnetic energy.

In the more general case, Q is the amount of energy dissipated in heat (Joule's dissipation) and the integral in (3.12) represents the energy that crosses the boundary of the volume per second. Therefore equation (3.12) states that the variation of the energy stored in the electromagnetic field equals the sum of three terms: the work done on the charges of the system by the field, the resistive dissipation, and the flow of energy across the boundaries of the system.

The vector $\vec{S}$ is called the *Poynting vector* and is interpreted as an energy flow (energy per second and unit area). Note that in general $\vec{S}$ only has physical significance when integrated over a closed surface. However, in a nonconducting system ($Q = 0$) where no work is done ($\delta A = 0$), equation (3.12) can be written as a continuity equation:

$$\frac{\partial w}{\partial t} + \vec{\nabla} \cdot \vec{S} = 0 \tag{3.18}$$

In this manner, the interpretation of the magnitude of $\vec{S}$ as the light intensity and its direction as the direction of propagation of light is reinforced.

3.2.1.4   The wave equation

Using cleverly Maxwell's equations (3.1)–(3.4) and the material equations (3.5)–(3.7), and making use of some identities involving $\vec{\nabla}$, it can be easily shown [61] that, for $\vec{j} = 0$ and $\rho = 0$, and a homogeneous medium ($\vec{\nabla}\varepsilon = \vec{\nabla}\mu = 0$) the electromagnetic field satisfies the wave equation:

$$\vec{\nabla}^2 \vec{E} - \frac{\varepsilon\mu}{c^2}\frac{\partial E}{\partial t} = 0 \tag{3.19}$$

$$\vec{\nabla}^2 \vec{H} - \frac{\varepsilon\mu}{c^2}\frac{\partial H}{\partial t} = 0 \tag{3.20}$$

That means that the electromagnetic field propagates as an *electromagnetic wave* with velocity:

$$v = \frac{c}{\sqrt{\varepsilon\mu}} \tag{3.21}$$

The quantity $n = \sqrt{\varepsilon\mu}$ will be called the *refractive index* of the material where light is propagating.

3.2.1.5   Scalar waves

Each rectangular component of $\vec{E}$ and $\vec{H}$ satisfies the homogeneous wave equation. We shall use $V(\vec{r},\, t)$ as any of the rectangular components.

$$\vec{\nabla}^2 V - \frac{\varepsilon\mu}{c^2}\frac{\partial V}{\partial t} = 0 \tag{3.22}$$

The simplest solution of the wave equation is the *plane wave*. If $\vec{s}$ is a unit vector in a fixed direction, then a solution $V = V(\vec{r} \cdot \vec{s},\, t)$ is a said to be a plane wave because the value of V is constant over each of the planes $\vec{r} \cdot \vec{s} = \zeta$, where ζ is a constant. If we choose a new coordinate system (ξ, η, ζ) with ζ in the direction of $\vec{s}$ then $\vec{\nabla}^2 = \partial^2 V/\partial\zeta^2$ and the wave equation depends only on one spatial variable:

$$\frac{\partial^2 V}{\partial\zeta^2} - \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} = 0 \tag{3.23}$$

Setting $\zeta - vt = p$, $\zeta + vt = q$, equation (3.23) can be written as:

$$\frac{\partial^2 V}{\partial p \partial q} = 0 \tag{3.24}$$

The general solution to the last equation is:

$$V = V_1(\vec{r} \cdot \vec{s} - vt) + V_2(\vec{r} \cdot \vec{s} + vt) \tag{3.25}$$

V_1 is clearly a propagation traveling with velocity v along the positive ζ direction, since the argument of V_1 is unchanged when changed from (ζ, t) to $(\zeta + v\tau, t + \tau)$. Similarly, V_2 is a propagation traveling with velocity v along the negative ζ direction.

Whatever the solution $V(\vec{r}, t)$, if we fix $\vec{r} = \vec{r_0}$ then $V(\vec{r_0}, t) = F(t)$ is a function of the time only. A particularly interesting case is when F is *harmonic*:

$$F(t) = a \cos(\omega t + \delta) \tag{3.26}$$

The constant $a > 0$ is the *amplitude* and $\omega t + \delta$ is the *phase*. The constant ω is the *angular frequency*. We define $\nu = \omega/2\pi = 1/T$ to be the *frequency*, and therefore T is the *period* of the vibration because F does not change when t is replaced by $t + T$.

A *harmonic plane wave* can be written as:

$$V(\vec{r}, t) = a \cos\left[\omega\left(t - \frac{\vec{r} \cdot \vec{s}}{v}\right) + \delta\right] \tag{3.27}$$

Clearly, the value of V is the same in all points of the plane $\vec{r} \cdot \vec{s} = $ constant and is harmonic. Moreover, V remains unchanged when $\vec{r} \cdot \vec{s}$ is replaced by $\vec{r} \cdot \vec{s} + \lambda$, with λ being the *wavelength* of the wave:

$$\lambda = v\frac{2\pi}{\omega} = vT \tag{3.28}$$

Notice that a wave of a given frequency will have a different wavelength depending on its velocity, which we saw depends on the material (see equation (3.21)). Often it will be convenient to use the *reduced wavelength*, which is the wavelength that a wave with the same frequency would have in vacuum:

$$\lambda_0 = cT = n\lambda \tag{3.29}$$

Another useful quantity is the *wave number*:

$$\kappa = \frac{1}{\lambda_0} = \frac{\nu}{c} \tag{3.30}$$

Related to the wave number, we have the *wave vectors* $\vec{k}$ and $\vec{k_0}$, whose direction is along $\vec{s}$ and their magnitude is:

$$k_0 = 2\pi\kappa = \frac{2\pi}{\lambda_0} = \frac{\omega}{c} \tag{3.31}$$

$$k = nk_0 = \frac{2\pi}{\lambda} = \frac{n\omega}{c} = \frac{\omega}{v} \tag{3.32}$$

From now on the term *wave number* will generally be used to refer to k_0 or k rather than κ unless stated otherwise.

A general time-harmonic wave can be written in the form:

$$V(\vec{r},\ t) = a(\vec{r})\cos[\omega t - g(\vec{r})] \tag{3.33}$$

The surfaces where $g(\vec{r}) = $ constant are said to be *wave surfaces*, i.e. they form a surface where the wave has the same phase. In the case of a plane wave it was already shown that $g(\vec{r}) = k \cdot \vec{r} \cdot \vec{s} + \delta = \vec{k} \cdot \vec{r} + \delta$.

In general, it is more convenient to write the wave as:

$$V(\vec{r},\ t) = \mathcal{R}[a(\vec{r})e^{ig(\vec{r})}e^{-i\omega t}] \tag{3.34}$$

Here $\mathcal{R}$ denotes the real part of the complex argument. Now define the *complex amplitude* as:

$$U(\vec{r}) = a(\vec{r})e^{ig(\vec{r})} \tag{3.35}$$

And equation (3.34) can be rewritten:

$$V(\vec{r},\ t) = \mathcal{R}[U(\vec{r})e^{-i\omega t}] \tag{3.36}$$

In this manner we separate the time and space dependencies of the wave. Substituting (3.36) into (3.22) we find that U must satisfy the equation:

$$\vec{\nabla}^2 U + n^2 k_0^2 U = 0 \tag{3.37}$$

For linear operations on V, the $\mathcal{R}$ symbol can be omitted and all operations applied to the complex wave. The real part of the result will be the quantity with physical meaning. That cannot be done with nonlinear operations such as squaring. In those cases, the real part has to be taken first and the operations applied on the real part. This restriction is not necessary when we are concerned with the averages of quadratic expressions of the field.

3.2.1.6   Vector waves

For a plane wave with $\vec{E} = \vec{E}(\vec{r} \cdot \vec{s} - vt)$ and $\vec{H} = \vec{H}(\vec{r} \cdot \vec{s} - vt)$ it can be shown [61] that the fields satisfy the following condition:

$$\vec{E} = -\sqrt{\frac{\mu}{\varepsilon}}\ \vec{s} \times \vec{H} \tag{3.38}$$

$$\vec{H} = \sqrt{\frac{\varepsilon}{\mu}}\ \vec{s} \times \vec{E} \tag{3.39}$$

Scalar multiplication of equations (3.38) and (3.39) with $\vec{s}$ gives:

$$\vec{E} \cdot \vec{s} = \vec{H} \cdot \vec{s} = 0 \tag{3.40}$$

Equations (3.38)–(3.40) tell us that $\vec{E}$, $\vec{H}$ and $\vec{s}$ form a right-handed orthogonal triplet of vectors and:

$$\sqrt{\mu}H = \sqrt{\varepsilon}E \qquad (3.41)$$

Here $E = |\vec{E}|$ and $H = |\vec{H}|$.

To find the energy per unit time that crosses a unit area perpendicular to the direction of propagation of the wave, consider a cylinder whose axis is parallel to the direction of propagation and whose base has area 1. Then the energy crossing the base of the cylinder in one second is the energy stored in a slice of the cylinder of length v, where v is the velocity of the wave. Therefore, the energy should be vw, where w is the energy density in the cylinder. Indeed, recalling equations (3.14), (3.15) and (3.17) and using (3.41) we have:

$$w = \frac{\varepsilon}{4\pi}E^2 = \frac{\mu}{4\pi}H^2 \qquad (3.42)$$

$$\vec{S} = \frac{c}{4\pi}EH\vec{s} = \frac{c}{4\pi}\sqrt{\frac{\varepsilon}{\mu}}E^2\vec{s} = \frac{c}{4\pi}\sqrt{\frac{\mu}{\varepsilon}}H^2\vec{s} \qquad (3.43)$$

Or, in a more compact form:

$$\vec{S} = \frac{c}{\sqrt{\varepsilon\mu}}w\vec{s} = vw\vec{s} \qquad (3.44)$$

That is the result that we expected and furthermore, it reinforces the interpretation of the Poynting vector as the flow of energy associated to the electromagnetic field, as we discussed in section 3.2.1.3.

3.2.1.7 The polarization of light

Let's focus now on a time-harmonic plane wave. Each component of the fields $\vec{E}$ and $\vec{H}$ has the form $a\cos(\tau + \delta) = \mathcal{R}[a\exp(-i(\tau + \delta)]$, with $a > 0$ and τ the non constant part of the phase, $\tau = \omega t - \vec{k} \cdot \vec{r}$.

For convenience we choose the z-axis in the direction of $\vec{s}$, the direction of propagation of the wave. Due to the transversality of the field, equations (3.38)–(3.40), $\vec{E}$ and $\vec{H}$ only have x and y components. In particular:

$$E_x = a_1 \cos(\tau + \delta_1) \qquad (3.45)$$
$$E_y = a_2 \cos(\tau + \delta_2) \qquad (3.46)$$

The points whose x and y coordinates are E_x and E_y respectively is the end point of the electric vector. The *polarization* of the wave is related to the curve which the end point of the electric vector describes in space. In the general case that curve will be an ellipse [61], and then the wave is said to be *elliptically polarized*. It is easily shown [61] that if the wave associated to the electric vector is elliptically polarized, then the wave associated with the magnetic vector is also elliptically polarized.

Two particular cases are of interest. When the ellipse degenerates into a circle the light is said to be *circularly polarized*. That happens when $a_1 = a_2$ and $\delta_2 - \delta_1 = m\pi/2$ for $m = \pm 1,\ \pm 3,\ \pm 5, \ldots$.

When the ellipse degenerates into a straight line the light is said to be *linearly polarized*, and the condition for that is $\delta_2 - \delta_1 = m\pi$ for $m = 0, \pm1, \pm2, \ldots$.

The linearly polarized wave is of particular importance because a wave of general polarization as described in equations (3.45) and (3.46) can always be interpreted as the sum of two linearly polarized waves, one described by equation (3.45) and the other described by equation (3.46).

That is all we need to know regarding the polarization of light. An in-depth treatment of the subject can be found in reference [61].

3.2.1.8 The laws of reflection and refraction

Consider a plane wave falling onto an interface between two different homogeneous and isotropic mediums. In order to satisfy the boundary conditions (3.8)–(3.11) a transmitted wave proceeding into the second medium and a reflected wave going back into the first medium have to be postulated. We will assume that both the transmitted and reflected wave are also plane.

For a plane wave, if the time behavior of the wave $F(t)$ is known at one point, then any other point separated from the first by a vector $\vec{r}$ will have the same behavior except for a phase delay, $F[t - (\vec{r} \cdot \vec{s})/v]$. Therefore, at any point in the boundary plane $z = 0$ where the three waves coincide, the time behavior must be the same. Using this fact it is easily demonstrated [61] that the following must be true:

$$\frac{\sin\theta_i}{v_1} = \frac{\sin\theta_r}{v_1} = \frac{\sin\theta_t}{v_2} \tag{3.47}$$

Here v_1 and v_2 are the velocities of propagation in the first and second mediums, respectively.

The *plane of incidence* is defined as the plane that contains the direction of propagation of the incident wave, $\vec{s}^{(i)}$, and the normal to the boundary (see figure 3.2). It can be demonstrated that the direction of propagation of the reflected and transmitted waves, $\vec{s}^{(r)}$ and $\vec{s}^{(t)}$ respectively, must also lie in this plane, that we will take to be the xz plane without loss of generality.

It is clear from equation (3.47) that $\sin\theta_i = \sin\theta_r$ and, since $\vec{s}_i$ and $\vec{s}_r$ are unit vectors and $\cos\theta_r$ has a sign opposite to $\cos\theta_i$ it must be $\cos\theta_r = -\cos\theta_i$. Those two relations combined give (see figure 3.2):

$$\theta_r = \pi - \theta_i \tag{3.48}$$

This last relation and the fact that $\vec{s}^{(r)}$ is in the plane of incidence constitute the *law of reflection*.

Now, recalling that $n = \sqrt{\varepsilon\mu}$ and using (3.47) we find:

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{v_1}{v_2} = \sqrt{\frac{\varepsilon_2\mu_2}{\varepsilon_1\mu_1}} = \frac{n_2}{n_1} = n_{12} \tag{3.49}$$

The last relation and the fact that $\vec{s}^{(t)}$ is in the plane of incidence constitute the *law of refraction* or *Snell's law*.
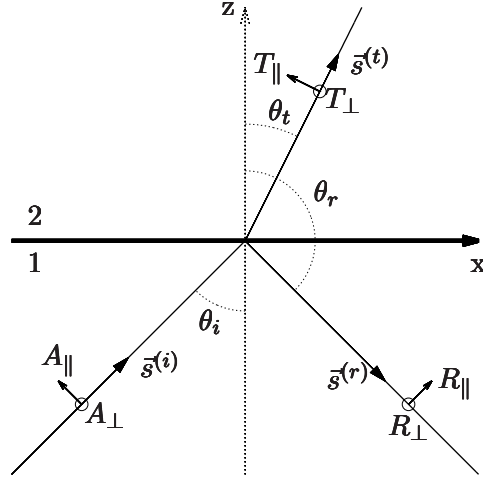
Figure 3.2: Schematic view of the system studied in section 3.2.1.8. The paper is the plane of incidence and the $\perp$ vectors point outwards perpendicular to the plane of incidence.

Now let's assume the mediums to be nonconducting and $\mu_1 = \mu_2 = 1$. With those assumptions we want to find how the amplitudes A and R of the reflected and transmitted waves, respectively, relate to the amplitude A of the incident wave. We will take A as a complex number whose phase is the constant part in the expression for the phase of the wave.

Each wave can be decomposed into two components, one parallel to the plane of incidence (subscript $\parallel$) and another perpendicular to the plane of incidence (subscript $\perp$). The positive directions for each vector are indicated in figure 3.2. With all those provisions it is not difficult to arrive to the well known *Fresnel formulae*:

$$T_\parallel = \frac{2n_1 \cos\theta_i}{n_2 \cos\theta_i + n_1 \cos\theta_t} A_\parallel \tag{3.50}$$

$$T_\perp = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_t} A_\perp \tag{3.51}$$

$$R_\parallel = \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} A_\parallel \tag{3.52}$$

$$R_\perp = \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} A_\perp \tag{3.53}$$

73

Using the law of refraction they can be rewritten as:

$$T_\parallel = \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t) \cos(\theta_i - \theta_t)} A_\parallel \tag{3.54}$$

$$T_\perp = \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t)} A_\perp \tag{3.55}$$

$$R_\parallel = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} A_\parallel \tag{3.56}$$

$$R_\perp = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} A_\perp \tag{3.57}$$

The coefficients between the reflected and incident amplitudes are the *reflection coefficients* $r_\parallel$ and $r_\perp$, whereas the coefficients between the transmitted and incident amplitudes are the *transmission coefficients* $t_\parallel$ and $t_\perp$.

Note that parallel and perpendicular components are not mixed in the Fresnel equations, which means that each one can be treated separately. The $\parallel$ wave (electric vector parallel to the plane of incidence) will be said to have *transversal magnetic (TM) polarization*, whereas the $\perp$ wave (electric vector perpendicular to the plane of incidence) will be said to have *transversal electric (TE) polarization*. Therefore, a wave can always be decomposed into its TE and TM components and each component treated separately.

If the angles θ_i, θ_r and θ_t are all real (the case when they are complex will be studied later), then it is clear that the coefficients between the reflected and transmitted amplitudes, as well as their corresponding incident amplitudes in the Fresnel formulae, are real numbers. In particular, the coefficients between the transmitted and incident amplitudes are positive real numbers, and therefore the phase of the complex numbers $T_\parallel$, $T_\perp$, $A_\parallel$ and $A_\perp$ is the same, and the phase of the incident and transmitted wave is the same too.

For the reflected wave the phase will depend on the relative values of θ_i and θ_t. According to the law of refraction, if $\varepsilon_2 > \varepsilon_1$ then $\theta_t < \theta_i$. In that situation, $R_\perp$ and $A_\perp$ have different signs and therefore their phases differ by π. In the case of $R_\parallel$, its phase differs by π from the phase of $A_\parallel$ when $\theta_i + \theta_t > \pi/2$. A similar analysis can be done for the case $\varepsilon_2 < \varepsilon_1$.

Now let's see how the energy of the incident wave is distributed to the transmitted and reflected waves. The energy of the incident wave in a unit area of the interface between the mediums is, according to the discussion about the Poynting vector in section 3.2.1.6 and using equation (3.43):

$$J^{(i)} = S^{(i)} \cos \theta_i = \frac{cn_1}{4\pi} |A|^2 \cos \theta_i \tag{3.58}$$

Similarly, the energy of the reflected and transmitted waves leaving a unit area of the interface is:

$$J^{(r)} = S^{(r)} \cos \theta_i = \frac{cn_1}{4\pi} |R|^2 \cos \theta_i \tag{3.59}$$

$$J^{(t)} = S^{(t)} \cos \theta_t = \frac{cn_2}{4\pi} |T|^2 \cos \theta_t \tag{3.60}$$

We define the *reflectivity* and the *transmittivity* as the ratios:

$$\mathcal{R} = \frac{J^{(r)}}{J^{(i)}} = \frac{|R|^2}{|A|^2} \tag{3.61}$$

$$\mathcal{T} = \frac{J^{(t)}}{J^{(i)}} = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_i} \frac{|T|^2}{|A|^2} \tag{3.62}$$

It is easily verified that the conservation of energy is fulfilled:

$$\mathcal{R} + \mathcal{T} = 1 \tag{3.63}$$

If we split the incident wave into TE and TM components, and let α_i be the angle of the incident electric vector with the plane of incidence:

$$A_\| = A \cos \alpha_i \tag{3.64}$$

$$A_\perp = A \sin \alpha_i \tag{3.65}$$

It is easily shown that:

$$\mathcal{R} = \mathcal{R}_\| \cos^2 \alpha_i + \mathcal{R}_\perp \sin^2 \alpha_i \tag{3.66}$$

$$\mathcal{T} = \mathcal{T}_\| \cos^2 \alpha_i + \mathcal{T}_\perp \sin^2 \alpha_i \tag{3.67}$$

And each coefficient verifies:

$$\mathcal{R}_\| = \frac{J_\|^{(r)}}{J_\|^{(i)}} = \frac{|R_\||^2}{|A_\||^2} = |r_\||^2 \tag{3.68}$$

$$\mathcal{R}_\perp = \frac{J_\perp^{(r)}}{J_\perp^{(i)}} = \frac{|R_\perp|^2}{|A_\perp|^2} = |r_\perp| \tag{3.69}$$

$$\mathcal{T}_\| = \frac{J_\|^{(t)}}{J_\|^{(i)}} = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_i} \frac{|T_\||^2}{|A_\||^2} = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_i} |t_\||^2 \tag{3.70}$$

$$\mathcal{T}_\perp = \frac{J_\perp^{(t)}}{J_\perp^{(i)}} = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_i} \frac{|T_\perp|^2}{|A_\perp|^2} = \frac{n_2 \cos \theta_t}{n_1 \cos \theta_i} |t_\perp|^2 \tag{3.71}$$

We are using the $|r|^2$ and $|t|^2$ notation to account for the case where the reflection and transmission coefficients $r_\|$, $r_\perp$, $t_\|$ and $t_\perp$ are complex numbers, as we will see later. Also, notice that the reflectivity $\mathcal{R}$ is always a real number, whereas the transmittivity $\mathcal{T}$ may be a complex number if any of the variables θ_i, θ_t, n_2 and n_1 is a complex number.

When the transmittivities are real, each component verifies the conservation of energy:

$$\mathcal{R}_\| + \mathcal{T}_\| = 1 \tag{3.72}$$

$$\mathcal{R}_\perp + \mathcal{T}_\perp = 1 \tag{3.73}$$

Finally, notice that for normal incidence there is no distinction between the TM and TE waves.

So far all the angles in the system have been considered real. However, the law of refraction can yield complex transmitted angles:

$$\sin\theta_t = \frac{n_1}{n_2}\sin\theta_i \tag{3.74}$$

Since θ_i is always in the range $[-1, 1]$, if $n_2 > n_1$ then $\sin\theta_t$ is always in the range $[-1, 1]$ and θ_t is therefore a real number. However, if $n_2 < n_1$ then it is possible to get $\sin\theta_t > 1$ and θ_t is therefore a complex number. The value of θ_i that makes $\sin\theta_t = 1$ is called the *critical angle*:

$$\sin\theta_c = \frac{n_2}{n_1} \tag{3.75}$$

When $\theta_i = \theta_c$, the transmitted wave exits the boundary between the mediums at 90 degrees (parallel to the boundary). For angles larger than θ_c, the angle of the transmitted wave does not bear the meaning of a true propagation angle anymore, and in fact there is no transmitted wave. All the light is reflected, and this situation is known as *total reflection*. However, while there is no transmitted wave, the field in the second medium is not zero. Instead, there is an *evanescent wave* that propagates along the boundary between mediums.

More details about this condition can be found in reference [61]. All that matters to us here is that, in this situation, θ_t is a complex number. Correspondingly, the transmittance is also a complex number and, indeed, it is easily verified from the Fresnel equations that $\mathcal{R}_\parallel = \mathcal{R}_\perp = 1$, i.e, all the energy of the incident wave is reflected back into the first medium.

3.2.1.9   Propagation in multilayered systems

In this section, a method for calculating the reflection and transmission coefficients of multilayered systems will be discussed.

A *multilayered system* or *multilayer* is a system composed of mediums with different optical properties stacked one on top of the other. If the z axis is normal to the boundaries between the mediums, then the permittivity and permeability of the system, $\varepsilon(z)$ and $\mu(z)$, depend only on z and in a discrete way, so that we have $\varepsilon_1, \ldots, \varepsilon_n$ and $\mu_1, \ldots, \mu_n$ for the mediums 1 to n composing the system.

If we take the plane yz as the plane of incidence, with z the direction perpendicular to the interfaces between mediums, and considering a plane, time-harmonic wave with TE polarization, we have $E_y = E_z = H_x = 0$. It can be shown [61] from Maxwell's equations that the components of the fields can be written as:

$$E_x = U(z)e^{i(k_0\alpha y - \omega t)} \tag{3.76}$$

$$H_y = V(z)e^{i(k_0\alpha y - \omega t)} \tag{3.77}$$

$$H_z = W(z)e^{i(k_0\alpha y - \omega t)} \tag{3.78}$$

Here, α is a certain constant, k_0 was defined in equation (3.31) and $U(z)$, $V(z)$ and $W(z)$ contain the dependence on z, and are complex functions in general, related to one another by a system of three equations, so that knowing two of them automatically determines the third [61].

Since Maxwell's equations remain unchanged when $\vec{E}$ and $\vec{H}$ are interchanged at the same time that ε and $-\mu$ are interchanged, any valid results for TE waves is also valid for TM waves after making those changes. Therefore, it follows immediately that for TM waves:

$$H_x = U(z)e^{i(k_0\alpha y - \omega t)} \tag{3.79}$$

$$E_y = -V(z)e^{i(k_0\alpha y - \omega t)} \tag{3.80}$$

$$E_z = -W(z)e^{i(k_0\alpha y - \omega t)} \tag{3.81}$$

U and V both satisfy a second-order linear differential equation [61], and therefore they can be expressed as a linear combination of two particular solutions, $U_1(z) = f(z)$, $U_2(z) = F(z)$, $V_1(z) = g(z)$ and $V_2(z) = G(z)$. We choose those functions so that:

$$f(0) = G(0) = 0, \qquad F(0) = g(0) = 1 \tag{3.82}$$

Then, the solutions with $U(0) = U_0$ and $V(0) = V_0$ can be expressed as:

$$U = FU_0 + fV_0 \tag{3.83}$$

$$V = GU_0 + gV_0 \tag{3.84}$$

The last system of two equations can be conveniently written in matrix form:

$$\mathbf{Q_0} = \mathbf{MQ} \tag{3.85}$$

$$\mathbf{Q} = \begin{pmatrix} U(z) \\ V(z) \end{pmatrix} \qquad \mathbf{Q_0} = \begin{pmatrix} U_0 \\ V_0 \end{pmatrix} \qquad \mathbf{M} = \begin{pmatrix} g(z) & -f(z) \\ -G(z) & F(z) \end{pmatrix} \tag{3.86}$$

The matrix $\mathbf{M}$ is called the *characteristic matrix* of the system, and it can be used to relate the x and y components of the electric or magnetic vectors in $z = 0$ to the components in any other plane $z = $ constant. Furthermore, according to the relations between U, V and W, knowing two of them is enough to determine the field completely.

This is a very important result, because it says that all we need to know to determine the propagation of a plane monochromatic wave through a multilayered system is its characteristic matrix.

Solving the second-order linear differential equations satisfied by U and V [61] for a single, homogeneous dielectric film (constant ε and μ and taking $n = \sqrt{\varepsilon\mu}$), and a TE wave propagating at an angle θ in the medium, gives: [61]:

$$\mathbf{M}(z) = \begin{pmatrix} \cos(k_0 n z \cos\theta) & -\frac{i}{p}\sin(k_0 n z \cos\theta) \\ -ip\sin(k_0 n z \cos\theta) & \cos(k_0 n z \cos\theta) \end{pmatrix} \tag{3.87}$$

where

$$p = \sqrt{\frac{\varepsilon}{\mu}} \cos \theta \tag{3.88}$$

For a TM wave in the same circumstances the result is the same, except p must be replaced by:

$$q = \sqrt{\frac{\mu}{\varepsilon}} \cos \theta \tag{3.89}$$

Now consider two stacked media, one from $z = 0$ to $z = z_1$ and another from $z = z_1$ to $z = z_2$. If $\mathbf{M_1}(z)$ and $\mathbf{M_2}(z)$ are their characteristic matrices:

$$\mathbf{Q_0} = \mathbf{M_1}(z_1)\mathbf{Q}(z_1) \tag{3.90}$$
$$\mathbf{Q}(z_1) = \mathbf{M_2}(z_2 - z - 1)\mathbf{Q}(z_2) \tag{3.91}$$

Therefore:

$$\mathbf{Q_0} = \mathbf{M_1}(z_1)\mathbf{M_2}(z_2 - z_1)\mathbf{Q}(z_2) \tag{3.92}$$

The matrix $\mathbf{M}(z_z) = \mathbf{M_1}(z_1)\mathbf{M_2}(z_2 - z_1)$ allows us to find the components of the fields at $z = 0$ from the components of the field at $z = z_2$. Hence, the characteristic matrix of the two-layer system is the matrix product of the characteristic matrices of each individual layer. This can be extended to a general case with N mediums:

$$\mathbf{M}(z_N) = \mathbf{M_1}(z_1)\mathbf{M_2}(z_2 - z_1) \cdots \mathbf{M_N}(z_N - z_{N-1}) \tag{3.93}$$

Now we will show that it is possible to get the reflection and transmission coefficients of a multilayered system if we know its characteristic matrix. For simplicity, suppose a single film of finite thickness sandwiched between two mediums with constants ε_1, μ_1 and ε_2, μ_2. The angles θ_1 and θ_2 are the angles between the normal to the interfaces (z axis) and the direction of propagation of the light in the first and second mediums, respectively. Furthermore, let A, R and T be the complex amplitudes of the incident, reflected and transmitted waves, respectively. It is easy to see that, for a TE wave:

$$U_0 = A + R \tag{3.94}$$
$$V_0 = p_1(A - R) \tag{3.95}$$
$$U(z_1) = T \tag{3.96}$$
$$V(z_1) = p_2 T \tag{3.97}$$

where:

$$p_1 = \sqrt{\frac{\varepsilon_1}{\mu_1}} \cos \theta_1 \tag{3.98}$$

$$p_2 = \sqrt{\frac{\varepsilon_2}{\mu_2}} \cos \theta_2 \tag{3.99}$$

If we apply the basic equation ([3.85](#)) we get:

$$A + R = (m_{11} + m_{12}p_2)T \tag{3.100}$$

$$p_1(A - R) = (m_{21} + m_{22}p_2)T \tag{3.101}$$

Here m_{11}, m_{12}, m_{21} and m_{22} are the elements of the characteristic matrix of the film between the two mediums. From the last relations we can trivially find the reflection and transmission coefficients:

$$r = \frac{R}{A} = \frac{(m_{11} + m_{12}p_2)p_1 - (m_{21} + m_{22}p_2)}{(m_{11} + m_{12}p_2)p_1 + (m_{21} + m_{22}p_2)} \tag{3.102}$$

$$t = \frac{T}{A} = \frac{2p_1}{(m_{11} + m_{12}p_2)p_1 + (m_{21} + m_{22}p_2)} \tag{3.103}$$

The reflectivity and the transmittivity are then:

$$\mathcal{R} = |r|^2 \tag{3.104}$$

$$\mathcal{T} = \frac{p_2}{p_1}|t|^2 \tag{3.105}$$

Note that if we fix one of the angles of propagation, say θ_1, then the angles in all the other mediums are fixed by Snell's law. Furthermore, r and t are complex numbers because they carry information about the difference in phase between the incident, reflected and transmitted waves. The phase δ_r of r is the *phase change on reflection* and is referred to the boundary between the first medium and the layer, whereas the phase δ_t of t is the *phase change on transmission* and is referred to the boundary between the layer and the second medium.

Of course, this formalism works the same if we use the characteristic matrix of a stack of several layers sandwiched between two mediums.

The results for TM waves are obtained from those for TE waves after substituting p_1 and p_2 by:

$$q_1 = \sqrt{\frac{\mu_1}{\varepsilon_1}} \cos\theta_1 \tag{3.106}$$

$$q_2 = \sqrt{\frac{\mu_2}{\varepsilon_2}} \cos\theta_2 \tag{3.107}$$

However, note that in this case the reflection and transmission coefficients refer to the ratios between the magnetic fields, not the electric fields, i.e the coefficients that we obtain by this formalism correspond to the following quantities:

$$r_{\text{TE}} = \frac{E^{(r)}}{E^{(i)}} \qquad\qquad t_{\text{TE}} = \frac{E^{(t)}}{E^{(i)}} \tag{3.108}$$

$$r_{\text{TM}} = \frac{H^{(r)}}{H^{(i)}} \qquad\qquad t_{\text{TM}} = \frac{H^{(t)}}{H^{(i)}} \tag{3.109}$$

Further on we will want r_{TM} and t_{TM} expressed in terms of the electric fields, not the magnetic fields. The conversion can be made making use of equation (3.41), $n = \sqrt{\varepsilon\mu}$, and taking $\mu = 1$:

$$r_{\mathrm{TM}} = \frac{H^{(r)}}{H^{(i)}} = \frac{n_i E^{(r)}}{n_i E^{(i)}} = \frac{E^{(r)}}{E^{(i)}} \tag{3.110}$$

Therefore, the reflection coefficient for TM waves obtained from the characteristic matrix is directly the quantity that we want. However, for the transmission coefficient:

$$t_{\mathrm{TM}} = \frac{H^{(t)}}{H^{(i)}} = \frac{n_t}{n_i} \frac{E^{(t)}}{E^{(i)}} \Rightarrow \frac{E^{(t)}}{E^{(i)}} = \frac{n_i}{n_t} t_{\mathrm{TM}} \tag{3.111}$$

That means that the transmission coefficient for TM waves expressed in terms of the electric fields differs by a factor n_i/n_t from the transmission coefficient obtained with the above method, which comes expressed in terms of the magnetic fields.

3.2.1.10 Propagation in conductive media

In the previous sections we have always considered our media to be nonconducting. Conductivity is related to energy dissipation through heat, as we saw in section 3.2.1.3, which means that in conductive media the electromagnetic field will be attenuated.

Proceeding as we did in the previous sections, but considering $\sigma \neq 0$, it is easy to show [61] that the electromagnetic field satisfies the same equations, except the dielectric constant ε must be replaced by the complex number:

$$\hat{\varepsilon} = \varepsilon + i\frac{4\pi\sigma(\omega)}{\omega} \tag{3.112}$$

Correspondingly, we also introduce a complex wave number $\hat{k}$, a complex phase velocity $\hat{v}$ and a complex refractive index $\hat{n}$:

$$\hat{k}^2 = \frac{\omega^2\mu}{c^2}\hat{\epsilon} \tag{3.113}$$

$$\hat{v} = \frac{c}{\sqrt{\mu\hat{\varepsilon}}} \tag{3.114}$$

$$\hat{n} = \frac{c}{\hat{v}} = \sqrt{\mu\hat{\varepsilon}} = \frac{c}{\omega}\hat{k} \tag{3.115}$$

In general, we will refer to the real part of the complex refractive index as *refractive index*, and to its imaginary part as *extinction coefficient*.

We will not go into more details here. Suffice to say that all the results we have shown so far are also valid for conductive media if the complex refractive index is used. In particular, Snell's law is valid:

$$\sin\theta_i = \frac{\hat{n_1}}{\hat{n_2}}\sin\theta_i \tag{3.116}$$
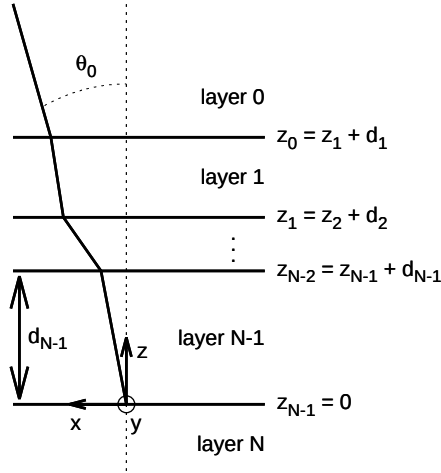
Figure 3.3: Schematic view of a general layered system.

And we can still get the reflection and transmission coefficients with the characteristic matrix formalism. Furthermore, the absorbance of the films manifests itself naturally when the extinction coefficient is nonzero so that $\mathcal{R} + \mathcal{T} \neq 1$, since there is energy to heat conversion in the medium.

3.2.2 The Crawford method

In reference [60] a method is developed to calculate the effect of a multilayered system on the emission of a dipole embedded in any of the layers. In this section we will briefly describe the method and discuss how to successfully apply it to our systems.

3.2.2.1 Basic theory

Figure 3.3 depicts our general multilayered system and the coordinate system that we are going to use. Basically, we have a system of $N + 1$ layers. The uppermost (layer 0) and bottommost (layer N) layers are in fact semi infinite mediums[2], whereas the intermediate ones, layers 1, ..., $N - 1$, have finite thicknesses d_1, ..., d_{N-1}. The position of each layer is taken as the z coordinate of its lower interface and the origin of the z axis is placed in the lowest interface of the system, i.e the interface between the bottom medium and layer $N - 1$. Hence, the position of any layer is $z_j = z_{j+1} + d_{j+1}$ with $z_{N-1} = 0$. Each layer is characterized by a (generally complex) refractive index $\varepsilon_j = (n_j + i k_j)^2$ and $\mu_j = 1$.

---

[2]Here our mediums go from 0 to N ($N + 1$ layers) instead of from 1 to N (N layers) as in reference [60]. That will change some subscripts and superscripts of the expressions for F_x, F_y and F_z.

Now consider a plane monochromatic wave coming from the uppermost medium with an angle of incidence θ_0. The angle of propagation θ_j in any other layer can be calculated using Snell's law. All angles are measured with respect to the z axis, as shown in figure 3.3. In what follows, the phase factor $\exp(-i\omega t)$ will usually be omitted, and it will be understood that all field amplitudes and dipole moment require it.

We can write the field in any layer j as the sum of the field propagating in the direction of decreasing z ($\downarrow$ superscript) and the field propagating in the direction of increasing z ($\uparrow$ superscript):

$$\vec{E}^q(\vec{r}) = E^\downarrow_{qj}\, \hat{\vec{u}}^\downarrow\, e^{i(\vec{k}^\downarrow_j \cdot \vec{r} - \omega t)} + E^\uparrow_{qj}\, \hat{\vec{u}}^\uparrow\, e^{i(\vec{k}^\uparrow_j \cdot \vec{r} - \omega t)} \tag{3.117}$$

The subscript $q = \parallel,\ \perp$ stands for the polarization of the wave (TE or TM). $E^\downarrow_{qj}$, $E^\uparrow_{qj}$ are scalar constants and $\hat{\vec{u}}^\downarrow$, $\hat{\vec{u}}^\uparrow$ are unit vectors.

Expressing each wave in terms of reflected and transmitted waves, one finds that for TE waves it is possible to express the field at any point in the system in terms of the incoming field [60]:

$$\vec{E}^\perp(\vec{r}) = \hat{y}\, F_y(z)\, E^\downarrow_{y0}\, e^{i\vec{k}^\downarrow_0 \cdot \vec{r}} \tag{3.118}$$

Here the function $F_y(z)$ clearly represents the ratio of the field with the multilayered system in place to the field at the same point in the absence of the multilayered system, i.e if all space was filled with the medium corresponding to layer 0. Depending on the position, $F_y(z)$ takes the following form:

$$F_y(z) = \begin{cases} 1 + r^{(N+1)}_{\perp 01} e^{2i\eta_0(z-z_0)} & z > z_0 \\ \dfrac{t^{(j+1)}_{\perp 0j}\left(1 + r^{(N-j+1)}_{\perp jj+1} e^{2i\eta_j(z-z_j)}\right)}{1 - r^{(N-j+1)}_{\perp jj+1} r^{(j+1)}_{\perp jj-1} e^{2i\eta_j d_j}} e^{i\eta_0(z-z_0)-i\eta_j(z-z_{j-1})} & z_{j-1} < z < z_j \\ t^{(N+1)}_{\perp 0N} e^{i\eta_0(z-z_0)-i\eta_N z} & z < 0 \end{cases} \tag{3.119}$$

Here, $0 < j < N$ and the coefficients $r^{(m)}_{qjk}$, where $q = \perp,\ \parallel$ and $k = j+1,\ j-1$, are the reflection coefficients of the subsystem composed by the m contiguous layers behind the interface jk, i.e. the ratio of the electric field coming in the direction $j \leftarrow k$ (reflected) to the field incident on the interface $j \rightarrow k$, the amplitudes being evaluated at the layer j. The coefficient takes into account reflections at the $m-1$ contiguous interfaces starting with interface jk. As an example, $r^{(N+1)}_{\perp 01}$ is the reflection coefficient of the whole stack for light incident into the first interface.

Similarly, $t^{(m)}_{qjk}$ is the transmission coefficient for light going from layer j to layer k crossing $m-1$ interfaces between j and k, i.e the ratio of the electric field in the direction $j \rightarrow k$ evaluated at k to the electric field in the same direction $j \rightarrow k$ evaluated at j. As an example, $t^{(N+1)}_{\perp 0N}$ is the transmission coefficient of the whole stack.

Finally, we have used:

$$\eta_j = \frac{\omega}{c}\sqrt{\varepsilon_j - \varepsilon_1 \sin^2 \theta_0} \tag{3.120}$$

Working in the same way, one can find similar expressions for the components of a TM wave:

$$F_x(z) = \begin{cases} 1 - r_{\parallel 01}^{(N+1)} e^{2i\eta_0(z-z_0)} & z > z_0 \\[2mm] \dfrac{t_{\parallel 0j}^{(j+1)}(1 - r_{\parallel jj+1}^{(N-j+1)} e^{2i\eta_j(z-z_j)})}{1 - r_{\parallel jj+1}^{(N-j+1)} r_{\parallel jj-1}^{(j+1)} e^{2i\eta_j d_j}} e^{i\eta_0(z-z_0)-i\eta_j(z-z_{j-1})} \dfrac{\cos\theta_j}{\cos\theta_0} & z_{j-1} < z < z_j \\[4mm] t_{\parallel 0N}^{(N+1)} e^{i\eta_0(z-z_0)-i\eta_N z} \dfrac{\cos\theta_N}{\cos\theta_0} & z < 0 \end{cases} \tag{3.121}$$

$$F_z(z) = \begin{cases} 1 + r_{\parallel 01}^{(N+1)} e^{2i\eta_0(z-z_0)} & z > z_0 \\[2mm] \dfrac{t_{\parallel 0j}^{(j+1)}(1 + r_{\parallel jj+1}^{(N-j+1)} e^{2i\eta_j(z-z_j)})}{1 - r_{\parallel jj+1}^{(N-j+1)} r_{\parallel jj-1}^{(j+1)} e^{2i\eta_j d_j}} e^{i\eta_0(z-z_0)-i\eta_j(z-z_{j-1})} \dfrac{\sin\theta_j}{\sin\theta_0} & z_{j-1} < z < z_j \\[4mm] t_{\parallel 0N}^{(N+1)} e^{i\eta_0(z-z_0)-i\eta_N z} \dfrac{\sin\theta_N}{\sin\theta_1} & z < 0 \end{cases} \tag{3.122}$$

Here once again $0 < j < N$.

A plane wave with arbitrary polarization at any point in the system can thus be written in terms of the incident wave as:

$$\vec{E}(\vec{r}) = \hat{y}\, F_y(z)\, E_{y0}\, e^{i\vec{k}_0^\downarrow \cdot \vec{r}} + (\hat{x}\, F_x(z)\cos\theta_0 - \hat{z}\, F_z(z)\sin\theta_0)\, E_{\theta 0}\, e^{i\vec{k}_0^\downarrow \cdot \vec{r}} \tag{3.123}$$

It can also be shown that, due to the continuity relations (3.8)–(3.11), the F functions evaluated on the two sides of an interface ij must satisfy:

$$F_{x,\,y}(z_{ij} + \Delta_i) = F_{x,\,y}(z_{ij} + \Delta_j) \tag{3.124}$$

$$\varepsilon_i F_z(z_{ij} + \Delta_i) = \varepsilon_j F_z(z_{ij} + \Delta_j) \tag{3.125}$$

Here Δ_i and Δ_j are small displacements into the layers i and j respectively. That explains the discontinuities observed for TM waves at the interfaces (see figure 3.11(b)).

So far an expression for the field in the multilayer in terms of the incident field has been found. However, we are interested in finding the field in the first layer (the upper medium) in terms of the field generated within any of the layers. In reference [60] is is shown how that can be related to the previous result by reciprocity, so that the field in a point $\vec{R}$ in the upper medium due to a dipole $\vec{p}$ located at $\vec{r}_a$ within any of the layers can be expressed as:

$$\vec{E}^{(a)}(\vec{R}) = \left[\hat{y}p_y F_y(z_a) + \hat{\theta}\,(p_x F_x(z_a)\cos\theta_0 - p_z F_z(z_a)\sin\theta_0)\right] \times$$
$$\times\, e^{i\vec{k}_0^\downarrow \cdot \vec{r}_a} \left(\frac{\omega}{c}\right)^2 R^{-1} e^{ik_0 R} \tag{3.126}$$

Here p_x, p_y and p_z are the components of the dipole with moment $\vec{p}$. The radiated power per unit solid angle can then be written as:

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega} = \frac{c}{8\pi} \left(\frac{\omega}{c}\right)^4 \varepsilon_0^{\frac{1}{2}} \times \left(|p_y F_y(z_a)|^2 + |p_x F_x(z_a)\cos\theta_0 - p_z F_z(z_a)\sin\theta_0|^2\right) \quad (3.127)$$

3.2.2.2   The effect of the layered system on TE and TM waves

The first term in the parentheses of equation (3.127) comes from the TE component whereas the second one corresponds to the TM component of the wave. Clearly, if we set $p_x = p_z = 0$ it follows that $|F_y(z_a)|^2$ is the ratio of the energy we get from a dipole oriented in the y direction to the energy we would get if there was no multilayer. A similar conclusion can be drawn for dipoles oriented along the x and z directions. Note that while the TE and TM components are independent, the x and z components of the TM wave are not independent.

As per the above discussion, it is clear that for TE waves the quantity we want to calculate is $|F_y(z_a)|^2$. However, for TM waves it is not so clear.

In order to arrive at equation (3.126) the emission of the dipole is expressed for the direction perpendicular to the orientation of the dipole. Moreover, remember that $F_i(z)$ is implicitly a function of λ and θ_0. That means that we can set a dipole oriented along the x direction in equations (3.126) and (3.127). The value of the intrinsic energy emitted by the dipole corresponds to the emission normal to the orientation of the dipole [60]. However, we can observe at an arbitrary angle θ_0. The result of evaluating equation (3.127) has two components: one comes from the fact that the layer is affecting the energy through $F_x(z_a)$, and another one comes from the fact that we are observing the radiation at an angle θ_0 and therefore only the fraction of the energy projected to that direction is being accounted for.

As an example, set $p_z = p_y = 0$ and $\theta_0 = \pi/2$. It is clear from equation (3.127) that $\mathrm{d}W/\mathrm{d}\Omega = 0$. The reason is not that the layer is absorbing all the energy through F_x, but rather that we are setting the direction of observation perpendicular to the direction of emission.

For TE waves this is not a problem because a dipole oriented along the y direction will emit the same amount of energy in all directions of the plane of incidence, which means that, regardless of θ_0, F_y will always be the ratio between two rays that come from the dipole with the same original energy and propagate with the same angle θ_0 in layer 0. In figure 3.4 the solid line corresponds to a ray that, after refraction in the first interface, travels at a certain angle θ_0 in the upper medium. The dashed line corresponds to a ray that, in the absence of the layered system, travels with angle θ_0 too. Although the angle of each ray as they exit the dipole is different, their original energy is the same due to symmetry considerations. That is why there is no θ_0 in the y term of equation (3.127), because for every direction there is a pair of rays with the same intrinsic energy, each traveling with different angles in the interior layers but with the same angle in layer 0.
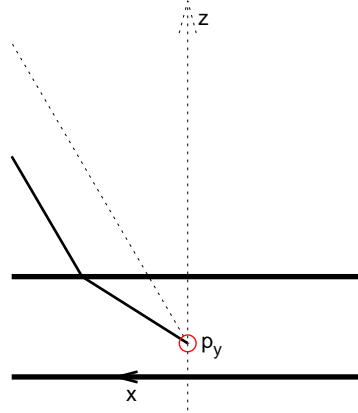
Figure 3.4: Dipole oriented along the y axis. The dashed line corresponds to emission without the multilayer. The solid line corresponds to a ray that travels in the upper medium in the same direction as the dashed line.

That is not the case when we consider a dipole with only p_x and p_z components. For example, consider the result that you get after setting $p_y = p_z = 0$ and set $F_x = 1$ in equation (3.127), which corresponds to no multilayer present.

Rays that exit the dipole in different angles have intrinsically different energies, so that the comparison does not make sense. In reference [60] the author calculates $|F_x(z)\cos\theta_0|^2$ for dipoles oriented along the x axis and $|F_z(z)\sin\theta_0|^2$ for dipoles oriented along the z axis. This corresponds to the ratio of the energy in the direction of θ_0 due to dipoles along the x and z directions, respectively, to the energy that a dipole with the same strength but oriented along the y direction would emit in the absence of the layered system. That makes sense, but leads to different results for dipoles p_x and p_y even though we set the angle θ_0 to the same value. We know why this happens, but we would like an expression that assigns a univocal value to each ray of light, because of course the layer affects each ray differently, but it should not depend on how this ray was generated.

The quantity that we will calculate to evaluate the effect of the multilayer on TM waves is $|F_x(z)\cos^2\theta_0 + F_z(z)\sin^2\theta_0|^2$, which corresponds to setting the orientation of the dipole normal to the direction of propagation that we are considering in layer 0. This way, removing the multilayer leaves us with a ray with angle θ_0 and normal to the orientation of the dipole. When we add the multilayer, that very same ray does not travel with angle θ_0 in layer 0 because of the refraction, and therefore the effect of the multilayer on the ray comes not only from F_x and F_z but also from the fact that now the incidence of the ray on an hypothetical detector is slanted.

In the end, it does not matter what we choose to calculate, but we have to make sure we clearly understand what is it that we are calculating. An in depth understanding of the process that leads to equations (3.126) and (3.127) in reference

[60] is important to fully understand the previous discussion.

3.2.2.3 Distribution of the radiative centers

In section 3.2.2.2 we arrived at the conclusion that, for TE waves, we want to calculate

$$K_{\text{TE}} = |F_y(z; \lambda, \theta_0)|^2 \tag{3.128}$$

Whereas for TM waves:

$$K_{\text{TM}} = |F_x(z; \lambda, \theta_0) \cos^2 \theta_0 + F_y(z; \lambda, \theta_0) \sin^2 \theta_0|^2 \tag{3.129}$$

Here the implicit dependence of F_p ($p = x, y, z$) in λ and θ_0 has been written explicitly. In general, two of the parameters will be fixed in order to study how the third one affects the emission of a radiative center. The quantities K_{TE} and K_{TM} will be called the *energy ratio* for TE and TM waves, respectively.

A very important case will be fixing z and θ_0 and varying λ. This essentially gives the factors by which the intrinsic spectrum of the radiative center must be multiplied to give the spectrum that will actually be observed at an angle θ_0 for a radiative center located at z, i.e:

$$f_{\text{obs}}(\lambda) = K_i(\lambda; \theta_0, z)\, f_{\text{int}}(\lambda) \tag{3.130}$$

Here $i = $ TE, TM and $f_{\text{obs}}(\lambda)$ and $f_{\text{int}}(\lambda)$ are the *observed* and *intrinsic* emission spectra, respectively. Note that we have written K as a function of λ while θ_0 and z are considered fixed parameters.

While considering a radiative center at position z in the multilayer can give useful information of a real system, in our devices we will not have all the radiative centers at the same z coordinate. Instead, the radiative centers will follow a certain distribution $g(z)$. As a simple example, let's consider a layered system with two radiative centers at positions z_1 and z_2 and intrinsic spectra $f_{\text{int}}^{(1)}(\lambda)$ and $f_{\text{int}}^{(2)}(\lambda)$. Each spectrum can be written as:

$$f_{\text{int}}^{(i)}(\lambda) = g_i\, \bar{f}_{\text{int}}^{(i)}(\lambda) \tag{3.131}$$

Here $\bar{f}_{\text{int}}^{(i)}$ stands for the spectrum *normalized to area 1*, while g_i is the total energy of the emission.

It is clear that the spectrum of the radiative center at z_1 will be affected by the quantity $K_p(\lambda; \theta_0, z_1)$ and similarly for the radiative center at z_2. The detected spectrum will therefore have two components:

$$f_{\text{obs}}(\lambda) = f_{\text{obs}}^{(1)}(\lambda) + f_{\text{obs}}^{(2)}(\lambda) =$$
$$= g_1 K_p(\lambda; \theta_0, z_1)\bar{f}_{\text{int}}^{(1)} + g_2 K_p(\lambda; \theta_0, z_2)\bar{f}_{\text{int}}^{(2)} \tag{3.132}$$

This can easily be extended to an arbitrary non discrete distribution of radiative centers $g(z)$ and a spectrum which is in general a function of z, $f_{\text{int}}(\lambda, z)$.

$$f_{\text{obs}}(\lambda) = \int_{-\infty}^{\infty} K_p(z; \lambda, \theta_0)\, g(z)\, f_{\text{int}}(\lambda, z)\, \mathrm{d}z \tag{3.133}$$

Note that $g(z)$ is the factor the weighs the normalized spectrum at each value of z (we will drop the bar notation and consider all the spectra to be normalized to area 1 from now on). If we have $g(z_1) > g(z_2)$, for our purposes it is not relevant whether g is larger in z_1 because there are actually more radiative centers there or because the recombination rate is higher. All that matters to us is that the emission in z_1 contributes more to the total than the emission at z_2. We will refer to $g(z)$ as the *distribution of radiative centers*.

In case we are interested in the total energy, an integral over λ must be included:

$$W_{\text{obs}} = \int_{-\infty}^{\infty} \int_{0}^{\infty} K_p(\lambda, z; \theta_0)\, g(z)\, f_{\text{int}}(\lambda, z)\, \mathrm{d}z\, \mathrm{d}\lambda \tag{3.134}$$

Of course, the intrinsic total energy radiated in the system is:

$$W_{\text{int}} = \int_{-\infty}^{\infty} \int_{0}^{\infty} g(z)\, f_{\text{int}}(\lambda, z)\, \mathrm{d}z\, \mathrm{d}\lambda \tag{3.135}$$

This clearly corresponds to setting $K_p(\lambda, z; \theta_0) = 1$ (i.e no multilayer present) in equation (3.134) and therefore W_{int} is the total energy that would be received by an observer in the far field if there was no multilayer present. We will generally refer to that as the *total intrinsic energy*, although strictly speaking it is *not* the total intrinsic emission, only the emission that would reach the observer in case there was no multilayer.

For convenience we want the total intrinsic energy W_{int} to be 1 since then W_{obs} will directly indicate how much of the intrinsic energy is actually observed. Therefore, the intrinsic distribution of energy in the layered system $F(z, \lambda) = g(z)\, f_{\text{int}}(\lambda, z)$ should be normalized to volume 1:

$$\int_{-\infty}^{\infty} \int_{0}^{\infty} g(z)\, f_{\text{int}}(\lambda, z)\, \mathrm{d}z\, \mathrm{d}\lambda = 1 \tag{3.136}$$

It is clear that, with the normalization condition (3.136), the observed spectrum $f_{\text{obs}}(\lambda)$ in equation (3.133) is also expressed relative to the emitted one.

Now let's discuss a couple of interesting cases.

3.2.2.4 Single active layer

Suppose a simple system consisting of many layers with only one of them, between z_i and z_f, $z_i < z_f$, being optically active. The distribution of radiative centers is:

$$g_T(z) = \begin{cases} g(z) \neq 0 & z_i \leq z < z_f \\ 0 & z < z_i,\ z_f \leq z \end{cases} \tag{3.137}$$

Equation (3.133) is then reduced to:

$$f_{\text{obs}}(\lambda) = \int_{z_i}^{z_f} K_p(z; \lambda, \theta_0) \, g(z) \, f_{\text{int}}(\lambda, z) \, \mathrm{d}z \tag{3.138}$$

It is reasonable to assume that all the radiative centers in the active layer will have the same spectrum if we are considering an homogeneous medium. Therefore, $f_{\text{int}}(\lambda, z)$ is actually $f_{\text{int}}(\lambda)$. Hence:

$$f_{\text{obs}}(\lambda) = f_{\text{int}}(\lambda) \int_{z_i}^{z_f} K_p(z; \lambda, \theta_0) \, g(z) \, \mathrm{d}z = f_{\text{int}}(\lambda) \, \bar{K}_p(\lambda, \theta_0) \tag{3.139}$$

Since $f_{\text{int}}(\lambda)$ can be left out of the integral, we can express the observed spectrum as the intrinsic one multiplied by an *effective* energy ratio $\bar{K}_p(\lambda, \theta_0)$. Therefore, if we know $\bar{K}_p(\lambda, \theta_0)$, then we can calculate the observed spectrum if we know the intrinsic one, or the other way around.

Furthermore, in this case the normalization condition is:

$$\int_{-\infty}^{\infty} \int_{z_i}^{z_f} g(z) \, f_{\text{int}}(\lambda) \, \mathrm{d}z \, \mathrm{d}\lambda = \int_{-\infty}^{\infty} f_{\text{int}}(\lambda) \, \mathrm{d}\lambda \int_{z_i}^{z_f} g(z) \, \mathrm{d}z = 1 \tag{3.140}$$

This condition is clearly satisfied if:

$$\int_{-\infty}^{\infty} f_{\text{int}}(\lambda) \, \mathrm{d}\lambda = 1, \qquad \int_{z_i}^{z_f} g(z) \, \mathrm{d}z = 1 \tag{3.141}$$

3.2.2.5 Double active layers

In this section we will be concerned with a layered system with two different active layers, which will usually be contiguous, although they don't need to be for the analysis that follows.

Let's assume the first active layer to range from z_{1i} to z_{1f}, whereas the second one ranges from z_{2i} to z_{2f} with $z_{1i} < z_{1f} < z_{2i} < z_{2f}$. The distribution of the radiative centers can be written as:

$$g_T(z) = \begin{cases} g_1(z) \neq 0 & z_{1i} \leq z < z_{1f} \\ g_2(z) \neq 0 & z_{2i} \leq z < z_{2f} \\ 0 & z < z_{1i}, \; z_{1f} \leq z < z_{2i}, \; z_{2f} \leq z \end{cases} \tag{3.142}$$

In this case, equation (3.133) is reduced to:

$$f_{\text{obs}}(\lambda) = \int_{z_{1i}}^{z_{1f}} K_p(z; \lambda, \theta_0) \, g_1(z) \, f_{\text{int}}(\lambda, z) \, \mathrm{d}z +$$

$$+ \int_{z_{2i}}^{z_{2f}} K_p(z; \lambda, \theta_0) \, g_2(z) \, f_{\text{int}}(\lambda, z) \, \mathrm{d}z \tag{3.143}$$

Now it is also reasonable to assume, as we did in section 3.2.2.4, that in each layer the spectrum does not depend on z. Therefore we have:

$$f_{\text{int}}(\lambda,\, z) = \begin{cases} f_{\text{int}}^{(1)}(\lambda) & z_{1i} \leq z < z_{1f} \\ f_{\text{int}}^{(2)}(\lambda) & z_{2i} \leq z < z_{2f} \end{cases} \tag{3.144}$$

And equation (3.143) becomes:

$$f_{\text{obs}}(\lambda) = f_{\text{int}}^{(1)}(\lambda) \int_{z_{1i}}^{z_{1f}} K_p(z; \lambda,\, \theta_0)\, g_1(z)\, \mathrm{d}z +$$
$$+ f_{\text{int}}^{(2)}(\lambda) \int_{z_{2i}}^{z_{2f}} K_p(z; \lambda,\, \theta_0)\, g_2(z)\, \mathrm{d}z \quad (3.145)$$

Since it is clear that $f_{\text{obs}}(\lambda)$ has two independent contributions, each originating in a different layer, we can treat them separately, i.e:

$$f_{\text{obs}}(\lambda) = f_{\text{obs}}^{(1)}(\lambda) + f_{\text{obs}}^{(2)}(\lambda) \tag{3.146}$$

where

$$f_{\text{obs}}^{(n)}(\lambda) = f_{\text{int}}^{(n)}(\lambda) \int_{z_{ni}}^{z_{nf}} K_p(z; \lambda,\, \theta_0)\, g_n(z)\, \mathrm{d}z =$$
$$= f_{\text{int}}^{(n)}(\lambda) \bar{K}_p^{(n)}(\lambda,\, \theta_0) \qquad n = 1,\, 2 \quad (3.147)$$

The last result can be readily extended to an arbitrary number of active layers. Also, note that we can always write equation (3.146). However, in order to place the intrinsic spectrum out of the integral in equation (3.147), which allows us to transform $f_{\text{obs}}^{(n)} \leftrightarrow f_{\text{int}}^{(n)}$ in either direction, we must first assume that $f_{\text{int}}^{(n)}$ does not depend on z.

Regarding the normalization, we have one equation like (3.140) for each active layer, and we want the sum of all of them to evaluate to 1, i.e we want the total intrinsic energy to be 1. For two active layers with spectra independent of z we have:

$$\int_{-\infty}^{\infty} f_{\text{int}}^{(1)}(\lambda)\, \mathrm{d}\lambda \int_{z_{1i}}^{z_{1f}} g_1(z)\, \mathrm{d}z + \int_{-\infty}^{\infty} f_{\text{int}}^{(2)}(\lambda)\, \mathrm{d}\lambda \int_{z_{2i}}^{z_{2f}} g_2(z)\, \mathrm{d}z = 1 \tag{3.148}$$

Here the relative weight of each term is related to the relative intensity of the emission of each layer.

Generally we will consider the spectra $f_{\text{int}}^{(n)}$ to be normalized to area 1 and then choose $g_1(z)$ and $g_2(z)$ according to the intensities of each layer relative to one another. However, it is entirely possible to choose $g_1(z)$ and $g_2(z)$ to be both normalized to area 1 and then choose the area of each spectrum accordingly. While the first approach is conceptually more straightforward, the latter approach is more convenient when we have experimentally determined each spectrum with their specific relative intensities. In that case all we have to do is multiply both spectra for a common factor so that their combined integrals add up to 1 and then normalize each distribution $g_n(z)$ to 1.

3.3   The Python implementation

In this chapter a Python implementation of the calculations presented in section 3.2.2 will be discussed. The code is well documented and publicly available for download at Github. See appendix B for more information on how to download and use the code.

3.3.1   Why Python?

Python [62] is an interpreted programming language and, as such, it is usually regarded as inadequate for science and engineering because interpreted languages lag behind compiled languages such as C, C++ or Fortran in terms of raw performance. In spite of that, Python is being used consistently in many research areas such as photonics or neuroscience. There are several reasons for that.

The most important reason for choosing Python over C, C++ or Fortran is that *scientist time is more important than computing time*, at least in most scenarios. Python has a very straightforward syntax, it takes care of the memory management and has a large amount of modules that provide high level functionality that removes a lot of clutter in the code. The result is that any program can be prototyped in Python much faster than in any of the aforementioned compiled languages. For a scientist, this is important because it means different approaches to the same problem can be explored in Python in the same amount of time that a single, albeit faster, approach could be explored in other languages.

Once a Python program is working there are many ways in which its performance can be improved, *if* that is in fact *necessary*.

The use of Numpy [63] works around most performance issues that a scientist might encounter in his Python programs. The vectorization capabilities of Numpy make working with arrays of data in Python as fast as in C or Fortran. There are other specific optimization tools like numexpr [64] or Cyton [65], and of course the code can be parallelized or accelerated via GPU (see figure 3.5).

An added advantage of Python over other programming languages used by scientists, such as Matlab, is that Python is free software (as in *free speech*). This is important because the ability of the programmer to run his own code does not depend on the availability of a license or on the conditions set in the EULA[3].

It is important to note, however, that there is no universal solution that works for everything. Each programming language has its strengths and weaknesses, but we believe that Python was the right choice for our problem.

3.3.2   The multilayers.py module

The multilayers.py module defines two classes. The `Medium` class generates a convenient object that stores the information about the optical characteristics of a given

---

[3]End User License Agreement
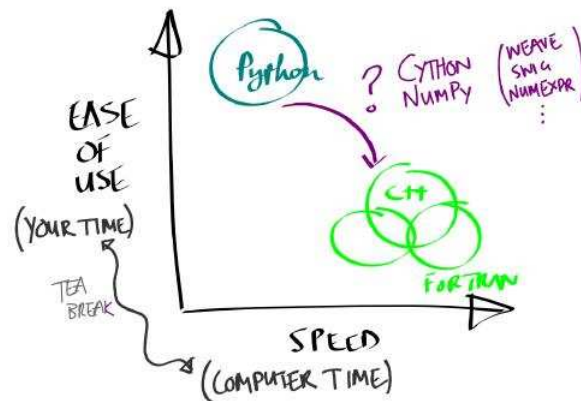
Figure 3.5: Speed versus ease of use of some popular programming languages. Extracted from a lecture given by Stéfan van der Walt during the *Scientific Programming in Python* summer school in Kiel, Germany, 2012.

medium, i.e its complex refractive index as a function of the wavelength. The table of wavelength, refractive index and extinction coefficient triplets is read from a plain text file. The program then automatically generates an interpolating function to the refractive index and the extinction coefficient, and stores the minimum and maximum interpolable wavelengths in instance attributes.

The decision of using an interpolator instead of the raw table comes with both advantages and drawbacks. In principle, since the program only needs to fetch the refractive indices, a dictionary with the wavelengths as keys and the complex refractive indices as values would be very efficient, since the Python dictionaries are implemented as hash tables. Using an interpolator means that there is a function call each time we need to fetch a refractive index.

On the other hand, if the raw table is used we need to make sure that all our tables corresponding to different materials are sampled at the same wavelengths. Different problems will require different samplings, which could be solved by having the tables sampled at very close values of the wavelength and then use only a subset of those points if a coarser sampling is needed. While it is entirely possible to do that, it is far more convenient to just import the table without worrying about the sampling and then be able to perform calculations at any wavelength because the interpolator takes care of everything automatically. To minimize the number of calls to the interpolator, the value of refractive index is calculated and stored in a variable when the wavelength of the system is set. Every time the value is needed, it is fetched from the variable instead of recalculated through a new call to the interpolator. Only when the wavelength of the system is changed the value is recalculated.

The second and most important class defined in the multilayers.py module is the `Multilayer` class. It stores the layout of the system, its state (working wavelength,

Medium			
__minWlength <float>	__maxWlength <float>	__nInterpolator	__kInterpolator

Multilayer

__workingWavelength
<float>

__minMaxWlength
<tuple>

__polarization
<string>

__charMatrixUpDown
<numpy.matrix>

__charMatrixDownUp
<numpy.matrix>

__coefficientsUpDown ⟶ <dict>	'r' <complex>	't' <complex>	'R' <complex>	'T' <complex>
__coefficientsDownUp ⟶ <dict>	'r' <complex>	't' <complex>	'R' <complex>	'T' <complex>

__stack ⟶ <list>	<dict>	<dict>	<dict>	<dict>	<dict>

'medium' <Medium>	'position' <float>	'thickness' <float>	'angle' <complex>	'matrix' <matrix>	'refindex' <complex>
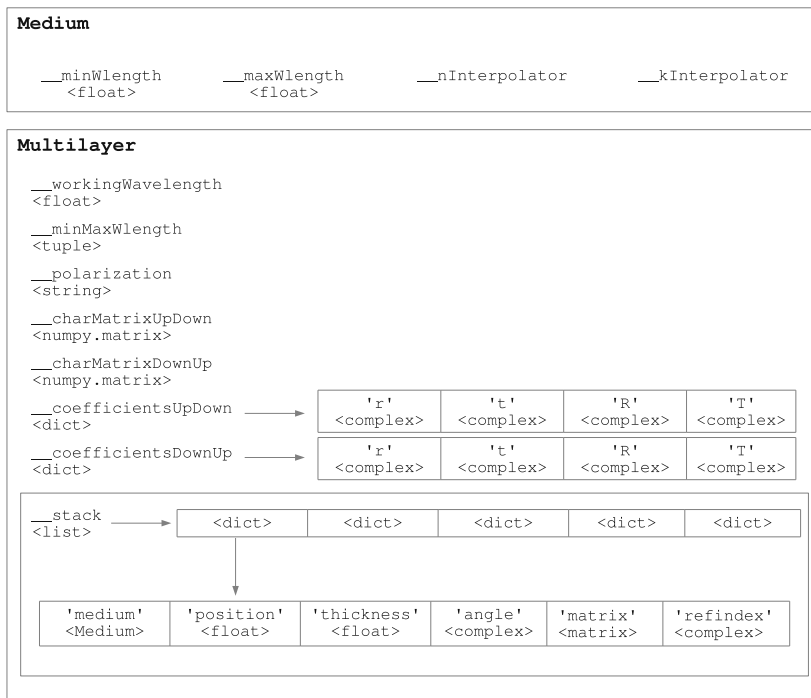
Figure 3.6: Data structures used in the multilayers.py module.

angles of propagation and polarization) and the optical constants for the current state (characteristic matrices and optical coefficients).

The values which are common to the global system are stored in instance attributes, while the values that change for each individual layer of the system (such as the propagation angle) are stored in the __stack list. See figure 3.6 for a diagram of the data structure.

The user loads a set of materials and builds the multilayer. Then the state of the system is specified, i.e. polarization, wavelength and propagation angle. The propagation angle is specified in any of the layers and then the angle in all the other layers is automatically calculated according to Snell's law. After the state is set, the user calculates the characteristic matrices of each layer and the global characteristic matrix of the system for both directions of propagation (up–down and down–up) using a couple of method calls. At the same time, the optical coefficients are automatically calculated from the characteristic matrix and they can be fetched with a method call that will return the dictionary where they are stored (see figure 3.6). Note that even if the angles and refractive indices are real, the program always stores them as complex numbers. This is because, if we work beyond the critical angle or have a metallic layer, the angle must be complex, and of course the refractive
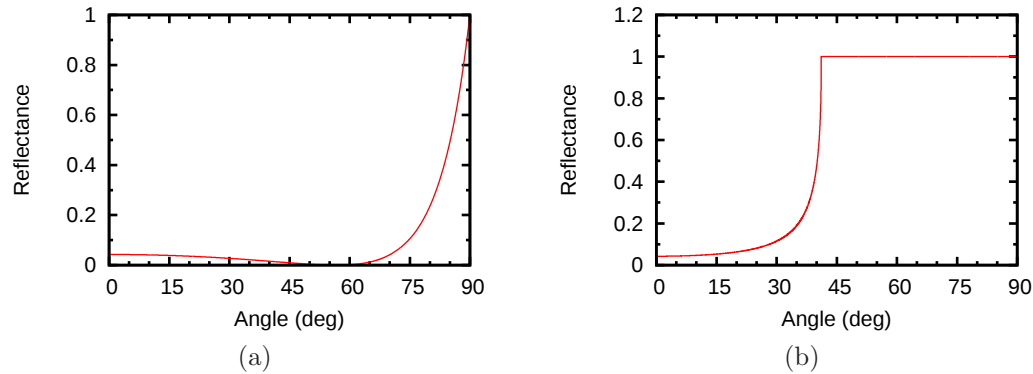
Figure 3.7: (a) Reflectance of the air/dielectric interface, note the Brewster angle; and (b) reflectance of the dielectric/air interface, note the critical angle. The dielectric has refractive index 1.52,

index in metallic layers will be complex too.

The F function discussed in section 3.2.2 can be calculated with a method call. The state of the system is automatically set according to the arguments given to F. Since the calculation of F requires the optical coefficients corresponding to different subsystems of the global multilayer (see equations (3.119), (3.121) and (3.122)), the program automatically builds the necessary subsystems and calculates the corresponding optical coefficients. Once all the required coefficients are known the F function is calculated.

Besides the F function, the optical coefficients of any layered system can be calculated, as already discussed. In particular, the simplest system can be defined: only the upper and bottom medium, which corresponds to a single interface. Therefore, quantities such as the Brewster angle or the critical angle can be calculated. Of course, it is far easier to calculate those angles by simply using the formulae, but this shows that the program is quite general in its scope. As an example, in figures 3.7(a) the reflectance of the interface between the air and a dielectric layer with refractive index 1.52 is calculated for TM waves. The Brewster angle appears at 56.66 degrees as expected. In figure 3.7(b) the same has been done for the interface between the dielectric layer and the air, and the critical angle appears at the expected value of 41.14 degrees.

If a system with absorbing layers is defined, the reflectance and the transmittance of the system can be easily calculated. As an example, figure 3.8 shows the reflectance and the absorbance of a polysilicon layer embedded in air as a function of its thickness for light with wavelength 500 nm.

Of course, there are multiple programs that can do this. The real usefulness of the approach implemented here is that the effect of the multilayer system on the light can be calculated through the F function for light generated *within any* of the layers at *any particular position*, not only for light coming from infinity and
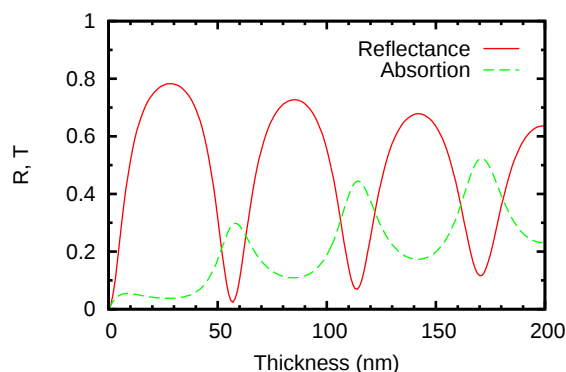
Figure 3.8: Reflectance and absorption of a polysilicon layer embedded in air for wavelength 500 nm. At this wavelength the refractive index of the polysilicon layer is $n \approx 4.4$ and its extinction coefficient $k \approx 0.1$.

incident on the first interface of the system. Results involving the calculation of the F function can be found in section 3.4.

For more information on the available method calls, see the reference manual included with the program. Appendix B shows how to install and use the module in custom Python scripts. The `examples` directory in the package contains example Python scripts that perform typical useful calculations. They can be easily customized to adapt them to any particular system.

3.4 General results

Figure 3.9 depicts the system we will be concerned with during this section. The system consists of a polysilicon gate about 300 nm in thickness and a 50 nm silicon dioxide or silicon rich oxide (SRO) layer on top of a silicon substrate.

In this section we will be interested in the results of the calculation of $F(z; \lambda, \theta)$ in different circumstances. However, what we will show are the results of calculating $|F_y|^2$ for TE waves and $|F_x \cos^2 \theta + F_y \sin^2 \theta|^2$ for TM waves, as already discussed, since these quantities refer to the ratio of the *energy* (rather than the *electric field*) that an observer receives with the layered system present to the energy he would receive in case the layered system was removed. We will refer generally to these quantities as *energy ratio*.

3.4.1 The energy ratio vs. the position of the radiative center

Let's study the dependence of the energy ratio on the position (z coordinate) of a single radiative center.

Air
z = 350 nm

Polysilicon

z = 50 nm
SiO_2 / SRO
z = 0 nm
Silicon
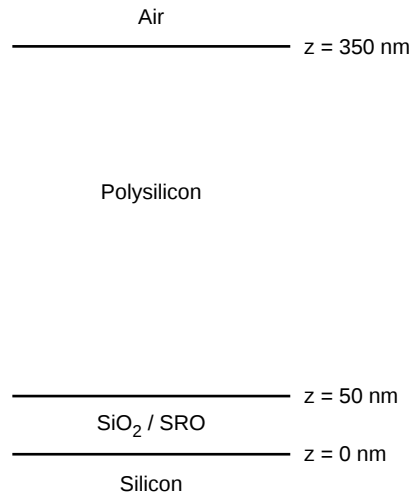
Figure 3.9: Schematic view of the layered system studied in section 3.4.

3.4.1.1 Effect of the wavelength

Figure 3.10 shows the typical standing wave behavior we expect due to interference effects. The period of the standing wave changes depending on the layer and the wavelength. For simplicity we study the case where the angle is zero (propagation normal to the interfaces) and therefore the TE and TM modes are equal. The layer between the polysilicon and the silicon is SiO_2 in this case. Since the SiO_2 layer is very thin, we don't get to see the first maximum within that layer. It can be seen that the radiative centers located close to the interface with the polysilicon will contribute more to the luminescence than those located close to the silicon substrate (that is, if all the radiative centers are considered to have the same intrinsic emission).

3.4.1.2 Effect of the angle

Figures 3.11(a) and 3.11(b) show how the angle affects the energy ratio for TE and TM waves, respectively.

If we concern ourselves with the SiO_2 layer it is quite apparent that for TE waves the energy ratio decreases as the output angle increases. For TM waves the behavior is not as straightforward. It is worth noting that for wide angles TM waves benefit from a higher energy ratio than TE waves. Moreover, notice the discontinuities at the interfaces.
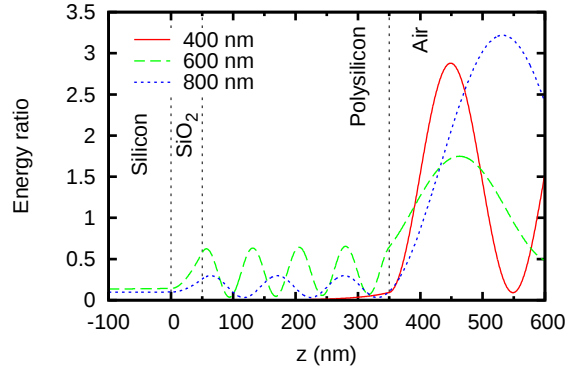
Figure 3.10: Emission perpendicular to the interfaces at different wavelengths (TE and TM modes are equal).
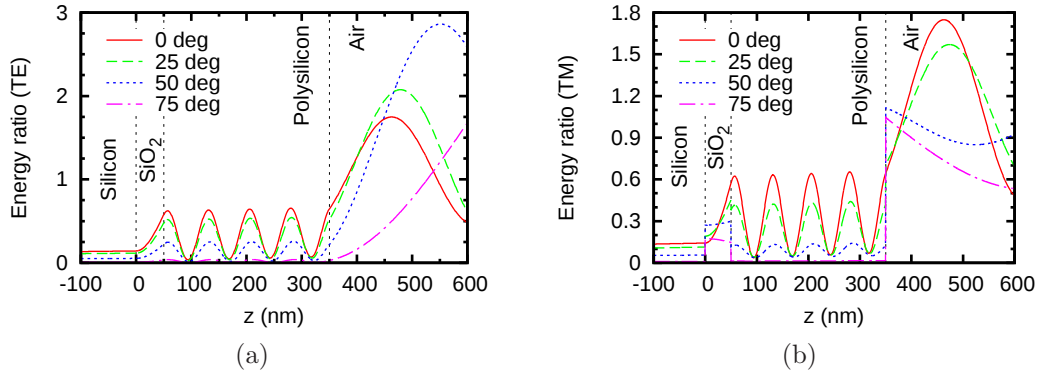
Figure 3.11: Emission at 600 nm with different exit angles for (a) TE and (b) TM waves.

3.4.1.3 Silicon dioxide vs. silicon rich oxide

The difference in the refractive indices of SiO_2 and SRO are shown in figure 3.12. In this case, the SRO was fabricated from a 40 nm thick thermally grown SiO_2 layer. The layer was then implanted with silicon ions with a dose of 3×10^{16} cm^{-2}. The refractive index and extinction coefficient as a function of the wavelength was measured by P. A. Postigo's group at the Instituto de Microelectrónica de Madrid.

As far as the energy ratio is concerned, the difference between stoichiometric SiO_2 and SRO is minimal, as can be seen in figure 3.13. At a wavelength of 500 nm the difference in the refractive indices is ≈ 0.08, but that hardly makes any difference in the calculated values of $|F_y(z)|^2$. That means that the typical difference in refractive index that we can expect from different implantation doses is essentially irrelevant to the simulations, since other parameters have a much greater influence on the energy ratio.
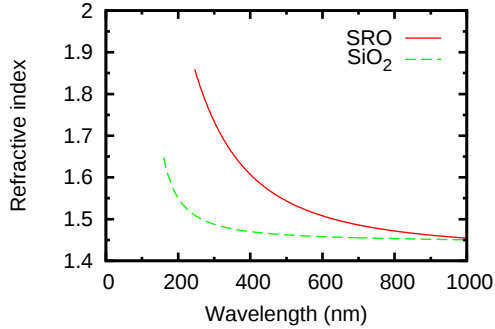
Figure 3.12: Comparison of the refractive indices of SiO_2 and silicon ion implanted SRO with an implantation dose of 3×10^{16} cm^{-2} on a layer with thickness 40 nm.
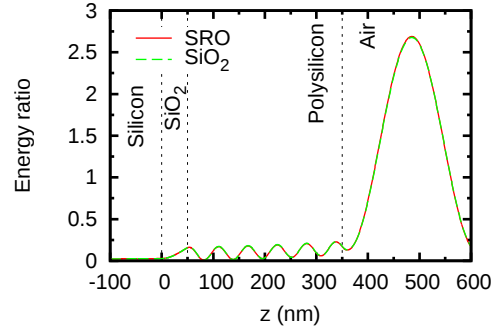
Figure 3.13: Energy ratio for silicon dioxide and SRO at an output angle of zero degrees and for wavelength 500 nm.

3.4.2 The energy ratio vs. the output angle

In this section we will restrict our analysis to radiative centers located within the oxide layer. In section 3.4.1 we saw the effect of the angle on the energy ratio, but for simplicity we only plotted the curves for a few angles. In the present section we will fix the position of the radiative center and the wavelength of light, and see how the energy ratio changes when the exit angle is increased.

In figure 3.14 we present the change of the energy ratio with the exit angle when the z coordinate of the radiative center is fixed at three different positions: close to the silicon substrate ($z = 0.1$ nm), in the middle of the oxide layer ($z = 25$ nm) and close to the polysilicon layer ($z = 49.9$ nm). The wavelength was fixed at 600 nm and the results are presented for both TE and TM waves.

Confirming what we could surmise from section 3.4.1.2, figure 3.14 shows that, for TE waves, the behavior is quite straightforward, as the energy ratio is monotonically decreasing with the angle. Moreover, radiative centers located closer to the polysilicon benefit from a higher energy ratio at all angles.

For TM waves the dependence of the energy ratio on the angle exhibits a maximum for radiative centers located up to 25 nm from the silicon substrate. For higher distances the dependence is monotonically decreasing. Moreover, only for angles smaller than ≈ 57 degrees it is true that radiative centers closer to the polysilicon interface show a higher energy ratio. For angles above ≈ 57 degrees the inverse is true, although the difference is small.

3.4.3 The energy ratio vs. the wavelength

In the present section we will discuss the effect of the wavelength on the energy ratio. Propagation angle and position of the radiative center will be fixed parameters.
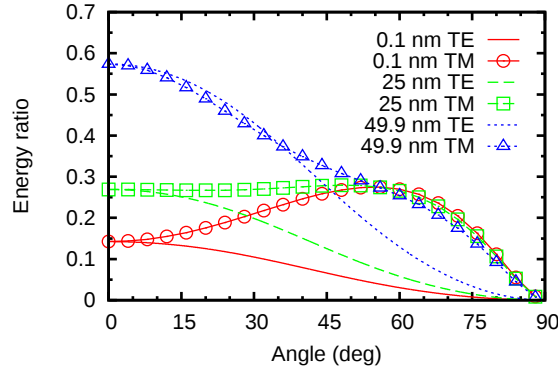
Figure 3.14: Energy ratio as a function of the exit angle for TE and TM waves. The position of the radiative center has been fixed at three different values within the oxide layer: close to the silicon substrate ($z = 0.1$ nm), in the middle of the layer ($z = 25$ nm) and close to the interface with the polysilicon ($z = 49.9$ nm). The wavelength was fixed at 600 nm.
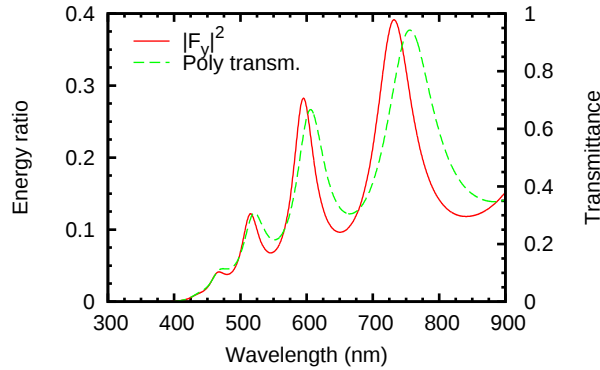
Figure 3.15: Energy ratio for emission at different wavelengths in the system depicted in figure 3.9. The radiative center is located in the middle of the silicon dioxide layer and the propagation direction is normal to the interface. The green dashed line is the transmittance of a 300 nm thick polysilicon layer at normal incidence with oxide as the input medium and air as the output medium.

3.4.3.1 Comparison to the transmittance of a polysilicon layer

In figure 3.15 we have plotted the transmittance of a 300 nm thick polysilicon layer at normal incidence with SiO_2 as the input medium and air as the output medium. For comparison, we have also plotted the energy ratio of a radiative center located in the middle of the SiO_2 layer of the system depicted in figure 3.9 for propagation normal to the interface.

It is clear that the contribution of the transmittance of the polysilicon layer to the energy ratio is important. However, there are some significant differences.
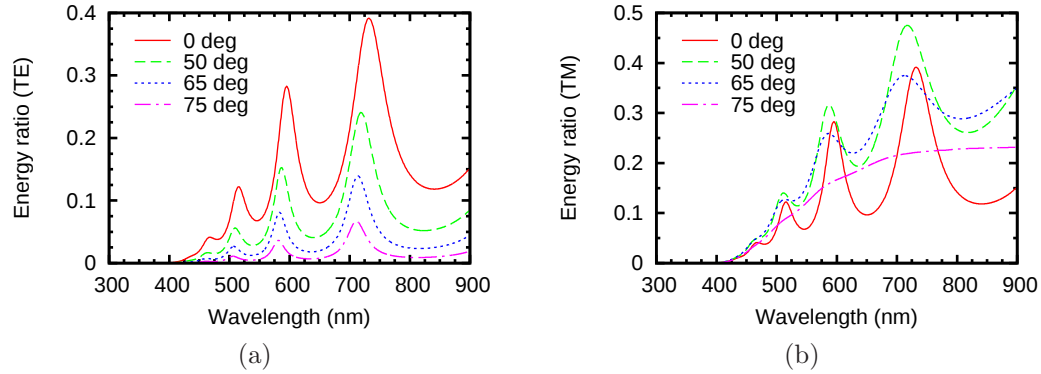
Figure 3.16: Energy ratio as a function of the wavelength at different output angles for (a) TE and (b) TM waves. The radiative center is located in the middle of the SiO$_2$ layer.

First, notice that, while the transmittance reaches a maximum value close to 1, the maximum value of the energy ratio is less than 0.4. Moreover, the peaks of the energy ratio are redshifted, but the shift is not the same for every peak (it increases with the wavelength). The relative height of the peaks is different too.

In conclusion, while the influence of the polysilicon transmittance is obvious, it is clear that the effect of the whole multilayer is significantly different than just the transmittance of its polysilicon layer.

3.4.3.2 Effect of the angle

Figures 3.16(a) and 3.16(b) show the energy ratio as a function of the wavelength for TE and TM waves, respectively, at different output angles. The location of the radiative center is fixed at 25 nm from the silicon substrate (i.e in the middle of the silicon dioxide layer).

For TE waves we had already shown that the energy ratio decreases when the angle is increased (see figure 3.11(a)). It is clear now from figure 3.16(a) that this result is consistent across the whole range of studied wavelengths. Note the redshift of the peaks when the angle is increased. Another significant feature is the widening of the valleys for large angles.

For TM waves the results are again less straightforward. In figure 3.14 (green squares) we had already seen that, for a radiative center in the middle of the silicon dioxide layer, the energy ratio would increase slightly up to ≈ 57 degrees. For larger angles it would decrease. In figure 3.16(b) it is clear that the energy ratio increases in all wavelengths when the angle is increased from zero to 50 degrees. However, when the angle is increased to 65 degrees the peaks become smoother and the global curve seems to decrease. The trend is confirmed after further increasing the angle up to 75 degrees. We then see clearly that the peaks have been smoothed out and
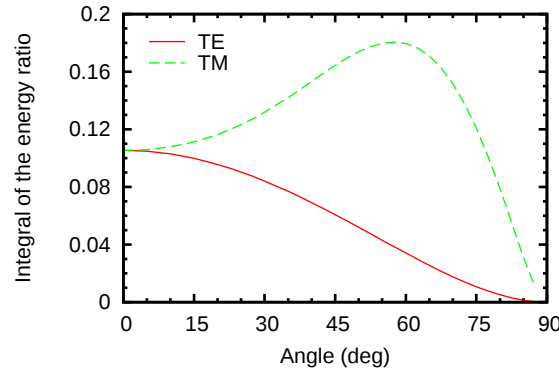
Figure 3.17: Integral of the energy ratio over the wavelength as a function of the propagation angle. The position of the radiative center is in the middle of the SiO_2 layer.

the energy ratio is globally lower.

It would be interesting to see how the integral of the curves shown in figures 3.16(a) and 3.16(b) evolves with the angle. This is shown in figure 3.17. The plots confirm the trend that can be surmised from figures 3.16(a) and 3.16(b). For TE waves the total amount of energy decreases when the angle is increased, while for TM waves it increases until it reaches a maximum at ≈ 57 degrees. This confirms that the trend we observed in figure 3.14 for a fixed wavelength of 600 nm can be extended to the integral of the energy ratio over the whole spectrum.

Notice that in the most favorable case the integral is ≈ 0.18, which means that we get as little as 18% of the energy we would get if we had no multilayer in place.

3.4.3.3   Effect of the position

In figure 3.18 we show that the energy ratio increases consistently in all wavelengths when the position of the radiative center is moved from close to the silicon substrate up to the polysilicon interface, as we already hinted in figure 3.10. From figure 3.14 we also know that this happens for all angles in TE waves. From the same figure we know that the trend is reversed for TM waves at angles larger than ≈ 57 degrees. Figure 3.19 confirms that this is valid for all wavelengths, although the energy ratio decreases very slightly.

3.4.3.4   Effect of the thickness of the active layer and polysilicon layer

Since the peaks observed in the plots of $F(\lambda)$ (figures 3.15, 3.16(a), 3.16(b), 3.18 and 3.19) arise from interference effects due to multiple reflections at the various interfaces of the system, we expect the position of the peaks to depend on the thickness of the layers.
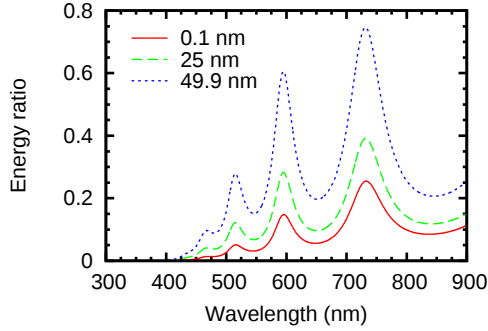
Figure 3.18: Energy ratio as a function of wavelength for different positions of the radiative center. Propagation is normal to the interfaces.
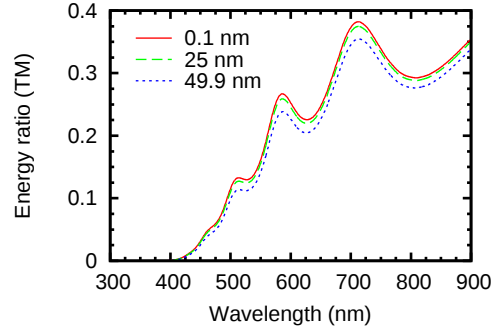
Figure 3.19: Energy ratio as a function of wavelength for different positions of the radiative center. Propagation is at 65 degrees and the results are shown for TM waves.
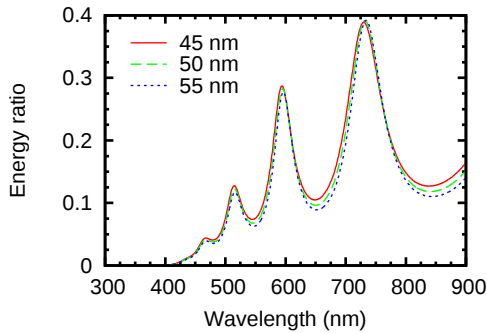
(a)
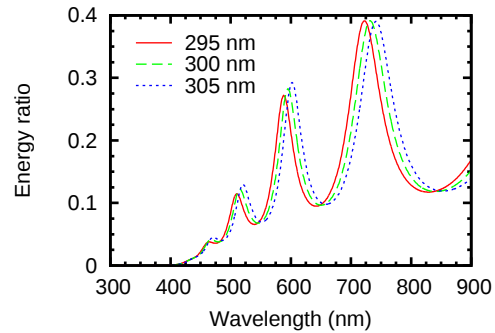
(b)

Figure 3.20: Variation of the energy ratio vs. the wavelength for three different thicknesses of (a) the active layer, with the polysilicon thickness fixed to 300 nm, and (b) the polysilicon layer, with the thickness of the active layer fixed to 50 nm. In all cases the propagation is normal to the interfaces and the radiative center is located at $z = 25$ nm.

In figure 3.20(a) we show the variation of the energy ratio for three thicknesses of the active layer. The thickness of the polysilicon layer is fixed at 300 nm. The positions of the peaks redshift about 4 nm when the thickness of the active layer is increased from 45 to 55 nm, although the amount of redshift is not constant across all wavelengths (it is slightly smaller for shorter wavelengths)

If we keep the thickness of the active layer fixed at 50 nm and change the thickness of the polysilicon layer from 295 to 305 nm we observe a redshift of about 20 nm in the position of the peaks, as shown in figure 3.20(b). Again, the amount of redshift is not constant but decreases slightly for shorter wavelengths.

According to these results, the system is not overly sensitive to small variations of the thickness of the active layer. However, the dependence of the position of the peaks on the thickness of the polysilicon layer is strong and therefore its precise knowledge is important.

3.4.4 Effect of the distribution of the radiative centers in the active layer

So far we have studied how the spectrum of a single radiative center is modified by the multilayer, and we have seen that it depends, among other parameters, on the position of the radiative center in the multilayer. However, in a real system we will not have a single radiative center or a two dimensional sheet of radiative centers at a given z coordinate, but rather a continuous distribution of radiative centers along the z axis. In this section we will study how different distributions of radiative centers within the SiO_2 layer affect its spectrum.

Throughout this section, we will consider that all the radiative centers have the same intrinsic emission spectrum regardless of their position. The intrinsic emission energy at different z coordinates will be adjusted through the distribution $g(z)$.

Let's start by considering the distribution $g(z)$ to be a Gaussian (mean μ and standard deviation σ) centered at the Si/SiO_2 interface ($\mu = 0$), at the middle of the SiO_2 ($\mu = 25$) and at the $SiO_2/poly$ interface ($\mu = 50$), with varying values of σ. We will also study the case of a uniform distribution.

Note that the distribution of the radiative centers can be interpreted as the distribution of the intensity of the emission and we want that to be normalized to area 1 in the given interval. Plots of the studied Gaussian distributions are shown in figures 3.21(a), 3.21(b) and 3.21(c).

In figure 3.22 we show the energy ratio calculated assuming that the radiative centers are distributed following the Gaussian distributions of figure 3.21(b), centered in the middle of the active layer. The results have been normalized in order to highlight possible differences in the shape of the energy ratio. In all cases the propagation is normal to the interfaces.

It is clear from figure 3.22 that the width of the distribution has very little impact on the shape of the energy ratio. This conclusion is also valid when the normal distributions are centered in the Si/SiO_2 and $SiO_2/polysilicon$ interfaces (plots not shown).

A uniform distribution of radiative centers across the active layer could be considered as a limit case of a Gaussian distribution with a very wide standard deviation, and therefore we expect little change when compared with a sharper Gaussian distribution. This is confirmed in figure 3.23.

Now we will fix the standard deviation and change the center of the normal distribution. The results are shown in figure 3.24 for $\sigma = 7$ and center in the middle of the active layer, at the Si/SiO_2 interface and at the $SiO_2/polysilicon$ interface. Propagation is normal to the interfaces and the results have been normalized to 1.

(a) Center at the Si/SiO$_2$ interface ($\mu = 0$ nm).

(b) Center in the middle of the SiO$_2$ ($\mu = 25$ nm).

(c) Center at the SiO$_2$/polysilicon interface ($\mu = 50$ nm).
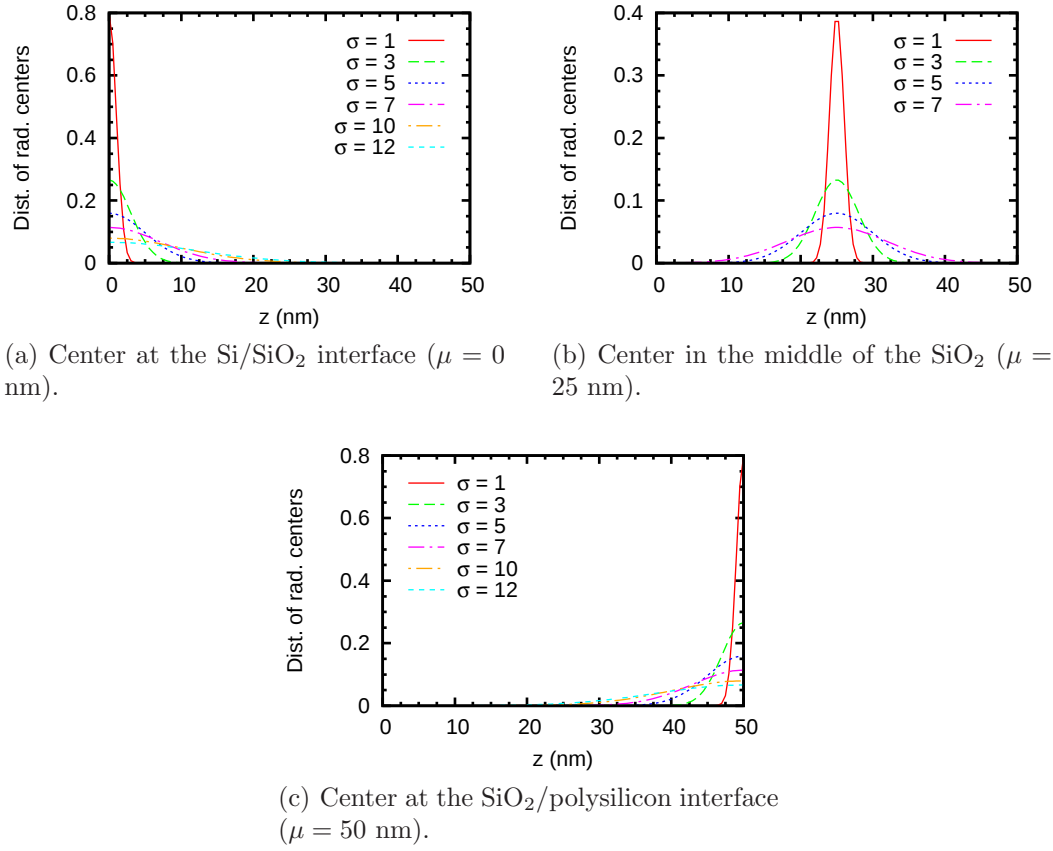
Figure 3.21: Gaussian distributions of the radiative centers, with the center μ of the distribution located at different positions of the SiO$_2$ layer, and with different deviations σ. The area under the curve is always 1.

It is clear that the position of the Gaussian has a significant effect on the shape of the energy ratio. The dominance of the highest peak increases as the center of the distribution moves deeper in the active layer.

3.4.5 Antireflective coating

We can use the our implementation of the Crawford method to calculate the optimal thickness of an antireflective coating deposited on top of the polysilicon gate. What we need to do is calculate $F(\lambda)$ for the given structure, including the antireflective coating, and then calculate the integral of $F(\lambda)$. The calculation is repeated for different thicknesses of the antireflective coating until we find the value that maximizes the integral.

For the following results, our basic structure will be as the one depicted in figure 3.9, plus the antireflective coating on top of the polysilicon. We will examine the
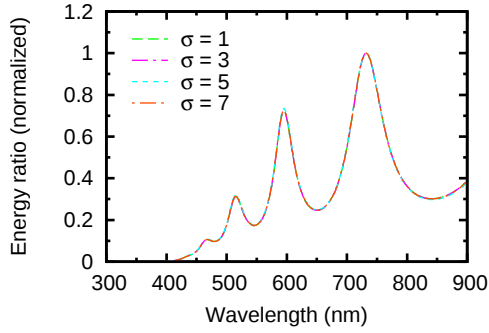
Figure 3.22: Normalized modification to the spectrum due to a Gaussian distribution of the radiative centers. The Gaussian is centered in the middle of the active layer and the results are plotted for different values of σ. Propagation is normal to the interfaces.
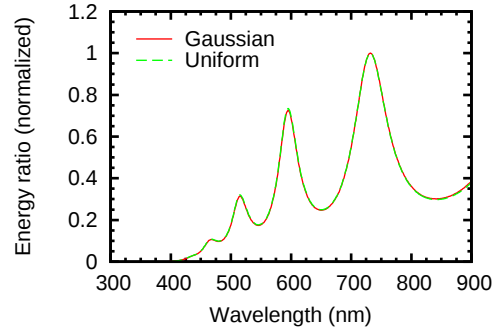
Figure 3.23: Normalized modification to the spectrum due to a Gaussian distributions with $\sigma = 7$ centered in the middle of the active layer and a normal distribution. Propagation is normal to the interfaces.
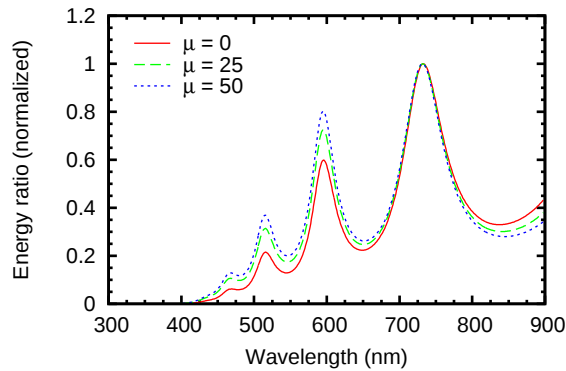
Figure 3.24: Normalized modification to the spectrum due to a Gaussian distributions of the radiative centers. The Gaussian distribution has $\sigma = 7$ and the results are plotted for different values of μ in the active layer. Propagation is normal to the interfaces.
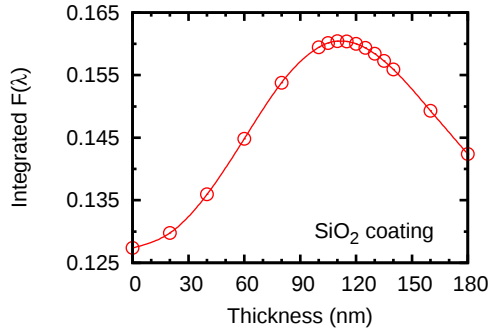
Figure 3.25: Values of the integral of $F(\lambda)$ as a function of the thickness of a SiO_2 layer on top of the polysilicon gate. The larger the value of the integral, the better the overall transmission.
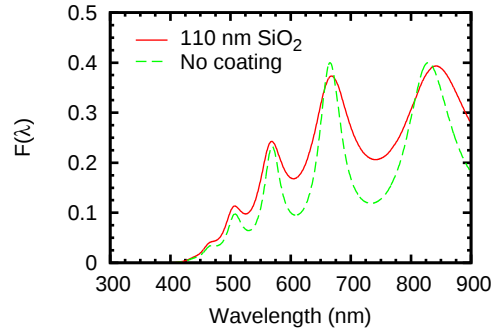
Figure 3.26: Comparison between $F(\lambda)$ with the optimal SiO_2 antireflective coating, and without it. Clearly, the energy is gained at the wavelengths corresponding to the valleys of F.

results for three antireflective coatings: a single SiO_2 layer, a single Si_3N_4 layer, and a $SiO_2 + Si_3N_4$ bilayer. For the calculation of $F(\lambda)$, the luminescence is considered to be uniform across the SRO layer and examined for the direction normal to the interfaces ($\theta = 0$). Note that we are not making any assumption regarding the spectrum of the emission, only that it is independent of the position in the SRO layer.

Figure 3.25 plots the integral of $F(\lambda)$ as a function of the thickness of a SiO_2 antireflective coating, in the conditions stated above. The maximum value of the integral, 0.160, corresponds to a thickness of 110 nm of the SiO_2 coating. Compared to the value 0.127 corresponding to the structure without the SiO_2 coating, that is an increase of roughly 25% in the total transmitted energy across the whole spectrum. Figure 3.26 plots $F(\lambda)$ for the structure without the coating and with the optimal 110 nm SiO_2 coating. The increase in the transmitted energy clearly comes from the fact that the peaks in $F(\lambda)$ are somewhat smoothed when the coating is present.

Figures 3.27 and 3.28 plot the same as figures 3.25 and 3.26 but for a Si_3N_4 coating layer instead of a SiO_2 coating. The maximum value of the integral of F is 0.167 for a thickness of 75 nm, which is a 31% increase compared to the structure without the coating. As in the previous case, the improvement of the transmission comes primarily from the partial filling of the valleys between peaks in the $F(\lambda)$.

Finally, we examine the case of an antireflective coating consisting of a Si_3N_4 layer on top of a SiO_2 layer. Figure 3.29 shows in a color map the values of the integral of $F(\lambda)$ for different combinations of the thickness of the two composing layers. With the selected color palette, white corresponds to the value of the integral of $F(\lambda)$ for the structure without any coating. The blue region represents the coatings that would give a worse result than not having a coating at all, whereas the red region represents the values where the coating improves the overall transmission. At the top of the color palette, a yellow region has been added to highlight
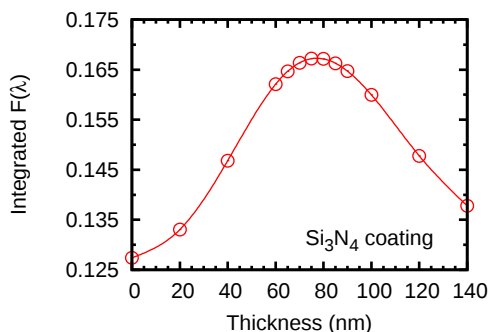
Figure 3.27: Values of the integral of $F(\lambda)$ as a function of the thickness of a Si_3N_4 layer on top of the polysilicon gate.
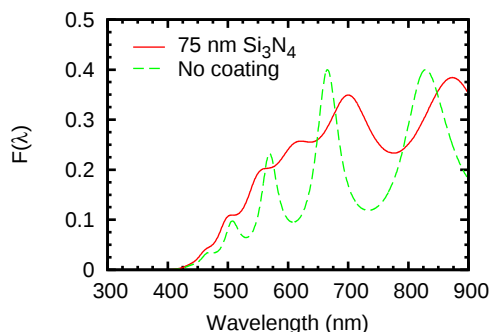
Figure 3.28: Comparison between $F(\lambda)$ with and without the optimal Si_3N_4 antireflective coating.
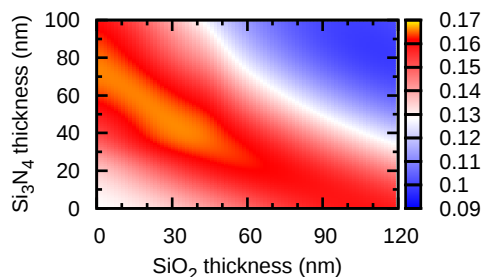
Figure 3.29: Color map representing the integral of $F(\lambda)$ for different thicknesses of a SiO_2/Si_3N_4 bilayer coating. The white band corresponds to the value for the uncoated system. Blue represents regions with worse overall transmission, whereas red represents regions with better overall transmission.
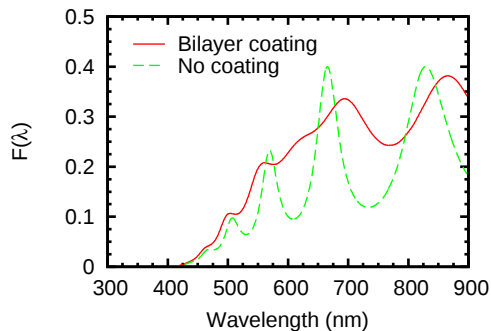
Figure 3.30: Comparison between $F(\lambda)$ with the optimal 20 nm $SiO_2/55$ nm Si_3N_4 bilayer coating, and without it.

the peak region. The best result is achieved for 20 nm of SiO_2 and 55 nm of Si_3N_4. This corresponds to a value of 0.168 of the integral, which is a 32% increment with respect to the uncoated structure. Figure 3.30 compares $F(\lambda)$ with and without the optimal coating.

Since the best increase in the overall transmission for the bilayer system is only marginally better than the value obtained for a single Si_3N_4 layer, the latter option is more convenient from the fabrication perspective.

It is worth noting that our analysis makes sense if the spectrum is wide. If we had a very narrow emission around 660 nm for instance, according to figure

Sample	Layers	Nominal thick. (nm)
	Air	∞
Ipl-3	SRO	40
	Silicon	∞
	Air	∞
Ipl-6	SRN	40
	Silicon	∞
	Air	∞
	Poly	350
Iph-1	SRN	30
	SRO	30
	Silicon	∞
	Air	∞
	SRN	30
Iph-9	SRO	30
	Silicon	∞

Table 3.1: List of the relevant parameters of the samples used to compare against the simulations. The silicon substrate is considered a semi-infinite medium in the simulations, although the wafers are 500 μm thick.

3.30 we would be better off without the coating, since the peak of relatively high transmission precisely at 660 nm gets actually lower with the coating. However, it is clear that, for our typically wide emission bands, the inclusion of the optimal coatings is beneficial to the light extraction.

3.5 Comparison to experimental results

In this section we will try to asses the validity of the simulations by comparing the experimentally obtained spectra with the results of the simulations. Table 3.1 lists the relevant parameters of the samples that will be studied in this section.

3.5.1 Photoluminescence measurements in single active layer systems

Since the intrinsic spectrum of the radiative centers is unknown, we usually compare the EL spectrum to the PL spectrum in search of similarities, and assume that the PL spectrum is very similar to the intrinsic emission. It is not far fetched to assume that the intrinsic emission should be the same in EL and PL experiments, and that the differences are due to the fact that the multilayer stack is different in both experiments, since we add a polysilicon layer in order to carry out EL measurements.
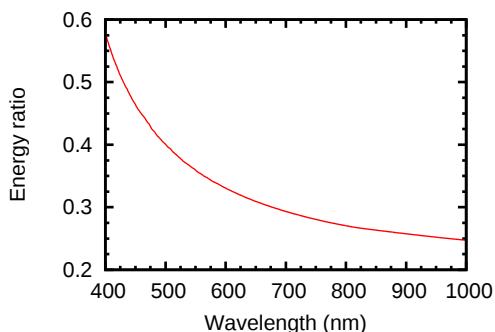
Figure 3.31: Energy ratio as a function of the wavelength for the Si/SRO/air system of sample Ipl-3.
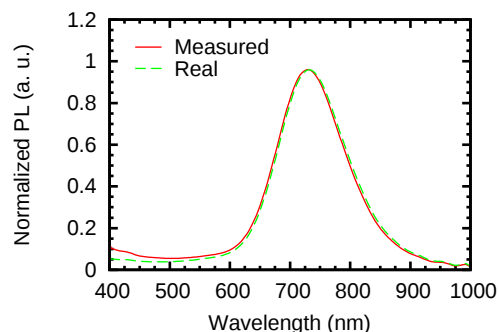
Figure 3.32: Comparison between the real (calculated) and measured PL spectrum of sample Ipl-3.

The first thing we want to do is to assess the effect that the layered system has on the PL measurements, to see how similar the PL is to the intrinsic emission.

For a system composed of a silicon substrate and a SRO layer, sample Ipl-3 (see table 3.1), and assuming a uniform distribution of the radiative centers in the SRO layer and emission normal to the interface, the energy ratio as a function of wavelength is plotted in figure 3.31. The refractive indices used for the calculation were measured in the same sample.

It is clear that the effect at wavelengths longer than 500 nm will be very small. If we eliminate the effect introduced by the multilayer on the PL of sample Ipl-3 (figure 3.32) we can confirm that measured spectrum and the calculated real spectrum are very similar. That means that we can consider the PL measurement to be close enough to the real one in SRO samples with emission typically above 600 nm.

A similar analysis has been carried out in the same conditions for the sample Ipl-6, which consists of a silicon substrate with a SRN active layer, see table 3.1. The energy ratio as a function of the wavelength for this system is plotted in figure 3.33

Even though the effect is similar to that of the previous sample, in this case the effect on the measured spectrum will be more significant, because most of the emission in SRN layers is located in the blue region of the spectrum, where the energy ratio changes more rapidly. As a result, there is a noticeable difference in the location of the peak emission between the measured and the calculated intrinsic spectrum, as shown in figure 3.34.

3.5.2 Photoluminescence measurements in bilayer systems

In this section we will study the SRO/SRN bilayer system of samples Iph-1 and Iph-9 (see table 3.1). We know that the total emission is the result of the emission in the SRO layer and the emission in the SRN layer. This is clear from the measured PL spectra (figure 3.35) and the typical emission of SRN and SRO layers seen in
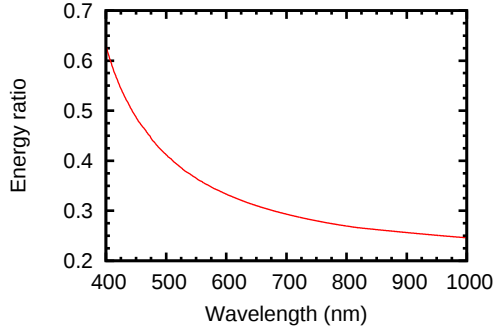
Figure 3.33: Energy ratio as a function of the wavelength for the Si/SRN/air system of sample Ipl-6.
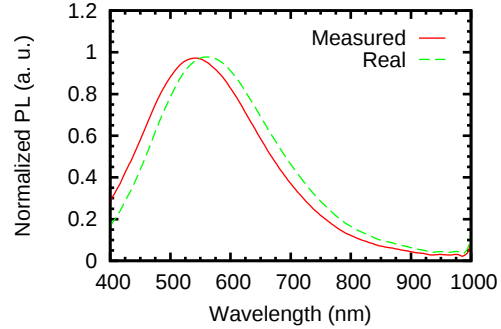
Figure 3.34: Comparison between the real (calculated) and measured PL spectrum of sample Ipl-6.

section 3.5.1 (figures 3.32 and 3.34).

In figure 3.35 we plot the PL of the SRO/SRN bilayer of sample Iph-9 (red dots) along with the components due to the SRO (green dashed line) and SRN (blue dotted line) layers as extracted from a multigaussian fit. The dashed-dotted line is the sum of the two components, which we can see fits well the overall emission.

If we proceed for each component of the emission as we did in section 3.5.1 for a single layer, we can find the intrinsic emission of the SRO and SRN layers. Figure 3.36 shows the calculated real emission for each band. Notice that the change is significant. The band corresponding to the SRO does not change much regarding the shape, but the intensity does change significantly. On the other hand, the band corresponding to the SRN shifts noticeably although the amount of energy is roughly the same. In figure 3.37 we plot the compound spectrum as measured (as in figure 3.35) and after removing the effect of the multilayer, that is the sum of the calculated real SRO and SRN bands shown in figure 3.36). Note that the difference between the measured PL and the calculated intrinsic emission is very significant.

Now that we know the real spectrum of each band and their energies relative to one another, we can calculate the modification experienced by each band when they are being generated in the system used for the EL measurements, i.e the Si/SRO/SRN/poly/air system of sample Iph-1. Figure 3.38 shows the results of the calculation considering a uniform distribution of the radiative centers in each layer (see the discussion in section 3.2.2.5) and using the experimentally determined values of the thickness of each layer. It can be seen that the SRN layer will be essentially the sole responsible for the emission up to 600 nm. From 600 nm to 700 nm we have significant contributions from both the SRO and SRN layers. From 700 nm and longer wavelengths, the SRO is responsible for the emission.

Finally, in figure 3.39(a) we show the calculated and experimentally determined spectrum of the Si/SRO/SRN/poly/air system of sample Iph-1. While there are inaccuracies, the calculated spectrum is a reasonably good approximation to the experimental curve.
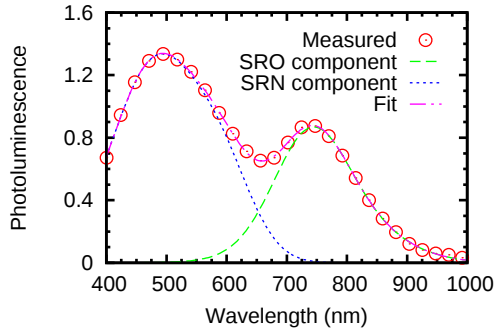
Figure 3.35: Measured PL spectrum of a SRO/SRN system of sample Iph-9 and its decomposition in two bands, one originating in the SRO layer and the other in the SRN layer.
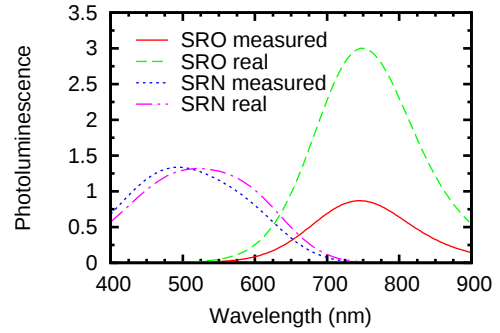
Figure 3.36: Change in the SRO and SRN PL bands of sample Iph-9 due to the effect of the layered system.
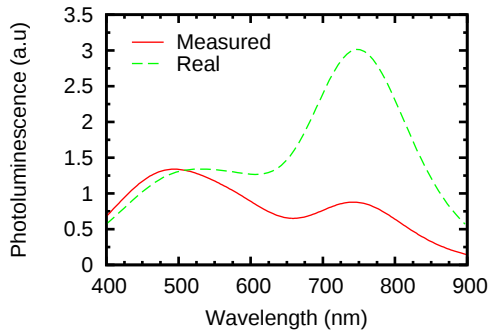
Figure 3.37: PL spectrum of the SRO/SRN system of sample Iph-9 compared to the calculated real spectrum after removing the effect of the multilayer.
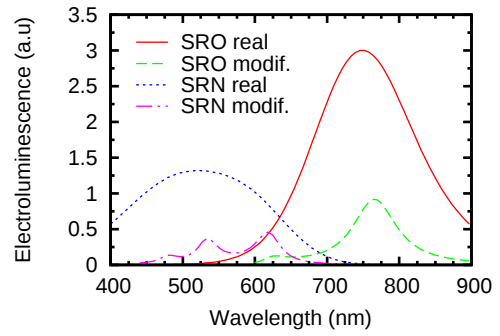
Figure 3.38: Modification of the SRO and SRN bands by the layered system used in the EL measurements.

There are two main discrepancies between the two curves: the peaks appear shifted and the peaks of the SRN contribution seem to be overweighted in the simulated spectrum.

Regarding the shift in the peaks, we saw in section 3.4.3.2 that the angle of propagation affects their position. Since we measure both the PL and EL in the direction perpendicular to the interfaces, the angle has been considered zero in all the simulations. Therefore, we can disregard any consideration as to the TE and TM components of the emitted light, since they are the same at normal propagation. However, the optical system used in the experimental measurements collects light from a considerably wide solid angle. Hence, the experimental spectrum carries contributions from different angles, each of them with unknown contributions from
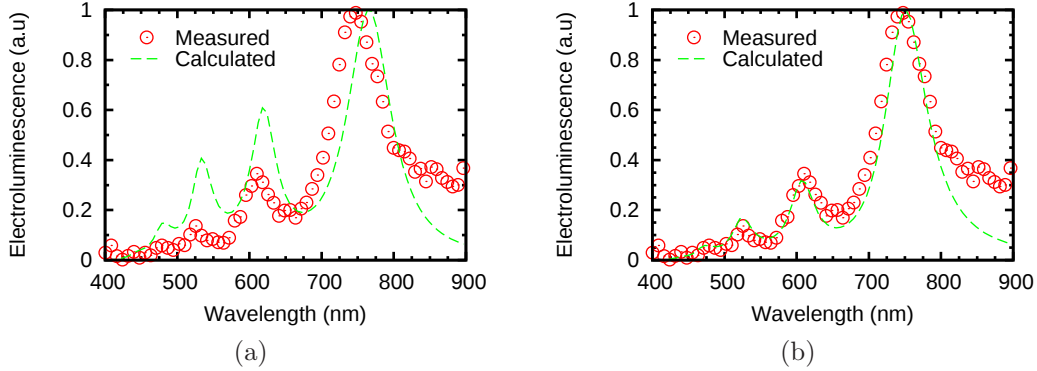
Figure 3.39: Comparison between the experimental and calculated EL spectrum of the SRO/SRN bilayer system of sample Iph-1. In (a) we use the measured polysilicon thickness and a uniform distribution of radiative centers in each layer. In (b) we have adjusted the polysilicon thickness and the distribution of radiative centers (see text).

TE and TM waves. Future work includes allowing for those contributions from different angles in the same way that we allow for contributions from emission at different positions in the layer, as seen in section 3.4.4.

Another factor affecting the position of the peaks is the thickness of the layers, particularly that of the polysilicon gate, as we discussed in section 3.4.3.4 (see figures 3.20(a) and 3.20(b)). This point is critical, since a reduction of 10 nm in the thickness of the polysilicon layer would be enough to eliminate the difference in the position of the peaks observed in figure 3.39(a). Unfortunately, the precise measurement of the thicknesses in multilayered systems is not an easy task. In particular, it involves previous knowledge of the thickness of the underlying layers. The thickness of the thermally grown oxide layer can be measured before the implantation. With that as a datum, the thickness of the silicon nitride layer can be measured. However, it is known that after the implantation and annealing the thickness changes [66]. Measuring after the annealing necessarily involves measuring a bilayer of which only the approximate values are known from the measurements previous to the annealing. Therefore, the measurement of the thickness of the polysilicon layer is based on the values of thickness of the underlying bilayer that are only known approximately. Moreover, as we already saw in section 2.2.2.3 of chapter 2, there is an oxynitride between the SRO and the SRN, which makes the task even more difficult.

Regarding the seemingly overweighted contribution of the SRN emission, consider that we assume that the distribution of the total radiated energy in the SRO and SRN band is the same when pumping the system optically and electrically. However, it is reasonable to assume that this will not be the case in general. In the EL measurements, the samples are being biased in the inversion regime. That means that electrons flow from the substrate to the SRO layer. It seems reasonable

to assume that the radiative recombinations will be more frequent in the SRO layer than in the SRN layer. Therefore, the relative proportion between the energy in the SRO and SRN bands will be more favorable to the SRO when pumping electrically than it is when pumping optically. In figure 3.24 we also saw that the position of the radiative centers clearly affects the relative intensity of the peaks. Therefore, a precise knowledge of the distribution of the emission within the layer is important if we want to achieve accurate results.

In order to better match the calculated spectrum to the experimental one, let us recalculate the EL spectrum as in figure 3.39(a), but this time we will consider the thickness of the polysilicon layer as a variable and assume a distribution of radiative centers of the form:

$$g(z) = a \, e^{\frac{-z}{d}} \tag{3.149}$$

We will use the normalized spectrum of each band, and $g(z)$ will also be normalized to area 1 in the range of interest (see discussion in section 3.2.2.5). Since we impose the integral of $g(z)$ to be unity, the constants a and d are not independent of each other. The reason for using a distribution of this kind is that it seems reasonable to expect such a distribution for the density of electrons in the stationary state, and the recombination rate should be proportional to the density of electrons in the active layer.

The best match between the simulation and the experimental measurement is shown in figure 3.39(b), which corresponds to $d = 21$ nm and a 305 nm thick polysilicon layer. The measured thickness of the polysilicon layer oscillates between 307 and 314 nm depending on the sample of fabrication Iph, although, as already discussed, there is a systematic error in all those measurements. Therefore, a value of 305 nm for our simulations is acceptable.

Summarizing, in this section we have calculated a remarkably good approximation to the EL spectrum of the Si/SRO/SRN/poly/air system of sample Iph-1, using as the only input data the PL of the corresponding Si/SRO/SRN/air system of sample Iph-9. We have seen that inaccurate information about the distribution of the emitted energy within the system, the approximation of collection normal to the interfaces and inaccuracies in the thickness of the layers, in particular that of the polysilicon gate, are the most probable causes of the mismatch between measurement and simulation.

3.5.3   Photoluminescence measurements in active layers with different substrates

In this section we will work with pairs of samples that share the same active layer but differ in the layers below. Comparison of the PL spectra between corresponding pairs shows large differences. We will try to tell if the observed differences are due to interference effects or not.

Figure 3.40 depicts the vertical layout of all the samples that will be studied in this section. The nominal values for the thickness of the layers are shown.
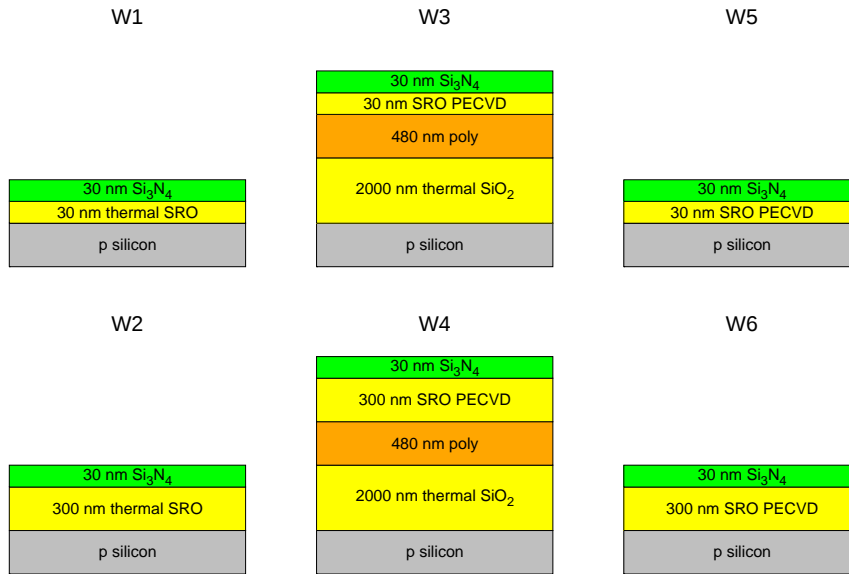
Figure 3.40: Vertical layout of the studied samples. Thicknesses are nominal.

Samples W1 and W2 feature a thermally grown layer of SiO$_2$, with a thickness of 30 nm in sample W1 and 300 nm in sample W2. In samples W5 and W6 the thermally grown silicon dioxide is substituted by a PECVD SRO with a low silicon excess. Samples W3 and W4 feature a thermally grown SiO$_2$ layer with a thickness of 2 μm. On top of that, a polysilicon layer with a thickness of 480 nm is deposited. After that, the same PECVD SRO layer used in samples W5 and W6 is deposited on top of the polysilicon. Finally, an LPCVD SRN layer with a thickness of 30 nm is deposited on all samples. A set of two silicon ion implantations, at 25 and 50 KeV, respectively, with a total dose 1.5×10^{16} cm^{-2} is carried out on all samples in order to introduce the silicon excess. Finally, all samples underwent a thermal annealing at 1240 °C for 60 minutes.

Since samples W4 and W6 share the same active layer, one would expect their PL to be the same. However, the measurements, presented in figure 3.41, show very different results. Using the same process outlined in previous sections, we find the intrinsic emission of sample W6 considering that all the emission comes from the SRO layer. Note that there is emission from the nitride layer in the region 400–600 nm, but its relative intensity is negligible. Then, assuming the intrinsic emission to be the same in sample W4, we calculate the effect of the multilayered structure of sample W4 to find what the PL of that sample should look like. The result is shown as the blue dotted line of figure 3.41.

It is clear that the difference between the measured spectra of samples W4 and W6 is massively dominated by interference effects. The calculation has been carried out considering the following thicknesses: 31 nm for the SRN; 325 nm for the SRO;

Sample	Si$_3$N$_4$ (nm)	SRO (nm)	Poly (nm)	SiO$_2$ (nm)
W1	31 ± 1	31 ± 1		
W2	31 ± 1	304 ± 1		
W3	31 ± 1	32 ± 2	490 ± 3	2032 ± 3
W4	31 ± 1	328 ± 10	490 ± 3	2032 ± 3
W5	31 ± 1	32 ± 2		
W6	31 ± 1	328 ± 10		

Table 3.2: Measured thicknesses of the set of samples being studied in the present section.

430 nm for the polysilicon; 2032 nm for the SiO$_2$. These values are in good agreement with the experimentally measured thicknesses, listed in table 3.2. Note that the thickness of the polysilicon was measured before removing the phosphosilicate glass which is formed during the doping process.

In the previous section we compared PL and EL measurements. Since the collection system is different in both kinds of measurements, we could not directly compare intensities and therefore we always compared normalized spectra. This is no longer the case. In figure 3.41 we are using arbitrary units but directly comparable values. Note that the calculated curve matches well not only the shape of the spectrum but also its intensity.

The PL of samples W3 and W5 should also differ only in the interference effects, since the active layer is the same in both samples. As with the previous pair of samples, in figure 3.42 we observe that the experimental PL is actually quite different in both samples. Note that sample W5 is very similar to the sample studied in section 3.5.2. Not surprisingly, the PL shows the same bands that were previously attributed to emission in the SRN and SRO layers. If we proceed like in previous analysis, we also achieve a good match between the measured PL of sample W3 and the calculated PL considering that its intrinsic emission is the same as that of sample W5, as confirmed by the blue dotted line in figure 3.42. Again, note that we are not normalizing the spectra, and we get a good match for the intensity. The thicknesses used for the calculation are: 31 nm for the SRN; 34 nm for the SRO; 442 nm for the polysilicon; and 2032 nm for the SiO$_2$. Again, those figures are in good agreement with the measured thicknesses.

Finally, it could be useful to see if the differences in the PL emission between samples W1 and W2 are related to interference effects. In this case the thickness of the SRN layer and the implantation parameters are the same, but the thickness of the SRO layer is different. Therefore, we expect the component of the emission coming from the SRN layer to be the same in both samples, whereas the emission in the SRO should be different because the same amount of implanted silicon will diffuse across a larger volume of SiO$_2$, and this will result in a different silicon concentration, and hence a different density and size distribution of the silicon nanoclusters. This
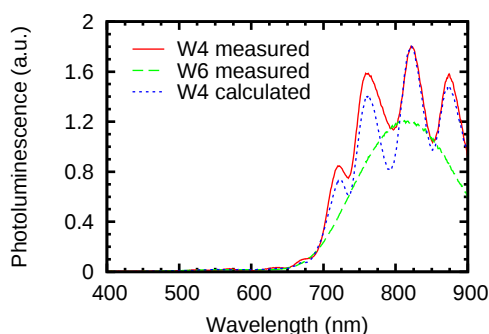
Figure 3.41: Measured PL of samples W4 and W6. The signal is quite different although the active layer is the same.
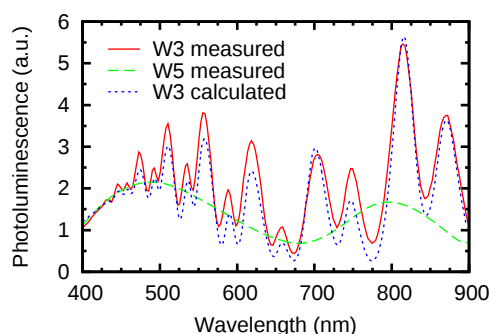
Figure 3.42: Measured PL of samples W3 and W5. The PL of sample W5 has been decomposed in two bands like in section 3.5.2.
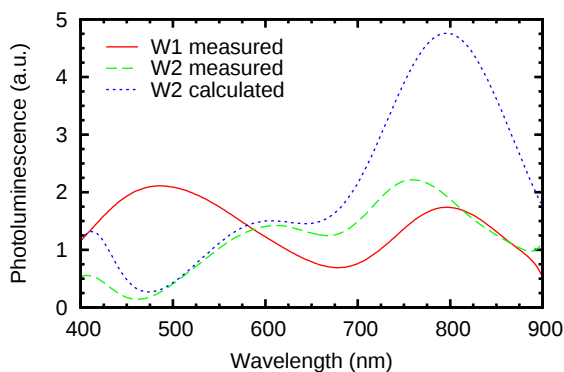
Figure 3.43: Measured PL of samples W1 and W3. Note how we can predict the features of the SRN band (blue region) but not the features of the SRO band (red region).

in turn should affect both the intensity and peak of the SRO component of the emission.

In order to confirm this we proceed as we just did with samples W3 and and W5: the emission of sample W1 is decomposed in two bands and we find their intrinsic spectrum. We then assume the emission to be the same in the SRN and SRO layers of sample W2. Note that here, since the thickness of the SRO layer is different, we will suppose that the same total energy is emitted in samples W1 and W2, but in sample W2 it is distributed across a larger volume, hence with a lower density. With this assumption, the observed emission of sample W2 is calculated. The thicknesses used in the calculations are: 31 nm for the SRO in sample W1; 304 nm for the SRO in sample W2; and 31 nm for the SRN in both samples. Those values are in good agreement with the measured thicknesses. The results are shown in figure 3.43.

The results confirm what we expected. We can essentially reproduce the observed

emission of sample W2 corresponding to the SRN band (400–650 nm), but the calculated emission in the region dominated by the SRO band (650–900 nm) does not fit, either in peak position or intensity. That confirms that the component of the emission originated in the SRN band is intrinsically the same and the apparent differences are only due to interference effects, while the emission originated in the SRO layer is intrinsically different.

The last result highlights the importance of considering the effect of the multilayered system on the emission. If one had to analyze the PL of sample W2, it would be straightforward to identify three different bands, one at 400 nm, a second one at 600 nm and another at 800 nm, and relate these to different radiative centers. However, the truth is that only two such bands exist, one around 500 nm and another around 750 nm. The bands at 400 nm and 600 nm observed in the spectrum of sample W2 are artifacts introduced by interference effects.

The small mismatch between the measured and calculated spectra in all the results shown in the present section may have different origins.

First and foremost, a recurrent problem when using this model to find the intrinsic emission from the measured one is that we do not know with absolute certainty how to decompose the observed emission into bands corresponding to different layers. A very good example of that is sample W2. If we want to obtain the intrinsic emission of sample W2 from the observed PL spectrum and we have reason to believe that two different components, originated in different layers, are found in the spectrum, we need to decompose the observed emission into said components, and in this case it would be impossible to do so because the two components are mixed in such a way that it is not possible to tell them apart. In sample W1 and other similar samples, we can come up with a reasonably good approximation to the real decomposition because there are two clear bands and they match the bands that can be found in samples featuring a single SRN layer and samples featuring a single SRO layer. However, even in these more favorable cases, the decomposition will not be exact, and therefore a certain degree of error can be expected in all the calculations.

Although this is not so important in systems with a single active layer, it is still true that the assumption of a constant spectrum and constant distribution of the energy across the layer might not be completely accurate. In the samples with 30 nm thick SRO, it may be quite accurate because the implantation is engineered to yield a reasonably flat profile across the thickness of a 30 nm thick layer, but this will no longer be the case when we implant in the samples with a 300 nm thick SRO layer while keeping the same implantation parameters. In those cases, it stands to reason that most of the energy could be concentrated in a sublayer of the SRO rather that equally distributed across the whole thickness.

Finally, we should not rule out the possibility of small differences in the intrinsic emission of samples with equal fabrication parameters.

3.6 Final notes and conclusion

In this chapter, we have presented a method originally developed by Crawford to calculate the modification of the emission of a radiative center embedded in a multilayer system. We have extended the method to account for a continuous distribution of radiative centers and have implemented it in a Python script, which has been used to analyze the effect of our typical multilayer systems on the emission of the active layer in a general form. It has also been shown that the calculation of the energy ratio F allows us to show how the differences between the measured PL and EL of our samples can be explained by the interference effects that take place in the stack.

As a conclusion to this chapter, we will show that, unfortunately, the method presented here is not a magic bullet that will allow us to calculate the intrinsic emission in every circumstance.

It has already been shown that, in order to accurately determine F, we need a precise description of the system and the distribution of the radiative centers. Lack of precision in the measurement of the thicknesses can be overcome by letting the thicknesses be a parameter that can be varied within the given boundaries, but the distribution of the radiative centers will most likely involve an educated guess. In section 3.5 we could validate our guess because we had reason to believe the intrinsic emission to be the same in PL and EL experiments, and we could arrive to a reasonably good match after factoring in the interference effects (figure 3.39(b)). However, if we only had one measurement and wanted to find the intrinsic emission, we would not have any means to validate our findings. In that case, an analysis of the change in the calculated intrinsic emission for different distributions could give us a measure of the degree of uncertainty in the result.

We have also seen that, in case of having different emission spectra at different positions in the stack, such as two different bands in two different layers, it is necessary to know in advance which part of the measured spectrum belongs to which band. If we lack that knowledge, we cannot remove the effect of the multilayer unless we separate the components of the spectrum based on a model, for instance fitting two Gaussian functions, one for each band. This will not be possible in all occasions, most importantly in the EL measurements, where the interference effects are strong. Even in PL measurements, where the bands can be approximated by Gaussian functions fairly well, when we fit the data with the model we lose any information that the model fails to capture. For instance, perhaps we had three bands instead of two, one of them being just a small shoulder on one of the main bands. When we fit the data with two bands instead of three and remove the interference effects from the modeled data, our result will of course lack that third component that we dismissed when we failed to add a third band to our fit. That might not be a problem if all we want to show is that the differences between the PL and EL measurements can be explained by interference effects, as we did in 3.5. However, if we want an accurate calculation of the intrinsic emission in order to draw conclusions as to its originating mechanisms, fitting the data with a model *before* removing the interference effects is the wrong thing to do. In that case, some
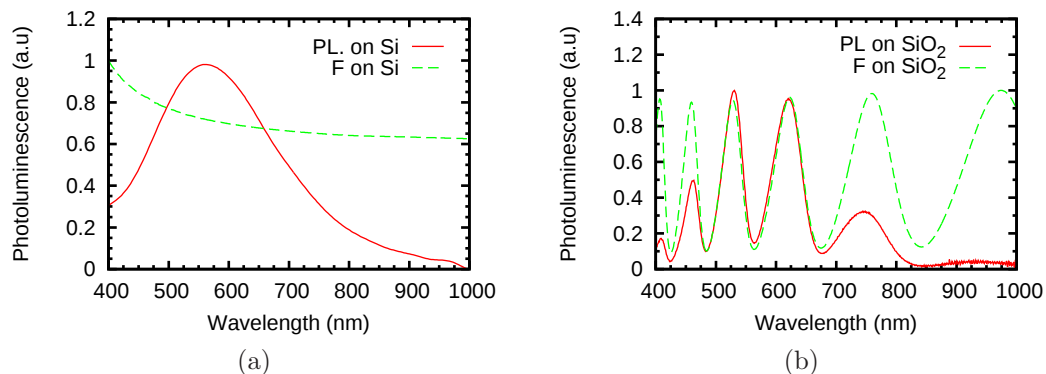
Figure 3.44: PL and F function for an SRN layer on top of (a) the Si substrate and (b) the 1.2 μm thick SiO_2 buffer layer.

way to experimentally measure each contribution separately is required, so that the effects of the multilayer can be removed from the raw data directly.

Finally, in section 3.5.2 it was shown that the differences between the EL and PL measurements could be explained by the interference effects in the multilayer. To do that, we calculated the intrinsic bands from the PL measurements (approximately, as per the remarks in the previous paragraphs), and then we applied the interference effects of the system used in EL measurements, only to find that the calculated emission matched the observed one. One could argue that it would have been less convoluted if we had removed the interference effects from both the PL and EL measurements and then compare the calculated intrinsic emissions to see if they were the same. There are two reasons why we avoided this approach.

The first one is that we have no way of separating the EL into its two components, as we have already noted. The second will be illustrated with the following example.

A 30 nm SRN layer was fabricated with two different substrates: directly on top of silicon and with a 1.2 μm SiO_2 buffer layer between the SRN and the silicon. The PL was measured in both configurations. Figure 3.44(a) plots the PL of the SRN on top of the silicon substrate and the F function for that system, considering a uniform emission in the SRN. Both functions have been normalized.

Figure 3.44(b) plots the PL of the SRN on top of the SiO_2 buffer and its corresponding F function, both normalized and again considering uniform emission.

Now, if we divide the experimental PL of the SRN on silicon by its corresponding F, we will obtain the intrinsic emission. If we then multiply that by the F function in figure 3.44(b), corresponding to the system with the buffer layer, we should obtain something very similar to the experimental PL in figure 3.44(b). Indeed, figure 3.45 compares the PL of the SRN layer on top of SiO_2, as measured and as calculated using this method. Even though there are small inaccuracies, it is not hard to sell that the difference between the PL measured on silicon and on SiO_2 is merely due to interference effects.
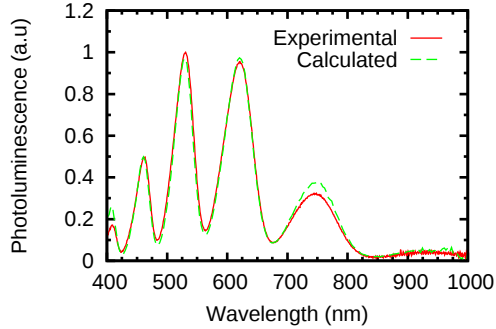
Figure 3.45: Comparison of PL of the SRN layer on top of SiO_2, as measured and as calculated from the intrinsic emission obtained from the PL on silicon.
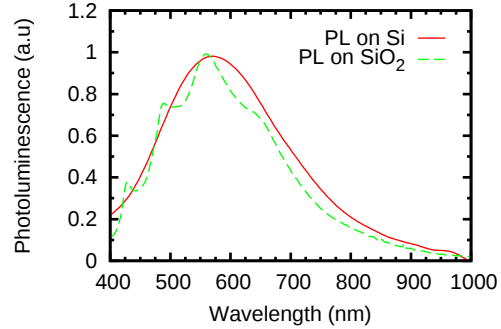
Figure 3.46: Comparison of the intrinsic emission as calculated from the PL measured on silicon and on SiO_2.

If we calculate the intrinsic emission by dividing the PL of the SRN on silicon by the corresponding F (the curves in figure 3.44(a)), and then do the same using the data from the system with the SiO_2 (divide de curves in figure 3.44(b)), we would expect both calculated intrinsic emissions to be the same. However, what we get in each case is what appears in figure 3.46. Even though the curves are similar, it would be tough to sell that these are actually the same.

The reason for the differences is not hard to understand. For the SRN on silicon, we find the intrinsic emission by dividing a smooth function by an even smoother and relatively flat F function. The result is obviously smooth. On the other hand, for the SRN on SiO_2, we find the intrinsic emission by dividing a peaky function by another peaky function in such a way that, in theory, every peak in the latter should exactly cancel a peak in the former in a smooth way. However, any small inaccuracy in the relative heights of the peaks or misalignment in their positions will have a dire effect in the calculated intrinsic emission. This is an accuracy problem, and that is why it is better to follow the more convoluted approach that we have been following.

That problem is akin to what is encountered in the computation of the difference of two nearly equal floating point numbers (*catastrophic cancellation*). To avoid the large inaccuracies involved in those subtractions, workarounds that substitute the subtraction for an addition are used whenever possible [67].

Of course, if finding the intrinsic emission is our goal, then it will be very difficult to find it accurately if our measurement is heavily distorted by interferences, like in the example of figure 3.44(b). Nevertheless, the dashed line in figure 3.46 is still a much better approximation to the real emission than the raw PL measurement of figure 3.44(b).

Transceiver

4.1 Introduction

The goal of this project is to integrate a light source, a waveguide and a light detector in a single monolithic process that can be carried out in a CMOS line, with the end goal of serving as a sensing platform. The design and fabrication procedure will be explained, accompanied of simulations of the most critical steps in the process. The results of the characterization of the finalized devices are also presented and discussed.

4.2 Design and fabrication

In this section the design of the proposed transceiver, as well as the processes involved in its fabrication, will be explained.

4.2.1 Working principle and vertical layout

Figure 4.1 shows the vertical layout of the proposed device. The emission takes place in a nitride layer which is also used as a waveguide. One of the ends of the nitride guide is implanted with silicon ions and annealed at high temperature in order to obtain the active layer. A p well below the active layer is used as a back contact. A bias V_E is applied between the contact on top of the active layer and the p well and the light generated in the SRN section of the guide travels sandwiched between two layers of silicon dioxide, which provide the necessary confinement. The light that reaches beyond the silicon dioxide trench can generate carriers in the PIN diode formed by the second p well and the n doped back of the wafer. A reverse bias V_D is used to ensure the full depletion of the $n-$ region. The light generated carriers will be swept away by the electric field generated between the p and n wells and will be detected. In a future use of the device as a sensor, the top SiO_2 slab of the waveguide would be used as the sensing platform. However, the successful integration of emitter, waveguide and detector has to be sorted out first.

121
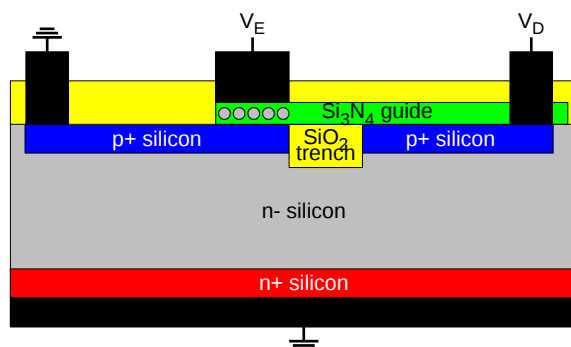
Figure 4.1: Vertical layout of the proposed device. Note that the vertical dimensions are scaled conveniently to make all features easily visible.

There are three main challenges to overcome in this configuration. The first one is the thickness of the nitride guide. On the one hand, one would like a thick layer in order to avoid the system to be dominated by edge phenomena. On the other hand, the thicknesses that can be used are limited by the electrical performance of the active layer. If the nitride is too thick then very high voltages are needed to operate the emitter. That forces the acceptable thicknesses of the nitride guide to be in the few tens of nanometers.

The second challenge is the planarity of the silicon dioxide trench. To avoid coupling of light from the nitride guide into the silicon substrate, the silicon trench must exceed 1 micron in thickness. On top of that trench a nitride layer of no more than a few tens of nanometers will be deposited. It is imperative that any surface feature in the region is smooth enough so as to not jeopardize the correct operation of the guide.

Finally, the thermal load during the fabrication will be dominated by the most aggressive annealing, which is the one required to form the active layer. That means that our ability to optimize the p and n wells is limited to the implantation parameters because the diffusion process will be utterly dominated by the annealing required for the formation of the active layer.

4.2.2   Mask design

The mask set used for the fabrication includes six levels in total:

1. Oxide trench

2. p-wells

3. Nitride guide

4. Implantation of the active layer

5. Windows in the field oxide

6. Aluminum contact pads

The masks are designed to be used in photolitographic steps with a positive photoresist (the illuminated photoresist is dissolved). Therefore, the levels that require opening windows in a material were fabricated with dark field, whereas the levels that require extrusion were fabricated with bright field. In the following sections we will briefly describe the features of each layer.

4.2.2.1 Layer 1: oxide trench.

This layer in the mask set is used to define the regions of the silicon substrate that will be etched away and later filled with oxide, in order to create the bottom slab of the guide.

Figure 4.2(a) depicts the layer corresponding to the oxide trench (only the area corresponding to a single chip is displayed). The width of the guide is the same in all six devices in any given chip, and slightly wider than the active layer. However, three different lengths of the trench are found in each chip, with the goal of studying how the path length in the guide affects the coupling between emitter and detector.

Since this layer is fabricated in a dark field, the rectangular features shown in the picture will be photoresist-free after the photolitography. Subsequent etching will open windows in the oxide with the rectangular shape depicted in 4.2(a). The windowed oxide will serve as a mask for the following etch of the underlying silicon substrate, and we will get trenches in the silicon with the rectangular shape shown in 4.2(a)

4.2.2.2 Layer 2: p-wells.

This level in the mask set is used to define the regions of the silicon substrate that will be implanted to generate the p wells. For each device in the chip there are two wells: one (bottom of figure 4.2(b)) is used as a back contact for the emitter and the other as the p region of the diode that forms the detector. There are three extra wells at the top of the layout, one intended for measurements of the resistivity of the wells, and two meant to be used as standalone detectors with two different area sizes, for testing purposes.

The implantation wells extend at both ends of the oxide trench. Therefore, the two wells in every pair, one for the emitter and the other for the detector of the same device, are separated by a length equal to the length of the trench (see figure 4.1).

In order to implant silicon in the substrate in the rectangular regions depicted in figure 4.2(b) we need to deposit an oxide layer that will serve as implantation mask, and then open windows in the oxide in the rectangular regions where we want to implant. Therefore, we also fabricate this level in a dark field.
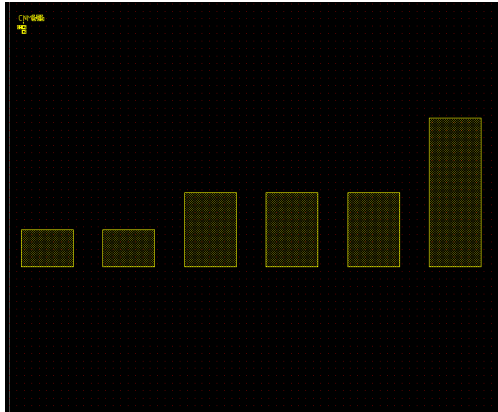
(a) Oxide trench.
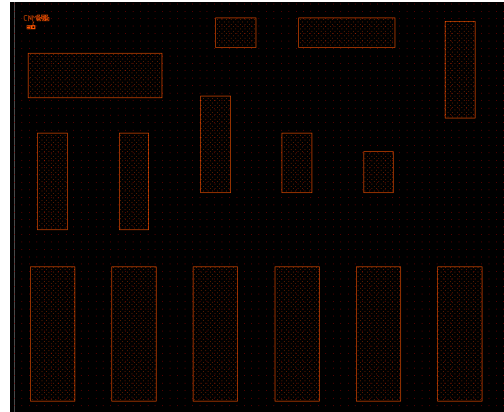
(b) p wells.
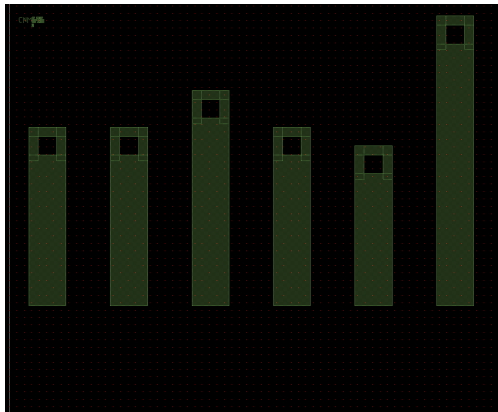
(c) Nitride guide.
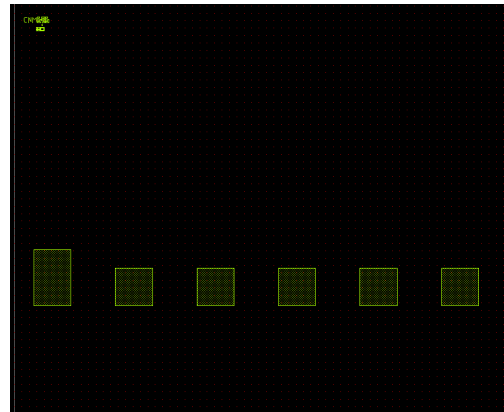
(d) Implantation of the active layer.
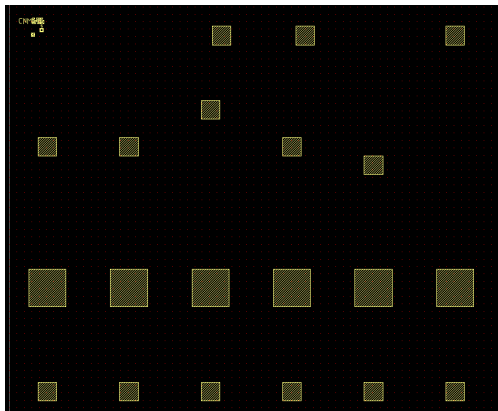
(e) Windows in the field oxide.

(f) Aluminum contacts.

Figure 4.2: Design of each of the levels of the mask for one complete chip.

4.2.2.3 Layer 3: nitride guide.

The third layer in the mask set corresponds to the nitride guide, and is depicted in figure 4.2(c). Since one end of the guide will act as an active layer, it must rest on top of one of the p wells, extend over the oxide trench, and end beyond the trench, on top of the detector well. Several lengths of the guide are fabricated. In combination with the length of the trench, that results in several combinations of guide length and detector area in each chip.

The hole at one end of the guide will be filled with aluminum, which will be in contact with the p well in the substrate and will allow us to reverse bias the detector.

For the waveguide we need a nitride layer with the shape depicted in figure 4.2(c). That means that we have to deposit a nitride layer and then remove it everywhere except in the rectangular regions shown in figure 4.2(c). Therefore, after the photolitographic step we need the photoresist protecting the region where the guide is going to be, and that is achieved by fabricating this layer in a bright field.

4.2.2.4 Layer 4: Implantation of the active layer.

This level, depicted in figure 4.2(d), corresponds to the implanted region of the nitride layer that will be used as an active area. The implanted region is a square at the emitter end of the guide. Therefore, the width of the guide is what defines the size of the square emitter. The innermost edge of the implanted area is aligned with the edge of the oxide trench. One exception is the leftmost device of the chip, in which the active area is pushed further inwards on top of the oxide trench. We want to see if that improves the coupling to the guide.

Since we need to implant a square region of the nitride layer on one side of the guide while protecting the rest of the wafer from the implantation, this level needs to be fabricated in a dark field so as to open a window in the masking layer in the square regions defined by the features of the mask.

4.2.2.5 Layer 5: Windows in the field oxide.

The whole structure is covered in a field oxide to protect the structure and isolate different chips, as well as to serve as the top slab of the guide. In order to place the aluminum contact pads we need to open windows in the field oxide that allow us to reach the surface that we want to contact. That means three windows for each device of the chip: one for the back contact of the emitter (small squares at the bottom of figure 4.2(e)), one for the top contact of the emitter, that is on top of the active area (big squares), and the third for the detector. The two extra windows at the top of figure 4.2(e) belong to the two test detectors.

For this level we need to define the windows in a dark field.
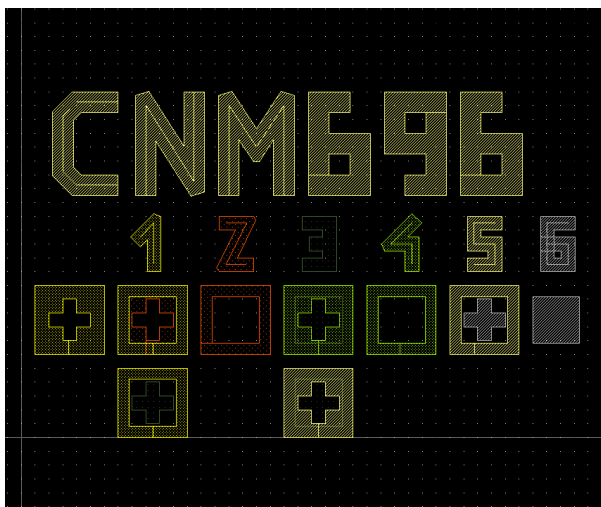
Figure 4.3: Alignment features used in the set of masks.

4.2.2.6 Layer 6: Aluminum contact pads.

The final level of the mask set corresponds to the aluminum pads, which are located in the same areas where the windows were opened on the oxide, as shown in figure 4.2(f). Additionally, the aluminum layer is also used to visually separate individual devices as well as to display labels that allow the identification of every device.

Since the fabrication of the pads involves depositing an aluminum layer and removing it everywhere except where the contact pads are going to be, we need to define the features of the mask in a bright field.

4.2.2.7 Alignment features.

Figure 4.3 shows the features used for the alignment of the masks in the set. To properly align a mask with the underlying wafer we add at least one square and one cross to each level. The cross in level N is placed within the square of level $N-1$. In figure 4.3 we can see how the orange cross corresponding to level 2 is placed within the square corresponding to level one. Since level 3 will also be aligned against level 1 (because the implantation corresponding to level 2 leaves no mark that one can use to visually recognize its corresponding square) we need to add a second square to level one, and the cross corresponding to level 3 is placed inside it. Levels 4 and 5 both are aligned against level 3 (again, the implantation corresponding to level 4 leaves no visual mark so the 5th level cannot be aligned against it). Finally level 6 is aligned against level 5. Additionally, we draw a cross for level 1 and a square for level 6 in case we need to add a level before 1 or after 6 in the future.

Note that in the levels that are going to be fabricated with dark field we draw the inverse of the alignment feature (see for instance the features corresponding to level 1 in figure 4.3), whereas in the levels that are going to be fabricated with
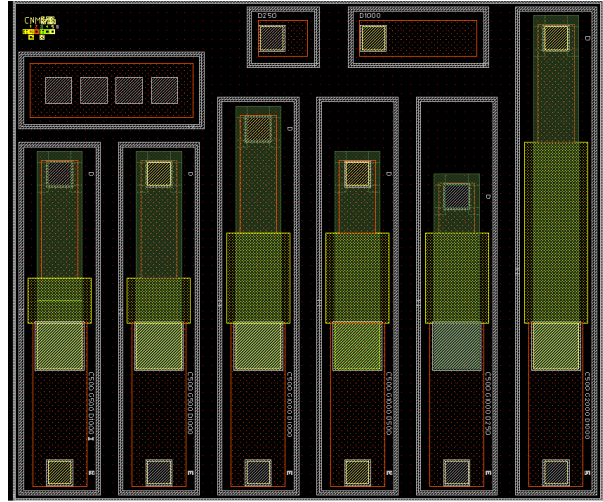
Figure 4.4: Complete view of the chip. The same chip is fabricated in three sizes, capacities 500×500, 1000×1000 and 2000×2000 μm. Note that all three sizes look exactly the same because they only differ in a scale factor.

bright field we draw the feature directly (for example level 6 in 4.3). That is for practical purposes. Since the technician carrying out the photolitography needs to align the cross of the current level with the square of some underlying level, it is necessary for him to actually *see* around the cross in order to tell if the cross is within the underlying square. That is why we always want the cross in the mask to be a dark feature surrounded by a transparent area. If we are working in a bright field then what we draw is dark and therefore we can draw the cross directly (as in level 6). However, if we are working in dark field then we what we draw is going to be transparent in the actual mask, and that is why we need to draw the area around the cross while leaving the cross itself as dark background.

In figure 4.4 the combination of all layers is presented, while figure 4.5 shows the distribution of the chips in the wafer.

4.2.3 Fabrication process

The fabrication consisted in seven wafers. One of the wafers was intended for the FIB analysis of the trench. The other six wafers can be separated into two groups, one with a 150 nm thick guide, and the other with 30 nm thick guide. The active layer is 30 nm thick in all cases. In each group, one of the three wafers is intended for photoluminescence studies of the active layer (in the group with 150 nm thick guide, the active layer is obtained from a 150 nm Si_3N_4 layer which is later etched down to 30). The other two wafers are have complete structures with polysilicon and aluminum contacts, respectively. See table 4.1 for a list of the wafers.

Unfortunately, during the fabrication process, the step that was supposed to etch
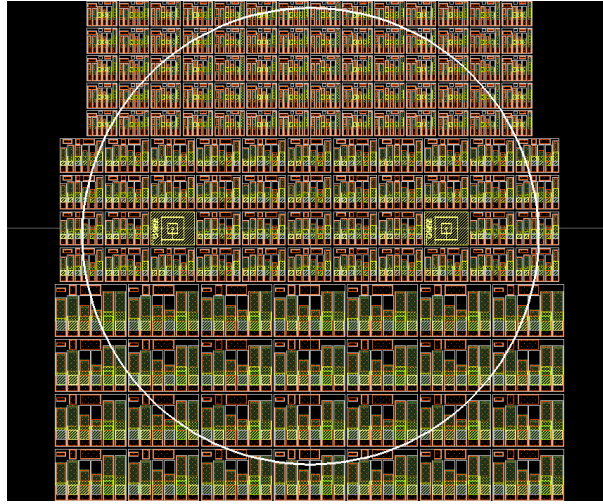
Figure 4.5: Distribution of the chips on the wafer.

Wafer num.	Pattern	Guide thick. (nm)	Contacts
1	Only active layer		
2	Complete devices	30	Poly
3	Complete devices	30	Al
4	Only active layer		
5	Complete devices	150	Poly
6	Complete devices	150	Al.
7	Only trench		

Table 4.1: List of the fabricated wafers.

the 150 nm thick Si_3N_4 layer down to 30 nm resulted in an over-etch that removed essentially all the active layer, and wafers 4, 5 and 6 were ruined in the process. The fabrication of the remaining wafers was successful and the thicknesses of the active layer were measured to be very close to the targeted 30 nm and with good uniformity.

The fabrication of the transceiver can be divided in five main stages:

1. Oxidation of the trench.

2. Implantation of the p and n wells.

3. Definition of the nitride guide.

4. Obtainment of the active layer.

5. Definition of the contacts.

The following subsections describe in detail the precise steps and processes involved in each stage. See figure 4.6 for an illustration of the basic stages of the fabrication process. Appendix C contains an exhaustive list of the steps required for the fabrication of the sample set.

4.2.3.1 Initial steps

A set of n-type wafers was selected with resistivity in the range 1–12 Ω cm. Their precise resistivity was measured and a subset of seven wafers with the most similar resistivities was selected (figure 4.6(a). After cleaning of the wafers, the steps to achieve the oxide trench were initiated.

4.2.3.2 Oxidation of the trench

A 36 nm thick SiO_2 layer was grown by dry thermal oxidation. Next, a 117 nm thick Si_3N_4 layer was deposited by LPCVD. The nitride layer will be used as a mask for the LOCOS (local oxidation of silicon) process, and the underlaying 36 nm SiO_2 layer is used as a buffer layer to avoid the strain generated by the nitride layer during the thermal processes from creating dislocations in the silicon substrate [15].

A photolitography with mask layer 1, see section 4.2.2.1, followed by a dry etching of the 117 nm nitride layer and a wet etching of the 36 nm oxide layer will open a window to the underlying silicon substrate. A 750 nm deep dry etching of silicon will create a rectangular trench in the silicon substrate (figure 4.6(b)).

A 1.6 μm wet oxidation, which is masked by the nitride layer, will overfill the trench with oxide. After that, the oxynitride layer generated on top of the nitride is eliminated by wet oxidation and then the nitride layer itself is stripped off by wet etching.

In order to planarize the remaining oxide layer, a PECVD layer of 130 nm of undoped TEOS (tetraethyl orthosilicate) and 1200 nm of BPTEOS (borophosphorous tetraethyl orthosilicate) is deposited on the wafer, and a fluidification of the BPTEOS at 1000 C planarizes the surface of the wafer. The dry etching of the whole layer of oxide ideally leaves a clean silicon surface and an oxide trench leveled with the silicon surface.

Finally, the back side of the wafer is stripped clean of all the layers that have been deposited there during the previous processes. One of the wafers is used to examine the wafer by FIB to assess how leveled the trench is with the surface of the wafer (figure 4.6(c)).

4.2.3.3 Implantation of the p and n wells

This section of the process involves the creation of the implantation wells in the wafer. Two of the wafers, which will be destined to photoluminescence characterization, will be implanted in the whole surface (no implantation mask will be used). In the rest of the wafers, the implantation will be masked so that only certain rectangular regions of the silicon surface will receive a dose.
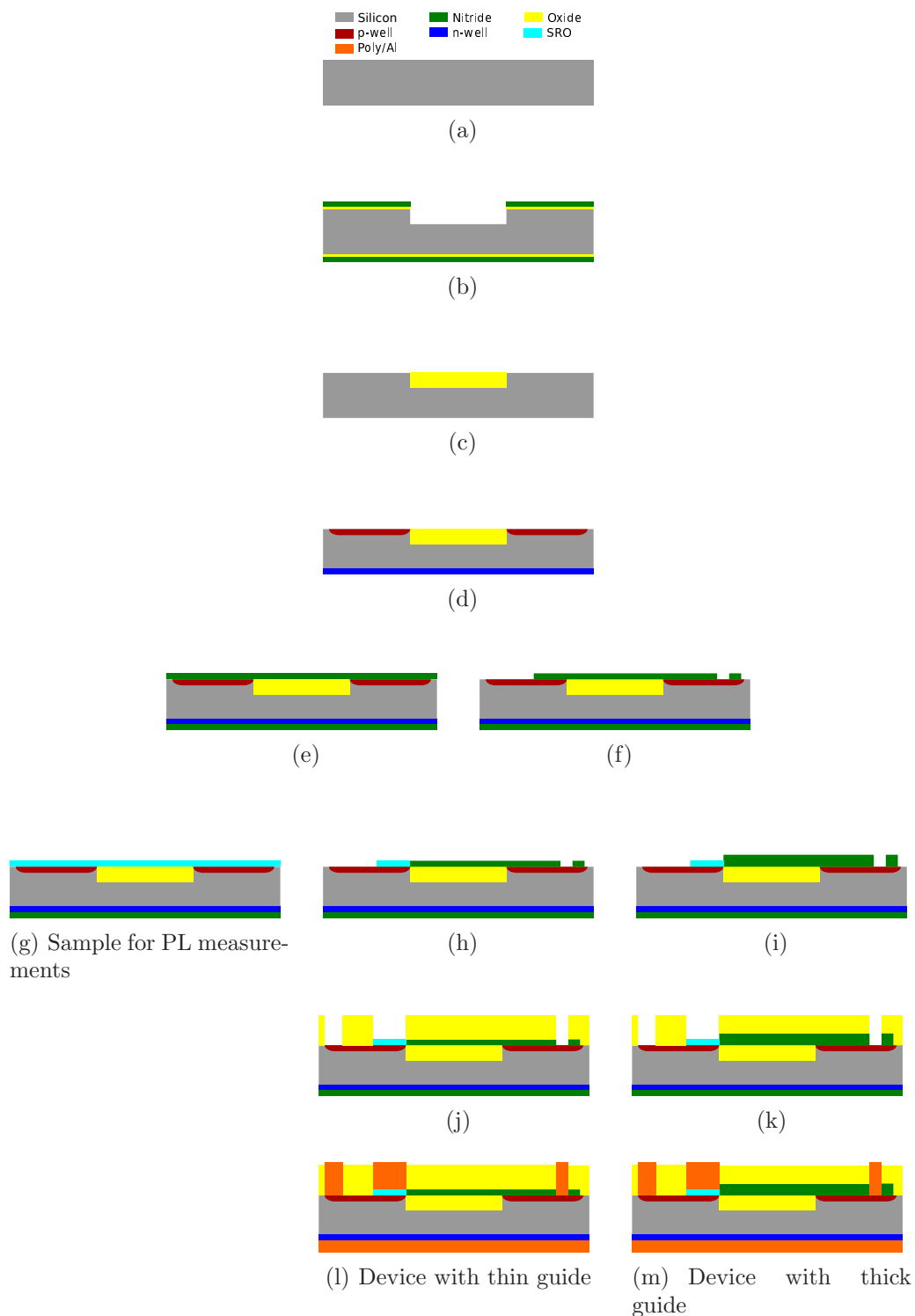
Figure 4.6: From top to bottom, different stages of the fabrication process. Different columns correspond to different groups of samples.

In order to create the implantation mask, a 300 nm dry thermal SiO_2 layer and a 117 nm LPCVD Si_3N_4 layer are deposited on the wafers. For the samples that will actually use the mask for implantation, a photolitography with mask layer 2, see section 4.2.2.2, opens the implantation window in the photoresist (the samples destined to PL skip the photolitography step). Then the two deposited layers, nitride and oxide, are etched away by dry and wet etching, respectively. In the samples with the photoresist, that will open the implantation window. In the samples destined to PL the layers are removed from the whole wafer.

The implantation of the top p wells is then carried out, followed by the etching of the nitride and oxide mask from the wafers that had it. Finally, the $n++$ implantation of the back side of the wafer is carried out (figure 4.6(d)). The implantation of the p regions is carried out at 50 KeV for a dose of 1.0×10^{15} cm^{-2}, whereas the $n++$ implantation is carried out at 100 KeV for a dose of 4.2×10^{15} cm^{-2}.

4.2.3.4 Definition of the nitride guide

The wafers are split in two sets, one with a 30 nm thick nitride guide and another with a 150 nm thick nitride guide. The first step in this section is to deposit the Si_3N_4 layer that will act as a guide. The following steps do not apply to the samples that are going to be used for photoluminescence characterization (figure 4.6(e)).

The geometry of the guide is defined by photolitography with level 3 of the mask set, see section 4.2.2.3, followed by a thermal annealing at 200 C for 30 minutes in order to harden the photoresist, and finished by dry etching of the nitride layer (30 nm or 150 nm, depending on the sample) (figure 4.6(f)).

At this point, the samples destined to photoluminescence differ from the others in that they are covered by a uniform nitride layer, whereas the others have the guide patterned in the the nitride layer.

4.2.3.5 Obtainment of the active layer

The active layer is fabricated after implanting silicon ions in a square region at one edge of the active layer. Again, the first steps of this section do not apply to the samples destined to photoluminescence characterization.

A photolitography with level 4 of the mask set, see section 4.2.2.4, leaves an implantation window open in a square region at one edge of the nitride guide. After hardening of the photoresist through a thermal annealing at 200 C for 30 minutes, the photoresist itself can act as an implantation mask.

At this point, the set of samples that had a 150 nm nitride layer, both the samples with the patterned guide and one of the samples for photoluminescence characterization, suffer a dry etching of 120 nm out of the total 150. In the case of the sample for PL, it leaves the nitride layer at 30 nm, just like the other sample for PL, whereas for the other samples the etching is masked by the photoresist so that only the square area of the guide that is going to be implanted sees its thickness lowered.

Finally, a 1×10^{16} cm^{-2} silicon ion implantation at 20 KeV is carried out in all the samples (figures 4.6(g), 4.6(h) and 4.6(i)) followed by a thermal annealing at 1100 C during 60 minutes.

The process ends here for the samples destined to PL measurements (figure 4.6(g)). The remaining steps apply only to the samples with full devices.

4.2.3.6 Definition of the contacts

Here the process begins with a PECVD deposition of 130 nm of undoped TEOS and 1370 nm of BPTEOS, which will act both as top cladding for the nitride layer and as passivation for the rest of the device. The deposition is followed by a photolitography with layer 5 of the mask set, see section 4.2.2.5 and a thermal annealing at 200 C for 30 minutes to harden the photoresist. The thick TEOS and BPTEOS layer is etched by two consecutive attacks, first dry etching and then wet etching, to open windows in the passivation layer. Finally, the photoresist is stripped off (figures 4.6(j) and 4.6(k)).

Next, the contact material is deposited. Here the samples will again be split into two groups, one with polysilicon contacts and another with aluminum contacts. The polysilicon contacts are intended as a control sample, since the performance of the active layer can be evaluated through the semi-transparent polysilicon contact.

In the case of the polysilicon contacts, the process continues as follows. A 350 nm thick polysilicon layer is deposited. The polysilicon is then doped with POCl$_3$. The phosphosilicate glass (PSG) created during doping is then removed.

For the aluminum samples, a 1 μm thick aluminum layer is deposited by sputtering.

For both groups of samples the next step is to define the contact pads through a photolitography with layer 6 of the mask set, see section 4.2.2.6. Then the polysilicon or aluminum are etched to define the contacts, wet etching for the aluminum and dry etching for the silicon.

At this point, all the layers deposited on the back of the wafers during the previous processes are stripped off.

Finally, and for all samples, a 1 μm thick aluminum layer is sputtered on the back of the wafer as a back contact and sintered at 350 C in N$_2$/H$_2$ atmosphere (figures 4.6(l) and 4.6(m)).

4.3 Simulations

In order to assess the feasibility of the proposed device, some simulations were carried out. In particular, the LOCOS process and the fabrication of the p and n wells were simulated with the Sentaurus suite.

4.3.1 Obtainment of the oxide trench

One of the most challenging steps in the fabrication process is the creation of the oxide trench. Achieving a flat trench with a smooth transition to the silicon substrate is important in order to keep the nitride guide that will be deposited on top as smooth as possible, which is required so as to minimize the propagation losses in the guide. Our first approach was to try a local oxidation of silicon (LOCOS) process. In this process, an oxidation is carried out in a localized region of the silicon. The localization is achieved by masking the oxide growth with a nitride layer (see figure 4.6(b)). However, since we want a relatively deep oxide trench, the exposed silicon is first etched so that the cavity will be filled with oxide after the oxidation.

Simulations of the LOCOS process with Sentaurus readily showed that planarity is a complicated feature to achieve. Since the oxide grows everywhere the silicon is exposed, including the walls of the trench, the oxide growing on the corner of the trench with the surface of the wafer is bumped up, leaving an inconvenient protuberance that cannot be removed by subsequent etching because the etching process is largely conformal. Figure 4.7(a) shows the effect in the case of a trench with vertical walls, achieved by dry etching.

It seems reasonable to expect this effect to be somewhat smoothened if the walls of the trench come at an angle instead of being vertical. That can be achieved with an anisotropic wet etching with TMAH (tetramethylammonium hydroxide). Sentaurus simulations show that this is indeed the case, as can be seen in figure 4.7(b). In that case the oxidation was simulated on a trench with walls a 54 degrees. While that result is an improvement over the vertical walls, it is still far from smooth and planar.

The natural extension is to work without digging a trench on the silicon, that is as in an ordinary LOCOS process. In that case a longer oxidation is needed in order to get the desired depth of the trench, and the result is shown in figure 4.7(c). While this gives a smooth profile, the result is still far from planar, and a subsequent etch makes things worse due to the conformality of the process. Figure 4.7(d) shows the simulation of an etching process on the result shown in figure 4.7(c). While the surface of the oxide and that of the silicon are leveled far from the edge of the trench, we get an unacceptable groove right at the edge, due to the aforementioned conformality of the etching process.

In order to test the validity of the simulations, a test sample was fabricated in order to examine the resulting trench with a focused ion beam.

Focused ion beam (FIB) is a lithography technique that uses metal ions (typically gallium) extracted from a liquid metal source by an electric field to bombard the sample to be patterned. The obtained ion beams hold high brightness levels and low energy spread and have diameters under 50 nm. Moreover, the patterning is maskless (software masks are used instead).

It is a very versatile technique that can be used for micromachining and ion milling, microdeposition of metals or maskless ion implantation, among other applications.

(a) Pre-etched trench with vertical walls.

(b) Pre-etched trench with walls at 54 degrees.

(c) Standard LOCOS without pre-etched trench.

(d) Standard LOCOS without pre-etched trench followed by a post etching of the oxide.
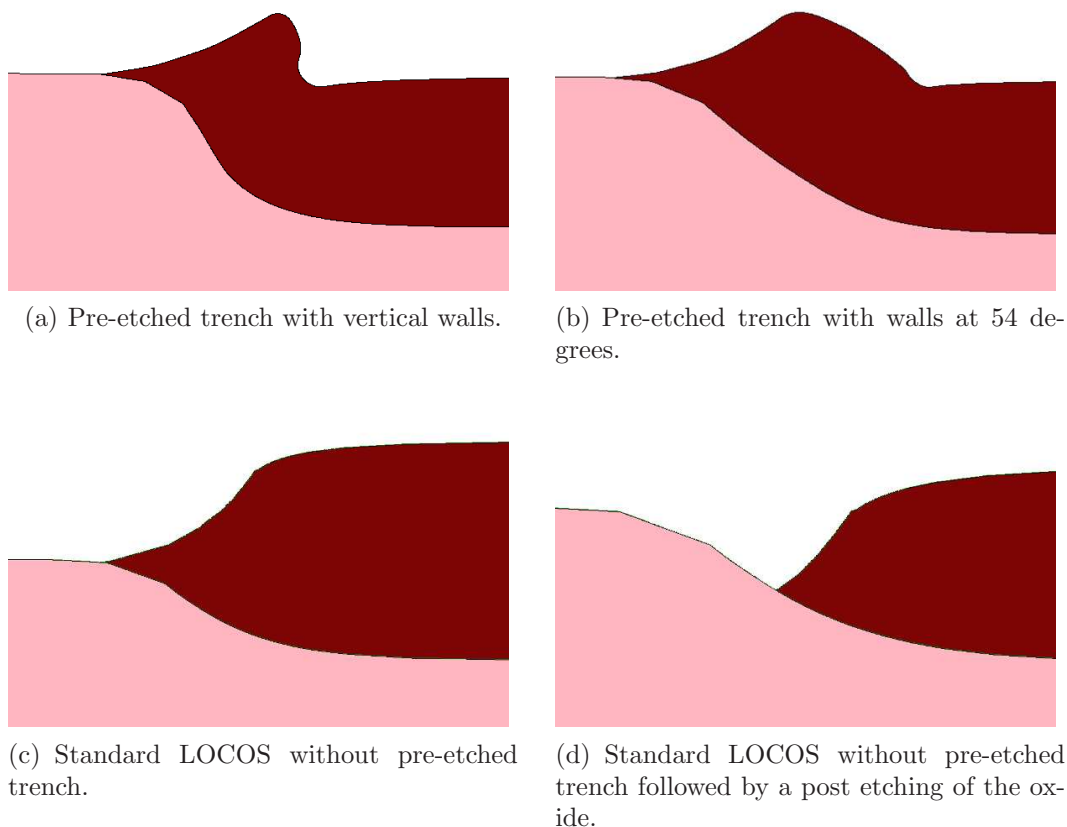
Figure 4.7: Simulated SiO_2 trench in different conditions.

A FIB column is usually combined with a SEM (scanning electron microscope) column to visually examine a vertical cut of the sample. The SEM column is oriented at a certain angle with the plane of the sample and the FIB column (figure 4.8). The SEM provides a visual feed of the milled region while the FIB mills a trench, all in real time.

A detailed treatment of FIB can be found in [68]. SEM and related techniques are treated, for example, in [69]. For this work, a Zeiss 1560XB FIB+SEM system was used.

Figure 4.9(a) shows the result after the LOCOS process for a trench 750 nm deep. It is clear that the simulations correctly predicted the bump at the edge of the trench.

In order to solve the problem, a layer of undoped TEOS and BPTEOS was deposited. Its thickness, 1.3 μm, is large enough to completely cover the bump. A fluidification process planarizes the layer, as can be seen in figure 4.9(b). Close inspection of that picture reveals the underlying bump buried under the BPTEOS.
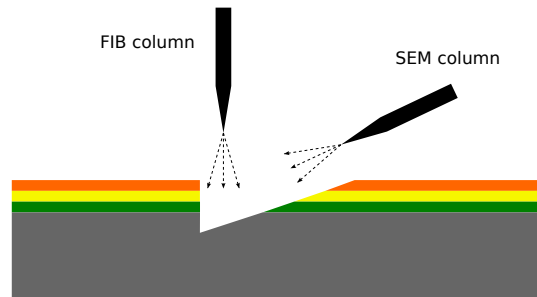
Figure 4.8: FIB and SEM columns to inspect the vertical profile of a sample. The FIB column mills while the SEM column examines the sample in real time.
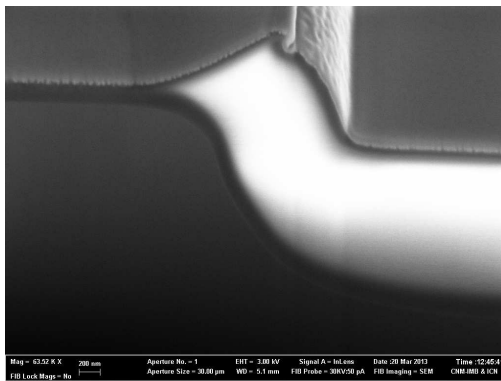
(a) After the LOCOS process.
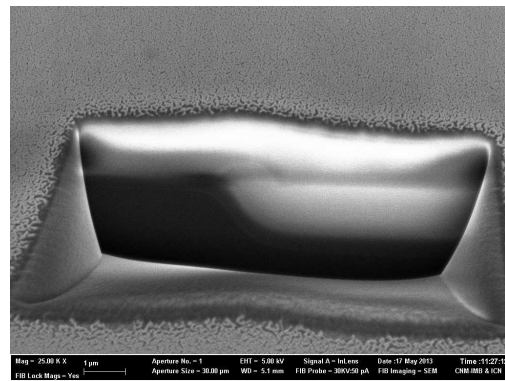
(b) After the TEOS+BPTEOS deposition and fluidification.
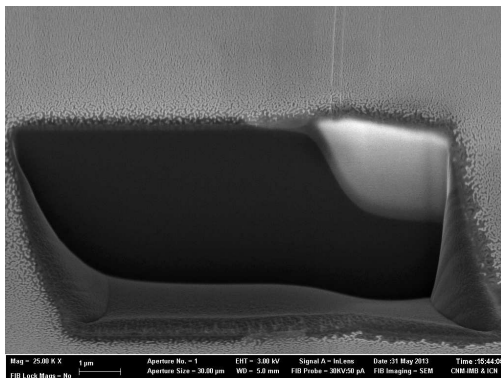
(c) After the etching.
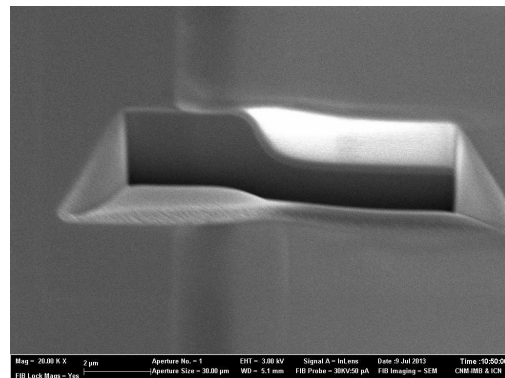
(d) Final set of samples.

Figure 4.9: FIB images of the trench at different stages of the process. Images (a), (b) and (c) correspond to a test sample, whereas image (d) corresponds to one of the samples of the final fabrication. All images by A. Mª Sánchez.

135
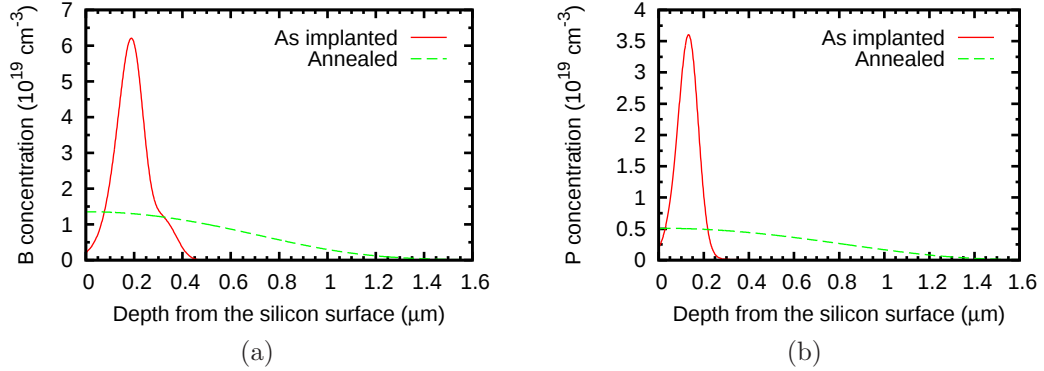
Figure 4.10: Implantation profiles before and after annealing at 1100 °C for 60 minutes corresponding to (a) boron in the p wells and (b) phosphorus in the n wells.

After etching the oxide a much smoother and planar structure is achieved, as shown in figure 4.9(c). Figure 4.9(d) shows the trench for the final samples that were actually used to fabricate the devices.

4.3.2 Implantation of the p-wells and the back of the wafer

The third challenge discussed in section 4.2.1 is related to the constraints that the thermal annealing required to generate the active layer imposes on our ability to tune the distribution of dopants in the implantation wells. In order to assess the feasibility of the PIN diode used as a detector, the implantation and diffusion were simulated in Sentaurus.

Figures 4.10(a) and 4.10(b) show the profiles of boron and phosphorous, respectively, in the p and n wells before and after the annealing.

According to the simulations, the junction line in the p wells should be about 2 μm below the silicon surface. Moreover, if we consider the diffusion length of electrons in p silicon with a concentration of acceptors of about $1 \times 10^{19} \, \text{cm}^{-3}$ (see figure 4.10(a)) to be $L_n \approx 1 \, \mu\text{m}$ [70], then we need a reasonable amount of photons to reach beyond 1 μm deep into the silicon. The electron-hole pairs generated before that point will not be able to diffuse long enough to reach the depletion region where the built-in electric field can drift them away and contribute to the photogenerated current.

Using $\alpha_{500} \approx 10^4 \, \text{cm}^{-1}$ for the absorption coefficient of light with 500 nm wavelength and $\alpha_{900} \approx 10^3 \, \text{cm}^{-1}$ for 900 nm wavelength [71], we can find the ratio of the incoming photons (those that penetrate the silicon surface) that can make it beyond the 1 μm goal using the well known Beer-Lambert law:

$$N(x; \, \alpha) = N_0 \, e^{-\alpha x} \tag{4.1}$$
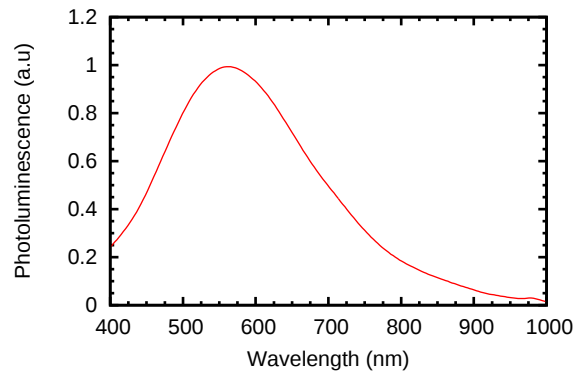
Figure 4.11: Photoluminescence spectra of the active layer.

where N_0 is the amount of photons at $x = 0$, x is the depth, α the absorption coefficient and N the remaining (unabsorbed) photons at x.

For α_{500} we find that 37% of the photons reach beyond the 1 μm limit, meaning that 37% of the generated electron-hole pairs will contribute to the photocurrent for light of 500 nm wavelength, whereas the figure goes up to 90% for α_{900}, that is when the light has 900 nm wavelength.

Even though that is just a rough calculation, it should be enough to show that the fabrication parameters are acceptable as far as the fabrication of the PIN diode goes.

4.4 Characterization

In this section we are going to review the results of the characterization of the transceiver. The PL of the SRN layer will be measured to check for consistency with previous results. The detectors will be characterized in order to determine their operation parameters. The resistance of the p wells will be measured to assess their adequacy as a substrate for the emitter. The IV characteristics and EL of the emitter will be measured and, finally, we will determine if there is coupling between the emitter and the detector.

4.4.1 Photoluminescence of the SRN layer

The PL of the active layer was measured with the setup shown in figure 2.9 in chapter 2. The PL spectrum, plotted in figure 4.11, was measured in several points of the wafer. The measurements in all points were very consistent both in spectrum and intensity, which is expected due to the uniformity in the thickness of the layer. The spectrum agrees well with that of other SRN layers fabricated during this work, such as the ones plotted in figure 2.12 in chapter 2.

The measurements in wafer 4 resulted in no PL observed at all since, as already noted in section 4.2.3, an error during fabrication removed the whole active layer.

(a) Long detector in a device with active area 1×1 mm^2.   Area of the detector 2.07 mm^2

(b) Short detector in a device with active area 0.5×0.5 mm^2. Area of the detector 0.22 mm^2.
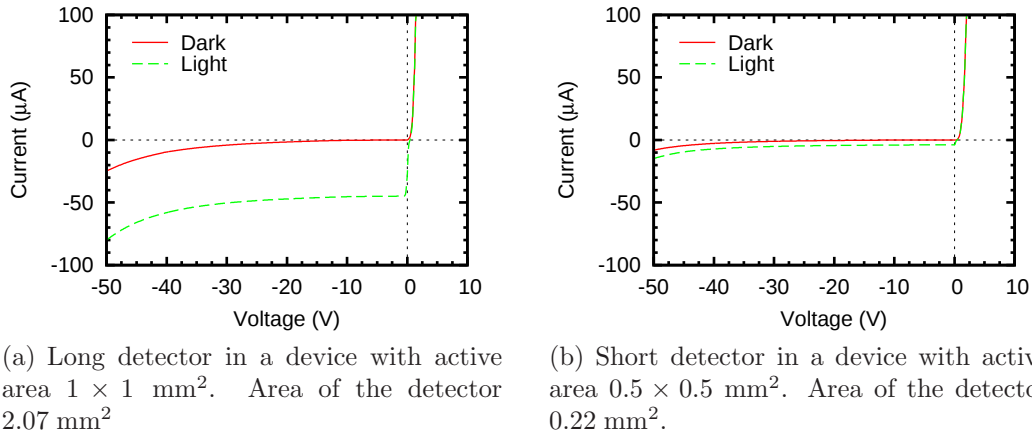
Figure 4.12: IV characteristics of two detectors of different size, with and without an external light.

4.4.2   Characterization of the detectors

The IV characteristics of the detectors was measured between -50 V (reverse bias) and 3 V (forward bias). The measurements were carried out in the dark and under illumination of a halogen lamp. The measurement setup was similar to that shown in figure 2.18 in chapter 2. The bias was applied between the detector contact and the back of the wafer (V_D in figure 4.1). The measurements across many detectors were consistent and the statistical deviations small.

Figure 4.12(a) shows the IV curves in dark and illuminated conditions, for a long detector of a device with active area 1×1 mm^2. The area of the detector is 2.07 mm^2. The IV curve corresponds with what is expected from a PIN diode. Furthermore, the addition of the halogen lamp to the measurement offsets the curve as expected.

The same measurement carried out in a smaller detector, in this case with area 0.22 mm^2, corresponding to a short detector of a device with active area 0.5×0.5 mm^2. The results, plotted in figure 4.12(b), show the same behavior. In this case the offset caused by the halogen lamp is smaller since the smaller area of the detector means that less light is collected.

In order to study the response of the detector to the external illumination, a laser source at 635 nm with variable power was directed towards the detector. The current was measured as a function of the laser power for two bias conditions, 0 V and -25 V. The results are shown in figure 4.13. The current in the detector increases with the power of the laser and is larger at reverse bias, just as expected.
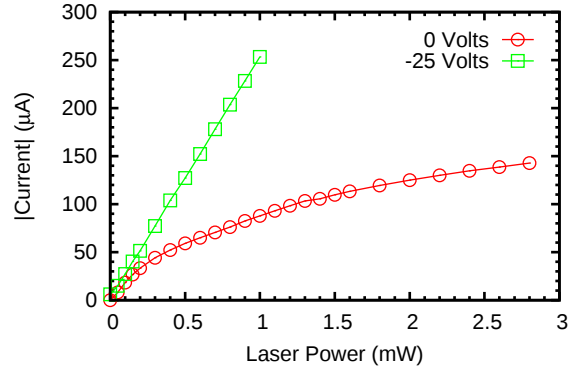
Figure 4.13: Photogenerated current for varying intensities of a 635 nm laser. The detector was reverse biased at 0 and -25 V.

4.4.3 Resistance of the p++ wells

The resistivity of the p wells should be measured in the four pads that were designed for that purpose. Unfortunately, a mistake was made in the design of level 5 of the mask set, corresponding to the windows in the passivation oxide. The windows are not opened where the four contact pads are. As a consequence, the aluminum pads rest on top of the passivation oxide.

As a workaround, the resistance between the two top contacts of the emitter (the V_E contact and the top ground in figure 4.1) was measured in several devices of wafer 6. Since we know that wafer 6 has no active layer, both aluminum contacts are directly in contact with the silicon p well.

Once the resistance is known, the sheet resistance can be calculated as:

$$R_s = R\frac{W}{L} \tag{4.2}$$

where R is the measured resistance, W is the width of the p well and L is the distance between the probes. The resistivity can be calculated from the sheet resistance as:

$$\rho = R_s \times t \tag{4.3}$$

where t is the depth of the p well, which we can gather approximately from the simulations presented in figure 4.10(a).

After the measurements, the figure of merit for the resistivity of the p wells is

$$\rho = 0.012 \pm 0.007 \, \Omega \, \text{cm}$$

The largest contribution to the error is by far the uncertainty in the depth of the well. The statistical deviation of the resistivity over several measurements is much smaller than the error. Therefore, the given mean value should be a good indicator of, at the very least, the order of magnitude of the resistivity, even if the measurements have been carried out in less than optimal circumstances.
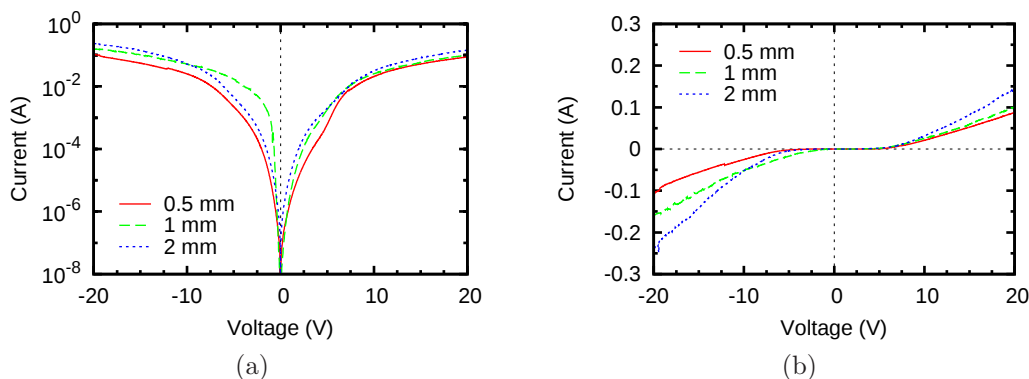
Figure 4.14: IV characterization of three emitters with different side lengths and polysilicon gate, corresponding to sample 2, in (a) logarithmic scale and (b) linear scale.

The measured best estimate of $\rho = 0.012\ \Omega\,\text{cm}$ is a good value for the resistivity of the p well, since the p substrates that have been regularly used for all our fabrications have resistivities in the range 0.1 and 1.4 $\Omega\,\text{cm}$.

4.4.4 IV characterization of the emitter

The IV characterization of the emitters has been carried out just like in every other electroluminescent device in this work, as described in chapter 2. Figures 4.14(a) and 4.14(b) show the typical IV characteristics of emitters with different sizes and polysilicon gate in a logarithmic and linear scale, respectively.

We can see clearly that the high currents, of the order of 10 mA at just 10 V, correspond to what we described in chapter 2 as a high conduction, or low resistance, state, and that we related with emission through points or edge emission.

Furthermore, we can see from the plot in a linear scale, figure 4.14(b), that, even though the current at any given voltage increases with the size of the emitter, it does not do so in the same proportion as the size. That is, the current in the biggest emitter is roughly twice as high as in the smallest one, whereas its area is 16 times bigger. If the current was flowing uniformly, then at any given voltage the current density would be the same in all devices, since the material is the same. Therefore, the total current in the 2×2 mm^2 emitters should be 16 times higher than the current in the 0.5×0.5 mm^2 emitters. The fact that it is only about twice as high instead of 16 times higher makes sense if we consider the current to be flowing through a finite number of highly conductive paths. Even though a larger area can potentially contain more paths, the relation does not have to be proportional.

For sample 3, with aluminum contacts, the IV characterization is plotted in figure 4.15. It shows a perfectly linear relation between the current and the voltage, which corresponds to an ohmic conduction and therefore indicates that the system
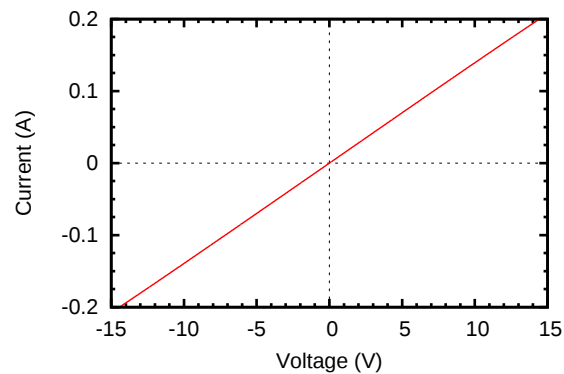
Figure 4.15: IV characterization of an emitter from sample 3, with aluminum gate.
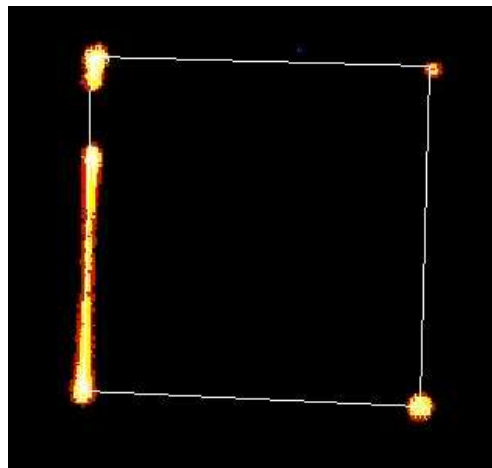
Figure 4.16: The EL was observed along the rim of the device. The white line marks the contour of the active area.

is short-circuited. This behavior was observed in all the devices that were measured, in different points of the wafer. That indicates that some problem occurred during the fabrication process.

4.4.5 Electroluminescence of the emitter

The EL of was observed along the rim of the devices with polysilicon gate, as illustrated in figure 4.16. That is very similar to the edge emission discussed in chapter 2. Given the high conduction of the devices, this kind of emission is not unexpected.

No EL was observed in the sample with aluminum gate, which is consistent with the short-circuit IV exhibited by that sample (figure 4.15).

The EL spectrum was measured in the same way as all other samples (see chapter
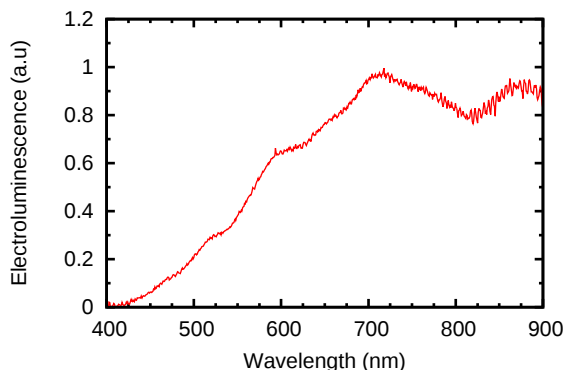
Figure 4.17: Typical EL spectrum of sample 2.

2). Figure 4.17 shows the the typical EL spectrum of sample 2, which is very similar to that of figure 2.24 in chapter 2.

4.4.6 Coupling between the emitter and the detector

In order to see if the transceiver works, we have to monitor the current in the reverse-biased detector while increasing the current in the emitter. Figure 4.18 depicts the connection scheme. V_s is the voltage/current source (Keithley 2430) that drives the emitter. V_b and I correspond to a Keithley 2100 source-measure unit that reverse-biases the detector at -30 V while measuring the current flow at the same time. The tests were carried out in a dark environment. The bottom panel in figure 4.18 shows an ideal schematic representation of the system, in which emitter and detector are electrically isolated from each other and the only possible coupling between the two is the light being transmitted through the nitride waveguide.

In figure 4.19, the lines with markers correspond to the measurements of the coupling between emitter and detector in several devices of sample 2, which we know emits light. It is clear that the current at the detector increases when the emitter current increases. In every case, it was checked that there was light emission from the active area during the test. The changes in the current at the detector, of the order of 10 nA, were similar to those observed after small changes in the residual ambient lightning, such as for instance switching on and off a computer screen in the otherwise dark laboratory room. Furthermore, it was observed that, in the case of device breakdown, the light emission would stop and a drop in the current at the detector would be immediately observed.

However, it is possible that the current in the detector is due to cross-talk between the emitter and the detector. In order to check that possibility, the same measurement was carried out on several devices of sample 5, which lack an active layer and therefore no light emission takes place. The results correspond to the lines without markers in figure 4.19. It is clear that there is no meaningful difference between the samples with and without light emission. Therefore, we conclude that
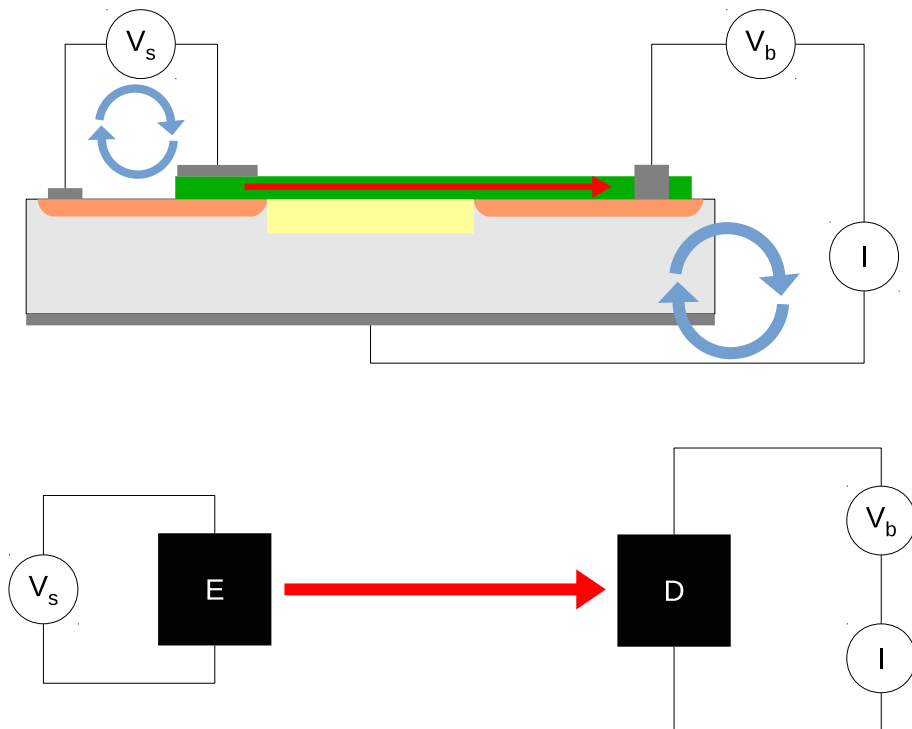
Figure 4.18: Top) Operation of the transceiver. V_s drives the emitter and V_b reverse-biases the detector. The photocurrent generated in the detector by the photons transfered along the guide (red arrow) is measured by I. Bottom) Schematic representation of the system. Emitter and detector are electrically isolated, with the only possible coupling between them being the light transmitted along the guide.
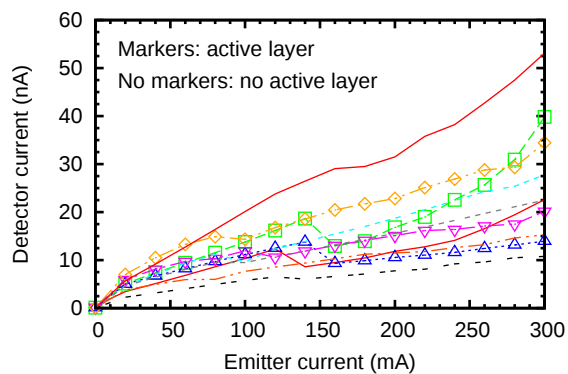
Figure 4.19: Current measured at the detector as a function of the current driving the emitter. The five curves with markers correspond to devices of sample 2, which have an emitting active layer. The five curves without markers correspond to devices of sample 5, which we know do not have an active layer.

no meaningful light detection is taking place and ascribe the current generated in the detector to carriers leaking from the p well of the emitter and making their way into the detector circuit.

4.5   Final thoughts and conclusion

Our first attempt at a monolithic, CMOS compatible, silicon based transceiver does not work. However, some important lessons can be drawn.

The basic design is correct as far as the electrical isolation between emitter and detector is concerned. Moreover, the detectors are much better than originally expected, especially considering that the amount of tweaking that can be done to their fabrication parameters is greatly limited by the requirements of the emitter.

Regarding the waveguide, there is unfortunately not much information that we can gather. Since the emission is poor, it is hard to say whether the guide is transmitting any light at all.

The weakest link in the system is the emitter. For one thing, even our best emitters show a poor emission, and we did not get close to fabricating one of our best emitters. This is in good part due to our inexperience with single SRN layers. Even though a previous fabrication indicated that we could achieve a reasonably good homogeneous electroluminescence in SRN layers, we lack control over their fabrication. At the beginning of this work it was not anticipated that the design that we were going to come up with for an integrated device would have to run on SRN instead of SRO.

In conclusion, we believe that, even though our design is promising, a more powerful emitter is needed in order to make it work at the prototype level.

Conclusion

We have characterized electroluminescent devices based on silicon rich oxide and/or silicon rich nitride. We have discussed the photoluminescence and structural characterization of the active layers and the electrical and electroluminescent characterization of full devices, all of them carried out in experimental setups built from scratch.

We have noted that the electroluminescence can appear in the form of discrete points scattered across the active area of the devices, in the form of emission along the rim of the active area, or homogeneously distributed across the area. These different kinds of emission have been related to the optical and electrical properties of the devices.

In the two former cases (emission through points an edge emission), the electroluminescence comes with high current densities, of the order of $1 \, \mathrm{A \, cm^{-2}}$, and low efficiencies of the order of 10^{-8}. On the other hand, the homogeneous emission comes with lower current densities, of the order of $0.01 \, \mathrm{A \, cm^{-2}}$, and better efficiencies, in the range $10^{-7} \sim 10^{-5}$. Therefore, we have concluded that the homogeneous emission is optimal in terms of efficiency. Furthermore, a simple model has been proposed to explain the appearance and occasional coexistence of the different kinds of emission. In this model, the homogeneous emission is related to uniform conduction in the device, whereas the emission through points takes place in low resistance preferential paths.

The effect of a nitride layer on top of the SRO has been explored, concluding that it helps in achieving a uniform conduction that favors the homogeneous emission in the active layer.

The conductivity states of the active layer associated with the different kinds of emission have been related with its CV behavior. The results of the study, which show that the homogeneous emission corresponds to well behaved CV curves, whereas the emission through points does not, reinforce the idea of a homogeneous conduction in layers which exhibit homogeneous emission, and a conduction through preferential paths in the case of emission through points.

The injection mechanisms in PECVD and ion implanted samples have been

studied and compared, concluding that no single emission mechanism can account for the injection at all regimes in the studied range of electric fields. Fowler-Nordheim tunneling or trap assisted tunneling have been found to play a significant role in PECVD samples. In implanted samples, Fowler-Nordheim tunneling dominates at low fields, whereas Poole-Frenkel is more likely to be the dominant mechanism at higher fields.

Comparison of the photoluminescence and electroluminescence spectra of bilayers SRO/SRN, allows us to conclude that each layer contributes a different band in the total emission, which results in a wider distribution of the energy across the visible spectrum. A wider emission spectrum provides a larger pool of wavelengths to choose from in case one wants to select a particular wavelength of the spectrum. Furthermore, being able to modulate two different bands gives more control over the final spectrum. However, more work is required on the SRO/SRN bilayer systems in order to control the emission of each band independently.

The comparison between the photoluminescence and electroluminescence has revealed massive differences in their spectra, which have been qualitatively attributed to interference effects.

A computer software based in the Crawford method for the study of the interference effects in multilayer stacks has been presented. The program has been used to quantitatively study the interference effects in the emission of our devices. We can conclude that the photoluminescence and electroluminescence spectra are indeed the same despite their apparent difference. Our analysis has also made it apparent that a quantitative understanding of the interference effects in the system is of utmost importance in order to draw valid conclusions regarding the origins of the luminescence based on the spectra of the active layer.

We have presented the design, fabrication and characterization of a CMOS compatible optical transceiver, and two main challenges in the integration of the emitter, waveguide and detector have been successfully overcome, namely achieving a reasonably flat and uniform silicon oxide trench and a good detector.

In the end, the transceiver has not worked as expected, most likely due to a poor SRN emitter. More work is required in order to better control the fabrication process of the SRN layers and therefore predict the emission of the fabricated layer. However, we believe the basic design to be valid, given the low electrical coupling detected between the emitter and the detector components of the transceiver.

Software used in this thesis

A good theoretical scientist must know the theories he works with and all their intricacies. A good experimentalist must know the experimental setups he works with and all the devices that intervene in his measurements, because it is important to know if the measurement apparatuses are working within their intended range and what they are actually measuring. Much in the same way, a scientist carrying out simulations must know the software and algorithms being used in his simulations, for instance their domain of applicability, in order to get the best results from them.

The theorist, the experimentalist and the simulator, all of them use a computer in their everyday work, be it for writing code to check theories or devise new simulations, programming experimental devices, analyzing and plotting experimental data, writing papers and reports, and managing the files generated for a particular project. All those tasks require some kind of software, and being proficient in its use is as important for all of them as being proficient in the theory is important for the theorist, in the experimental setup for the experimentalist, and in the algorithm for the simulator.

Since a scientist spends a good amount of time using some kind of software, it is reasonable to reserve a few lines to outline the software that has been used for the work that has resulted in this thesis.

Most of the software used for the present work has been run under Linux, in particular Debian/GNU Linux[1]. Some tasks have been carried out under Windows XP or Windows 7[2], most notably the work in the lab, and a few tasks here and there have been accomplished under Mac OS X[3].

Whereas any major operating system has the capabilities to run any of the software necessary for carrying out the tasks required in a scientist' everyday life, Windows is by far the most undesirable of the lot. Too much time is spent fixing stuff that used to work the day before instead of just accomplishing tasks. Even in the very static systems that are usually found in the lab, where the computers are

---

[1]`www.debian.org`
[2]`windows.microsoft.com`
[3]`apple.com/osx`

used for a very limited set of tasks, things break down more often than they should. Fewer Windows licenses would also mean more money available for other areas.

That being said, there is still some software that may not be available for Linux or Mac OS X, for instance the FIMMWAVE[4] waveguide mode solver.

As far as the experimental setup goes, the control software has been written in LabView[5]. Although LabView is an impressing piece of technology and it certainly has its uses, I reckon mostly in the industry, it is possibly the worst software I have found myself using in a long time. I must say that I have hated every minute I have spent programming in LabView. There is absolutely nothing intuitive about the graphical programming paradigm it endorses.

According to the website, LabView allows you to *"program the way you think"*. See, different people think in different ways, so no matter which programming language you use, it is *always* the programmer who shapes his mind to accommodate the language being used and the available programming structures, not the other way around. And LabView is certainly *not* an exception. Saying otherwise is just lame marketing nonsense.

Again according to the website, LabView *"is a development environment for problem solving, accelerated productivity"*. Sadly, I spent way too much time solving problems generated *by* LabView, rather than the original problem I was trying to solve. Simple tasks that would have been solved in literally seconds with a traditional, text based programming language, took hours in LabView. And while being more experienced in the language mitigated the issues *somewhat*, LabView is inherently problematic in some areas like moving data around between different VIs. That turns the whole *"accelerated productivity"* claim into a bad joke, and that is before factoring in the well known *"spaghetti code"* that any program beyond a lame textbook example inevitably turns into.

In addition, National Instrument's Linux support leaves a lot to be desired. A lot. At the time of this writing, the latest NI.488.2 drivers for Linux are 18 months old, and therefore are only supported in 18+ months old distributions. Furthermore, according to the readme file, they can only be used from 32-bit applications. You read that right. It is 2014. 32-bit applications.

For all those reasons, my recommendation as far as instrument control goes is: find the drivers for your interfaces (National Instrument's GPIB drivers are fine if you can manage to make them work in Linux using a Kernel other than the exactly precise version they were compiled for and under the exactly precise distribution they are supported for; otherwise, go for the linux-gpib package at `linux-gpib.sourceforge.net` and you will spare yourself a lot of frustration), and write your control program in C or whatever language you can find bindings for. At the end of the day controlling instruments is just opening a device, storing the handler in a variable, and send text commands to the handler, which you can find in the documentation that comes with the device. Believe me, you will save *a lot*

---

[4]`http://www.photond.com/products/fimmwave.htm`
[5]`ni.com/labview`

of time. The only scenario in which you *might* actually get things done faster in LabView is if you want a graphical user interface for your program, which in most cases is far from necessary and adds very little, if anything, to the usefulness of the program.

Another software used in the lab is SpectraSuite[6], now being replaced by Ocean-View[7], for capturing spectra with the Ocean Optics spectrometers.

In a related area, all the data generated by our measurements has been dumped into plain text files (no directly generating Excel files or similar shenanigans). The reasons are obvious. A plain text file is software independent, platform independent, architecture independent, and pretty much anything-independent. Once you have your plain text file with the data, it can be manipulated as required in order to be treated by whatever software you need to use.

In order to perform non mathematical manipulation of data files (such as adding comment markers to files generated by programs that do not add comment markers to the comment lines, for instance SpectraSuite) I have used Bash[8]. Bash is a wonderful scripting language that you will find in pretty much any Unix-like operating system out there. There is no need to be an absolute guru, but an advanced beginner level of knowledge in Bash is a good investment no matter what you do with a computer.

For mathematical manipulation and plotting of data files I have used either Python[9] (see the discussion in section 3.3.1) followed by Gnuplot[10], or Qtiplot[11].

Qtiplot is an open source plotting and data analysis tool similar to Origin[12]. It does pretty much everything Origin does, but for free (or very little money if you want to purchase precompiled binaries). There is absolutely no reason to drop \$69 for a 1-year student license (or \$500 for 1-seat license for academic use, \$800 in the case of Origin Pro) when Qtiplot probably does exactly what you need for free. Unless you need some killer feature that Origin has and Qtiplot does not, it would be insane to pay that ridiculous amount of money. It is hard to believe how Origin has become so ubiquitous. Plus, it only works in Windows, whereas Qtiplot runs natively on Windows, Mac and Linux.

Although Qtiplot is quite useful, as of late I have been leaning more and more toward a split work flow consisting of Python for data manipulation and analysis and Gnuplot for plotting. It is entirely possible to streamline such a work flow to make it efficient although, admittedly, in some scenarios the flexibility of Qtiplot is more convenient.

Gnuplot is an open source plotting program that works in almost any computer that has ever been fabricated. It is a wonderful application but it is important to

---

[6]http://www.oceanoptics.com/Products/spectrasuite.asp

[7]http://www.oceanoptics.com/oceanview/homepage.asp

[8]http://www.gnu.org/software/bash/

[9]www.python.org

[10]gnuplot.info

[11]soft.proindependent.com/qtiplot.html

[12]http://originlab.com/

note what it is *not*. It is *not* a data analysis tool. It is *not* a jack of all trades. It does one thing, plotting data, but it does it well. It can be used for *graphical* analysis and it allows curve fitting, but it remains essentially a plotting program. All the graphs in this thesis, as well as some diagrams like those in figures 2.2, 3.2, 3.3 or 4.1, among others, have been generated with Gnuplot.

As far as writing reports goes, LaTeX[13] has been the chosen engine in the vast majority of occasions, including the present thesis. Beamer[14] (a LaTeXpackage) has been used for slide composition. Only in very few occasions the LibreOffice[15] suite has been used instead. As for plain text editing, Kile, Kate and Vim have been my top choices (Vim being among my favorites as of late).

Finally, as already discussed in section 3.3.1, all the major coding has been done in Python, except for a script to fit injection models to the current-voltage measurements, which was coded in Matlab[16] (for no other reason than getting to know the language). Sage[17] has been used occasionally, but mostly as a high level Python interpreter, rather than an open source Mathematica replacement.

I have tried to rely on free (as in *free speech*) software as much as possible. I believe free software is particularly important in the research world. You do not want *your* code, *your* data, or *your* reports to be hostages of a private company, and you do not want the *access* to your code, data or reports to depend on your access to a license of a particular software. For example, all the code that you write in Matlab or Mathematica will be available for execution by you, its creator, as long as you pay your license. You may have a license available today, but who knows if you will have it tomorrow. Same goes for the reports you write in Microsoft Word, the presentation you compose with Microsoft PowerPoint, or the data you save in some proprietary binary format used by some measurement software for your instrumentation.

---

[13]http://www.latex-project.org/
[14]https://bitbucket.org/rivanvx/beamer/overview
[15]libreoffice.org
[16]http://www.mathworks.com/products/matlab/
[17]sagemath.org

A tutorial to the multilayers Python module

B.1   Introduction

The multilayers Python module allows you to calculate the effect of a multilayered system on the propagation of light. In particular, you will be able to calculate the characteristic matrices of the whole system or any of its layers, the reflection and transmission coefficients, the reflectance and transmittance [61].

Furthermore, consider a radiative center at some point of the multilayered system whose emission is detected by an observer at infinity. The multilayers module will allow the user to calculate the ratio of the energy detected by the observer in the presence of the multilayer to the energy he would receive if there was no multilayer (i.e. if the whole system was made of the same medium the observer is standing on). In order to calculate this, the module makes use of a method developed by O. H. Crawford and published in ref. [60].

In what follows, the inputs and outputs on the screen will be typed in `verbatim` font. The dollar sign (`$`) will represent the Bash prompt, with the current directory being printed before the `$` sign. The `>>>` symbol will represent the Python prompt.

B.2   Dependencies

In this section you will find what you need to have in order to use the multilayers module.

B.2.1   Python

You will need Python to use the multilayers.py module. In the web page [62] you will find links to installers for all major operating systems. Linux users can just install it from the package manager of their distribution.

To follow this tutorial it is not strictly necessary to know Python because the syntax of the statements we will use is very straightforward. However, a basic knowledge of some programming language is required in order to at least grasp the idea of what we're doing.

B.2.2 Numpy

Numpy is a Python module that provides a suitable array type for numerical calculations. In the web page [63] you can find installers for all major operating systems. As with Python, Linux users can get it from their package manager.

B.2.3 Scipy

Scipy is a Python module for mathematics, science and engineering. As with Numpy, in the web page [72] you can find installers for all major operating systems. Linux users can get it from their package manager.

B.2.4 Git

Git is a version control system. You will need this to get the multilayers module from github [73]. You can find installers for all major operating systems on the web [74]. Linux users can get it from their package manager.

B.3 Installation

To install the multilayers module you just have to clone it from a public repository in github [73]. If you already have git installed, cloning the repository is as easy as typing:

```
~$ git clone git://github.com/tortugueta/multilayers.git
```

That will create a `multilayers` directory and store all the files of the project inside. See the file `README.md` to learn more about the contents of the project.

Before you begin, if you plan to work in a directory other than the default `multilayers` (which is most likely) you will need to add that directory to your `PYTHONPATH` environment variable. In Bash you can do this by typing:

```
~$ export PYTHONPATH=${PYTHONPATH}:/home/user/multilayers
```

Replace `/home/user/multilayers` for the appropriate directory. This will allow you to import the multilayers module even if you are not in the directory where multilayers.py is located. If you want to make this change permanent, add that line in the `.bashrc` file located in your home directory.

Once this is done, you can go to the `tests` directory and execute the unit tests for the module. This checks that everything is working as expected:

```
~/multilayers/tests$ python test_multilayers.py

[some output lines skipped]

----------------------------------------------------
Ran 26 tests in 0.986s

OK
~/multilayers/tests$
```

There will be many error lines in the output. That is fine. The important thing to note is the end of the output, where it should say OK. If it says FAIL, then something is wrong and you should report that.

B.4 Usage

Here you will learn how to use the multilayers module. You can start by launching the Python interpreter:

```
~$ python
Python 2.6.6 (r266:84292, Dec 26 2010, 22:31:48)
[GCC 4.4.5] on linux2
Type "help", "copyright", "credits" or "license" for more
information.
>>>
```

You must import the multilayers module before you can use it:

```
>>> import multilayers as ml
>>>
```

We will also use Numpy so we will import it as well:

```
>>> import numpy as np
>>>
```

B.4.1 Defining the mediums

A medium will be defined by its complex refractive index $\tilde{n}$, which is a function of the wavelength λ. From now on, I will use the term *refractive index* to refer to the real part n of the complex refractive index $\tilde{n}$. The term *extinction coefficient* will be used to refer to the imaginary part k of the complex refractive index $\tilde{n}$.

In order to define a medium, you will need a plain text file with three columns: one for the wavelength, another for the refractive index n and a third for the extinction coefficient k. In the examples subdirectory of the multilayers directory you can find example files for air, silicon, silicon dioxide, silicon nitride and polysilicon.

To load a medium just type:

153

```
>>> dielectric = ml.Medium("examples/sio2.dat")
>>>
```

Here we load the table of refractive indices stored in `examples/sio2.dat`. Replace that for the path to the file you want to use. This creates a `Medium` object and we assign this object to a variable called `dielectric`.

If your file has more columns than necessary, they are in an order other than $(\lambda \; n \; k)$, they are separated by characters other than `TAB` or the comments start by something other than `#`, you can pass some options to the previous call that will allow you to tweak the loading of the table. See the reference to the `Medium` class for more information about that.

Once we have the medium loaded, we can do a number of things with it. First, note that even though your file has the refractive indices sampled at a discrete number of wavelengths, you can get the complex refractive index at any wavelength within the range defined by the maximum and minimum wavelengths of your sampling. That is because the program builds an interpolator with the provided data. This is important because that frees you of the necessity to make sure that all your materials are sampled at precisely the same wavelengths.

To recover the range where you can interpolate the refractive index, just type:

```
>>> dielectric.getMinMaxWlength()
(160.0, 2500.0)
>>>
```

You can check that 160.0 nm and 2500.0 nm are the minimum and maximum wavelengths in the file `examples/sio2.dat`.

To find out the complex refractive index at any wavelength within the interpolable range, for instance 502 nm, type:

```
>>> dielectric.getRefrIndex(502)
(1.4622175947399658+0j)
>>>
```

Notice that we get the complex refractive index, although in this case the extinction coefficient is zero. Also, notice that the interpolating function passes through the sampled points, which means that you can recover exactly the values at the sampled wavelengths (within the roundoff error)

```
>>> dielectric.getRefrIndex(500)
(1.4622999999999993+0j)
>>>
```

You can check in the `examples/sio2.dat` file that the value of the refractive index at 500 nm is 1.4623.

The units of the wavelength in your files may be anything, but you have to be consistent during the program and use always the same units.
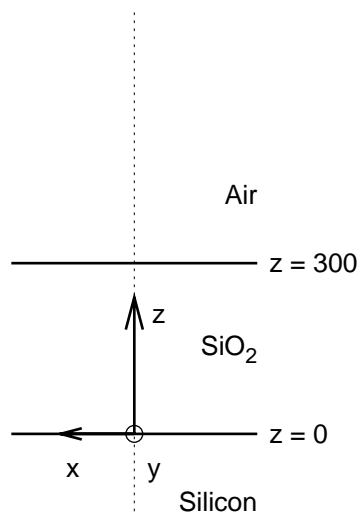
Figure B.1: Schematic view of our layered system

B.4.2 Defining the multilayer system

Before building our multilayer system, let's load two more materials.

```
>>> air = ml.Medium("examples/air.dat")
>>> substrate = ml.Medium("examples/silicon.dat")
>>>
```

Now we want to work with a simple multilayer system consisting of a dielectric layer 300 nm thick on top of a silicon substrate. The medium above the layer will be air[1]. See figure B.1. We can generate the `Multilayer` object as follows:

```
>>> multilayer = ml.Multilayer([
...         air,
...         [dielectric, 300],
...         substrate])
>>>
```

You can type the whole thing in a single line, but typing it like we just did makes the structure that we are creating more visually apparent. We have a *top medium* made of air, then a 300 nm thick layer of dielectric, and a silicon substrate (that we will also refer to as *bottom medium*). We have assigned the newly created `Multilayer` object to a variable called `multilayer`.

---

[1]Here we will consider the silicon substrate to be a semi infinite medium just like the air above the dielectric. Since typically the substrate will be much thicker than the layer, it is fine to make that approximation. We could also define the layer strictly and consider the silicon substrate to be a layer of finite thickness and use another layer of air as the bottom medium instead of the silicon.

Important! Note that the thickness must be in the same units as the wavelengths.

B.4.3 Setting the state of the system

Before we can calculate anything, we have to set the state of the system, i.e. the wavelength of the light propagating across the system, the angle of propagation and the polarization. Note that setting the wavelength fixes the complex refractive index in effect, and that affects the propagation angles (through Snell's law). Therefore, you must set the wavelength *before* the propagation angle. The polarization can be set first or last, since it does not affect the other variables.

Set the wavelength to 500 nm with the `setWlength` method, which gets the wavelength as its only argument:

```
>>> multilayer.setWlength(500)
>>>
```

We can check that we are currently working at 500 nm with the `getWlength` method:

```
>>> multilayer.getWlength()
500.0
>>>
```

Now we can set the propagation angle. The angles are measured with respect to the z axis. So $\theta = 0$ corresponds to propagation perpendicular to the interfaces. What we do is fix the propagation angle in one of the layers of the system with the `setPropAngle` method, which expects two arguments: the angle in radians and the index of the layer where we are fixing the angle. Each layer in the system is identified by a consecutive index starting with 0 in the top medium. Therefore, in our system the air will have index 0, the dielectric layer index 1, and the silicon substrate index 2.

The propagation angle in the other layers will be automatically calculated according to Snell's law. Here we will be interested in the light that comes out of the system at 25 degrees. Therefore, we will fix the propagation angle to 25 degrees in the top medium (index 0):

```
>>> multilayer.setPropAngle(np.deg2rad(25), 0)
>>>
```

Note that the `setPropAngle` method expects the angle in radians. We can use the `deg2rad` method in `Numpy` to convert from degrees to radians. If the index is not specified 0 will be used as a default value.

Now we can check that the propagation angle is correctly set to 25 degrees (≈ 0.436 rad) in the top layer and all the other angles have been calculated correctly. We do that with the `getPropAngle` method, which gets the index of the layer as an argument.

```
>>> multilayer.getPropAngle(0)
(0.43633231299858238+0j)
>>> multilayer.getPropAngle(1)
(0.29319178759065995-0j)
>>> multilayer.getPropAngle(2)
(0.09847162694380529-0.0016750238752748126j)
>>>
```

Now, notice that the output is always a complex number. Internally, the program always works with complex angles because in some situations the angle *must* be complex, for example when light propagates beyond the critical angle. In layers 0 and 1 we can see that the angle is formally real in the sense that the imaginary part is zero, as it should be. However, in layer 2 (the silicon substrate) the angle has a nonzero imaginary part. That is because in that layer we have a nonzero extinction coefficient. In that kind of medium the propagation angle does not really bear the meaning of a true propagation angle, just like the complex angle we get beyond the critical angle is not really a propagation angle. You can find more on that in ref. [61].

Now only the polarization remains to be set. We use the `setPolarization` method to accomplish this, which gets either `'te'` or `'tm'` (case insensitive) as an argument:

```
>>> multilayer.setPolarization('te')
>>>
```

We can check the current polarization with the `getPolarization` method:

```
>>> multilayer.getPolarization()
'TE'
>>>
```

B.4.4 Calculating the matrices

Once we have set the wavelength, propagation angle and polarization, we can calculate the characteristic matrices of each individual layer in the system. We do that with the `calcMatrices` method, which gets a list with the indices of the layers whose characteristic matrices we want to calculate. An empty list or no argument at all will result in the characteristic matrices of all layers to be calculated.

Since we need all the characteristic matrices, we do:

```
>>> multilayer.calcMatrices()
>>>
```

Note that, since the thickness of the top medium and the substrate is infinite, the characteristic matrix does not make sense in those two layers. Therefore, when we instruct the program to calculate the matrices of all the layers as we have just

done, in fact it will only calculate the matrices of the intermediate layers (in our case only the silicon dioxide) but not those of the top and bottom mediums.

Now we can see the matrix of each layer if we want. We use the `getMatrix` method, which gets the index of the layer as its only argument. The matrix of the silicon dioxide layer is:

```
>>> multilayer.getMatrix(1)
matrix([[ 0.535502+0.j       ,   0.000000+0.603282j ],
        [ 0.000000+1.182260j,   0.535502+0.j        ]])
>>>
```

The characteristic matrix of the top and bottom mediums is internally stored as the `None` value. If we try to print it we get nothing:

```
>>> multilayer.getMatrix(0)
>>> multilayer.getMatrix(2)
>>>
```

Once we have the characteristic matrices of all the individual layers of the system, we are ready to calculate the global characteristic matrix of the multilayered system, which in fact is the matrix product of the matrices of all the layers. We do that with the `updateCharMatrix` method:

```
>>> multilayer.updateCharMatrix()
>>>
```

The `updateCharMatrix` method calculates two matrices. One corresponding to the propagation in the direction from the top medium to the substrate, and another in the reverse direction. The matrices are in general different since the matrix product is not commutative.

The methods `getCharMatrixUpDown` and `getCharMatrixDownUp` return the matrices. The matrix corresponding to propagation from the top medium to the substrate is:

```
>>> multilayer.getCharMatrixUpDown()
matrix([[ 0.535502+0.j       ,   0.000000+0.603282j ],
        [ 0.000000+1.182260j,   0.535502+0.j        ]])
>>>
```

For propagation from the substrate to the top medium we have:

```
>>> multilayer.getCharMatrixDownUp()
matrix([[ 0.535502+0.j       ,   0.000000+0.603282j ],
        [ 0.000000+1.182260j,   0.535502+0.j        ]])
>>>
```

Of course, in this very simple system the characteristic matrix of the whole system is equal to the matrix of its only layer (remember that the top medium and the substrate do not count towards the calculation of the characteristic matrix), and it is trivially the same in both directions too.

B.4.5 Getting the coefficients

When we call the `updateCharMatrix` method, the program calculates not only the characteristic matrices, but also the reflection and transmission coefficients, the reflectance and the transmittance.

The methods `getCoefficientsUpDown` and `getCoefficientsDownUp` are used to fetch all those parameters. These methods return a dictionary. The reflection coefficient is stored with the key 'r', the transmission coefficient with the key 't', the reflectance with 'R' and the transmittance with 'T'. In the following lines I will store the dictionary in a variable called `coefUpDown` and then recover each key in different statements:

```
>>> coefUpDown = multilayer.getCoefficientsUpDown()
>>> coefUpDown['r']
(-0.053133453578214605+0.47742391532777739j)
>>> coefUpDown['t']
(0.24601384152987718-0.32014850303761927j)
>>> coefUpDown['R']
0.23075675881605304
>>> coefUpDown['T']
(0.76924324118394694+0.01316999164591307j)
>>>
```

The reflection and transmission coefficients are always complex numbers because they carry information on the phase of the wave. The reflectance will always be a real number. However, the transmittance may be real or complex. It will be complex whenever we have a complex angle or when the refractive index of either the top or bottom mediums have nonzero extinction coefficient.

In this case, we have a reflectance $R \approx 0.23$, which means that if we have a wave propagating from the air towards the SiO_2–on–silicon system, 23% of the energy will be reflected back towards the air (that is for the wavelength 500 nm and incidence at 25 degrees that we set earlier).

Now we will do the same for the propagation from the silicon towards the SiO_2–on–air system. This time we will skip the intermediate step of storing the coefficients in a variable.

```
>>> multilayer.getCoefficientsDownUp()['r']
(0.45063140231719617-0.14947982025241136j)
>>> multilayer.getCoefficientsDownUp()['t']
(1.1867461704360733-1.490830824900075j)
```

```
>>> multilayer.getCoefficientsDownUp()['R']
0.22541287741705593
>>> multilayer.getCoefficientsDownUp()['T']
(0.76924324118394671-0.013169991645913069j)
>>>
```

Notice that, although the characteristic matrix was the same for both directions of propagation, the coefficients are not since the system is not at all symmetric (the input and exit mediums are interchanged).

B.4.6 Changing the state

So, we have seen that for TE polarized light with $\lambda = 500$ nm and incidence at 25 degrees from air towards a 300 nm thick silicon dioxide layer on top of silicon substrate the reflectance is 0.23. What if we now want to see what happens when $\lambda = 600$ nm?

Basically, we have to set the state again, following the same order: first the wavelength (because that will change the propagation angles), then the propagation angle and finally the polarization (again, the polarization can be set at the beginning; the only requirement is that if we change the wavelength, we have to set the propagation angles again). In this case, since we want to keep the polarization in TE mode, we don't need to change it.

Every time we change the state we also have to recalculate the characteristic matrices. Therefore, we type:

```
>>> multilayer.setWlength(600)
>>> multilayer.setPropAngle(np.deg2rad(25), 0)
>>> multilayer.calcMatrices()
>>> multilayer.updateCharMatrix()
>>>
```

Now let's see how the reflectance has changed:

```
>>> multilayer.getCoefficientsUpDown()['R']
0.13013272824116065
>>>
```

So, at 600 nm the reflectance drops to 13%.

To finalize this section, let's see how to scan the reflectance of this system for a range of wavelengths, say between $\lambda = 400$ nm and $\lambda = 800$ nm every 50 nm.

```
>>> # Generate the list of wavelengths we want to scan
>>> wlength_list = np.arange(400, 800 + 50, 50)
>>>
>>> # Initialize the array where the results will be
>>> # stored
```

```
>>> reflectance = np.empty(wlength_list.size)
>>>
>>> # Loop over the wavelengths
>>> for (index, wavelength) in enumerate(wlength_list):
...     multilayer.setWlength(wavelength)
...     multilayer.setPropAngle(np.deg2rad(25), 0)
...     multilayer.calcMatrices()
...     multilayer.updateCharMatrix()
...     reflectance[index] = \
...         multilayer.getCoefficientsUpDown()['R']
...
>>> # Print the results
>>> reflectance
array([ 0.49229979,  0.42306659,  0.23075676,
        0.09724455,  0.13013273,  0.22453337,
        0.29978883,  0.34321062,  0.360822  ])
```

Note that the result is not a 3×3 matrix but a 9 element array with its contents spread across three lines.

B.4.7 The F function

In order to understand what the F function is and what is involved in its calculation, have a look at ref. [60]. In short, imagine the layered system we built in section B.4.2 and a coordinate system where z is in the direction perpendicular to the interfaces between layers (figure B.1).

Now imagine there is a dipole in $z = z_0$ oscillating along the direction i, $i = x, y, z$ and emitting light at wavelength λ. $F_i(z_0, \lambda, \theta)$ is the ratio between the electric field that an observer at $z = \infty$ would measure with an angle of incidence θ with respect to the z axis, to the electric field he would measure if the whole system was the same medium that we have at ∞ (in this case air). So, if $F_i(z_0, \lambda, \theta) = f$ it means that the observer measures an electric field f times what he would measure if there was no layered system (everything air).

F is a complex number because it carries information about the phase of the wave. In general we will want to square the modulus of F to get the ratio of the energy rather than the electric field.

The methods `calculateFx`, `calculateFy` and `calculateFz` are used to calculate the F function for dipoles oscillating along the x, y and z axis respectively. They receive three arguments: the z coordinate of the dipole, the wavelength of the emitted light, the angle of propagation and the index of the layer where we are fixing the angle (0 if not specified).

When we execute the methods, the state of the system is automatically set according to the arguments. That means that, once we have the multilayer defined as we did in section B.4.2, we can readily execute `calculateFx`, `calculateFy` or

`calculateFz` without setting the state. It will be set or changed automatically if necessary. Also, note that `calculateFy` will set the polarization to TE while `calculateFx` and `calculateFz` will set the polarization to TM.

So, for a dipole oscillating along y, located at $z_0 = 150$ and considering light at 25 degrees in the air we have:

```
>>> multilayer.calculateFy(150, 500, np.deg2rad(25), 0)
(0.0044385213156725055+0.68657782956390412j)
>>>
```

We are probably more interested in the energy stored in the TE and TM waves. For the TE wave let's calculate $|F_y|^2$, but now for propagation perpendicular to the interfaces:

```
>>> theta = 0
>>> fy = multilayer.calculateFy(150, 500, theta, 0)
>>> np.absolute(fy)**2
0.34198120861635478
>>>
```

That means that only $\approx 34\%$ of the energy of a TE wave at 500 nm propagating perpendicular to the interfaces and generated by a dipole in the middle of the silicon dioxide layer would reach us compared to what we would get if there was no multilayer. Interference in the layered system is responsible for the loss.

For a TM wave we want to calculate $|F_x \cos^2 \theta + F_z \sin^2 \theta|^2$:

```
>>> fx = multilayer.calculateFx(150, 500, theta, 0)
>>> fz = multilayer.calculateFz(150, 500, theta, 0)
>>> np.absolute(fx * np.cos(theta)**2 + \
...       fz * np.sin(theta)**2)**2
0.34198120861635478
>>>
```

Notice that the results for TE and TM waves are the same. This is expected since we are considering propagation perpendicular to the interfaces, and in this circumstance the TE and TM modes are indistinguishable.

We could repeat the calculation for different states by adding one or more nested `for` loops that cycle over different values of z, λ or θ.

B.4.8 A note about exceptions

The multilayers module will not let you do something that would result in a wrong result. For instance, suppose that you change the state of your system, let's say the wavelength. We know that the propagation angles will change, and so will the characteristic matrices. So, what happens if we change the wavelength of the system and then request the reflectance of the system without updating the angles and the matrices first?

```
>>> multilayer.setWlength(600)
>>> multilayer.getCoefficientsUpDown()['R']
>>>
```

We get nothing. That's because when we change the wavelength, the angles and characteristic matrices are reset to `None`. Now, what if you try to update the characteristic matrices without first setting the angles again?

```
>>> multilayer.calcMatrices()
Error: the propagation angle is not set
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
File "multilayers.py", line 934, in calcMatrices
raise ValueError
ValueError
>>>
```

Note the first error line after the statement: `Error:  the propagation angle is not set`. So, we cannot calculate the matrices because we have to update the angles first. Let's do it.

```
>>> multilayer.setPropAngle(0)
>>>
```

Now say that we want to update the global characteristic matrix but we forget to calculate the individual matrices first:

```
>>> multilayer.updateCharMatrix()
Error: the characteristic matrix cannot be calculated
because one of the individual matrices has not been
calculated
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
File "multilayers.py", line 1019, in updateCharMatrix
raise ValueError
ValueError
>>>
```

So, according to the error, we forgot to calculate the individual matrices. Let's do everything now:

```
>>> multilayer.calcMatrices()
>>> multilayer.updateCharMatrix()
>>> multilayer.getCoefficientsUpDown()['R']
0.096126056894384776
>>>
```

In short, the module will force you to do things properly before you can get a result. Note that when calculating the F function everything is updated automatically, so you don't need to manually set the state and calculate the matrices, as we saw in section B.4.7.

Step list for the fabrication of the transceiver

Table C.2 lists in detail all the steps carried out for the fabrication of the transceiver described in chapter 4. The columns labeled with the numbers 1–7 correspond to each different wafer in the run. A circle indicates that a particular step is carried out for the wafer corresponding to that column. In table C.1 the intended use of each wafer is described.

Num.	Use
1	Photoluminescence characterization
2	Transceiver – Poly gate (30 nm guide)
3	Transceiver – Al gate (30 nm guide)
4	Photoluminescence characterization
5	Transceiver – Poly gate (150 nm guide)
6	Transceiver – Al gate (150 nm guide)
7	FIB analysis

Table C.1: Intended use of the wafers.

Num.	Code	Description	1	2	3	4	5	6	7
1	INICIO	Initiate run	○	○	○	○	○	○	○
2	ESPECIAL	Measure doping level	○	○	○	○	○	○	○
3	MARC-NAA	Select n-type wafers	○	○	○	○	○	○	○
4	NETG-GEN	Wafer cleaning	○	○	○	○	○	○	○
5	OXINAAA	36 nm dry th. SiO_2	○	○	○	○	○	○	○
6	DNITAAB	117 nm LPCVD Si_3N_4	○	○	○	○	○	○	○
7	NETG-SIM	Wafer cleaning	○	○	○	○	○	○	○
8	FOTO-STU	Phot. TRENCH	○	○	○	○	○	○	○
9	PQ1NIC25	Dry etch Si_3N_4	○	○	○	○	○	○	○

Num.	Code	Description	1	2	3	4	5	6	7
10	QGOXRXXX	Wet etch SiO_2	○	○	○	○	○	○	○
11	P601POLI	Dry etch Si	○	○	○	○	○	○	○
12	DEC-RESI	Strip photoresist	○	○	○	○	○	○	○
13	NETG-SIM	Wafer cleaning	○	○	○	○	○	○	○
14	OHC-XXX	1.6 μm wet SiO_2	○	○	○	○	○	○	○
15	GHUM-ESP	Wet etch oxynit.	○	○	○	○	○	○	○
16	QDNITXXX	Wet etch Si_3N_4	○	○	○	○	○	○	○
17	NETG-SIM	Wafer cleaning	○	○	○	○	○	○	○
18	DBPTEOSX	TEOS deposition	○	○	○	○	○	○	○
19	NETG-SIM	Wafer cleaning	○	○	○	○	○	○	○
20	OFLDFAAD	Fluidification	○	○	○	○	○	○	○
21	MES-NANO	Thick. measurement	○	○	○	○	○	○	○
22	AMRIEING	Dry etch TEOS	○	○	○	○	○	○	○
23	CMOS-MNC	To contaminated							○
24	AFIB-ING	FIB							○
25	FIN-OBL	Finish wafer							○
26	PROT-GEN	Protect components	○	○	○	○	○	○	
27	QDOXTXXX	Wet etch SiO_2	○	○	○	○	○	○	
28	DEC-RESI	Strip photoresist	○	○	○	○	○	○	
29	NETG-SIM	Wafer cleaning	○	○	○	○	○	○	
30	OHC-XXX	300 nm dry th. SiO_2	○	○	○	○	○	○	
31	DNITAAB	117 nm LPCVD Si_3N_4	○	○	○	○	○	○	
32	NETG-SIM	Wafer cleaning		○	○		○	○	
33	FOTO-STU	Phot. WELLS		○	○		○	○	
34	PQ1NIC25	Dry etch Si_3N_4	○	○	○	○	○	○	
35	QGOXRXXX	Wet etch SiO_2	○	○	○	○	○	○	
36	DEC-RESI	Strip photoresist		○	○		○	○	
37	IF&DPC25	Implant. p wells	○	○	○	○	○	○	
38	QDNITXXX	Wet etch Si_3N_4		○	○		○	○	
39	QGOX-XXX	Wet etch SiO_2		○	○		○	○	
40	IF&DNC25	Implant. n well	○	○	○	○	○	○	
41	QPRMEC25	Dip in HF	○	○	○	○	○	○	
42	DNIT-ING	30 nm LPCVD Si_3N_4	○	○	○				
43	DNITXXX	150 nm LPCVD Si_3N_4				○	○	○	
44	FOTO-STU	Phot. GUIDE		○	○		○	○	
45	ESPECIAL	Anneal 200 C, 30 min		○	○		○	○	
46	AMRIEING	Dry etch Si_3N_4		○	○		○	○	
47	DEC-RESI	Strip photoresist		○	○		○	○	
48	FOTO-STU	Phot. ACTIVEIMP		○	○		○	○	
49	ESPECIAL	Anneal 200 C, 30 min		○	○		○	○	
50	AMRIEING	Dry etch Si_3N_4				○	○	○	
51	IMIMCING	Silicon implantation	○	○	○	○	○	○	
52	DEC-RESI	Strip photoresist		○	○		○	○	

Num.	Code	Description	1	2	3	4	5	6	7
53	QDRACXXX	Strip photoresist		○	○		○	○	
54	NETG-GEN	Wafer cleaning	○	○	○	○	○	○	
55	O&R-ESP	Anneal 1100 C, 60 min	○	○	○	○	○	○	
56	DBPTEOSX	TEOS deposition		○	○		○	○	
57	FOTO-STU	Phot. CONTACTWIN		○	○		○	○	
58	ESPECIAL	Anneal 200 C, 30 min		○	○		○	○	
59	AMRIEING	Dry etch TEOS		○	○		○	○	
60	QGOXRXXX	Wet etch TEOS		○	○		○	○	
61	DEC-RESI	Strip photoresist		○	○		○	○	
62	NETG-GEN	Wafer cleaning		○			○		
63	DPOLAAB	350 nm polysilicon		○			○		
64	ODPP0AAA	$POCl_3$ polysilicon doping		○			○		
65	QDPSGC25	Strip PSG		○			○		
66	QPRMEC25	Dip in HF			○			○	
67	MEVAC175	1 μm Al			○			○	
68	FOTO-ALH	Phot. CONTACTS		○	○		○	○	
69	QGALCXXX	Wet etch Al			○			○	
70	AMRIEING	Dry etch polysilicon		○			○		
71	DEC-RESI	Strip photoresist		○	○		○	○	
72	PROT-GEN	Protect components side		○	○		○	○	
73	ESPECIAL	Anneal 200 C, 30 minutes		○	○		○	○	
74	AMRIEING	Dry etch polysilicon		○			○		
75	AMRIEING	Strip layers from back side		○			○		
76	DEC-RESI	Strip photoresist		○	○		○	○	
77	MZ550ING	1 μm Al		○	○		○	○	
78	NETGH2OC	Wafer cleaning		○	○		○	○	
79	ORCALAAA	Al sintering		○	○		○	○	
80	FINAL	End	○	○	○	○	○	○	○

Table C.2: List of steps used in the fabrication of the transceiver (chapter 4.)

Resum en català

Els sistemes fotònics basats en silici han atret molta atenció en les últimes dues dècades, particularment per l'interès de la indústria microelectrònica en resoldre els problemes de coll d'ampolla en les interconnexions entre dispositius individuals que apareixen quan s'assoleixen els alts nivells d'integració amb els que treballa la indústria actualment. Tanmateix, les aplicacions d'una font de llum basada en silici i compatible amb la tecnologia CMOS s'estenen des de les comunicacions fins al sensat i la il·luminació.

Actualment, gairebé tots els dispositius fotònics necessaris per a la fabricació d'un sistema fotònic complet basat en silici, com ara guies d'ona, interruptors i moduladors òptics, i fotodetectors, s'han demostrat amb èxit. Malauradament, encara cal una font de llum eficient basada en silici i compatible amb la tecnologia CMOS. El present treball forma part d'un esforç continuat per assolir una font de llum basada en silici i fabricada amb un procés compatible CMOS, que pugui ser integrada monolíticament en una plataforma de sensat.

Les nanoestructures de silici permeten resoldre la qüestió del *bandgap* indirecte, que és la principal limitació del silici pel que fa a la seva eficiència com a emissor, a través del confinament quàntic. Per tal d'obtenir nanoestructures de silici, el nostre grup ha estat utilitzant un mètode de fabricació basat en l'òxid de silici enriquit en silici (SRO) i el nitrur de silici enriquit en silici (SRN). Després d'un recuit tèrmic d'alta temperatura, les capes de SRO i SRN sofreixen una separació de fases que dóna com a resultat nanoaglomerats de silici pur incrustats a l'interior de la matriu dielèctrica. Altres mètodes per obtenir emissió eficient en silici són la fabricació de silici porós, superxarxes de SiO/SiO_2, que permeten un control molt precís de la mida dels nanocristalls, o punts quàntics col·loïdals de silici.

Pel que fa a la integració, a dia d'avui no s'ha publicat cap transceptor compatible amb la tecnologia CMOS amb tots els components integrats monolíticament en el mateix procés. Tot i així, sí que s'ha aconseguit integrar fonts de llum del grup III-V en plataformes de silici.

En el nostre grup, el treball previ en aquest camp va centrar-se en la fabricació de les capes actives de SRO i SRN, en l'estudi de la relació entre els paràmetres de

fabricació i les propietats estructurals i fotoluminescents de les capes, i en la demostració d'electroluminescència (EL) en dispositius complets que utilitzaven aquestes capes com a material actiu.

L'objectiu del present treball és doble. En primer lloc, un cop les propietats estructurals i òptiques dels materials luminescents han estat estudiades i la EL ha estat demostrada, cal estendre el treball a l'estudi detallat de les propietats òptiques i elèctriques dels dispositius complets. Cal esperar que tal estudi resulti en pistes que ens permetin determinar estratègies per tal de millorar el rendiment dels dispositius, en particular la potència òptica emesa, la seva eficiència i durabilitat, així com millorar el control de l'espectre d'emissió.

El segon objectiu d'aquest treball és proposar un disseny per un transceptor òptic en el qual tots els components (emissor, guia d'ones i detector) estiguin monolíticament integrats en un procés CMOS i que pugui utilitzar-se com a plataforma de sensat, fabricar el disseny proposat i, finalment, caracteritzar-lo.

El primer objectiu es tracta en els capítols 2 i 3 d'aquesta tesi.

El capítol 2 descriu el procés de fabricació que seguim per obtenir les capes actives i els dispositius complets. S'hi presenten els resultats de la caracterització estructural i de la fotoluminescència (PL) de les capes, i es relacionen amb els resultats de la caracterització òptica i elèctrica dels dispositius complets. La caracterització s'ha dut a terme en dispositius experimentals que s'han preparat des de zero específicament per aquest treball.

La comparació de la PL i EL de bicapes SRO/SRN ens ha permès concloure que cada capa contribueix una banda diferent a l'emissió total, cosa que dóna lloc a una distribució de l'energia en un rang més ampli de l'espectre visible. A més a més, disposar de dues bandes d'emissió dóna més flexibilitat a l'hora de controlar l'espectre final obtingut. Tanmateix, calen més estudis per tal de trobar la manera de modular la intensitat de cada banda de forma independent.

Hem determinat que la EL pot presentar-se en tres formes diferents: a través de punts luminescents discrets distribuïts per l'àrea del dispositiu, com a emissió a la vora del dispositiu, o bé homogèniament distribuïda en tota la seva àrea. Aquests tres tipus d'emissió s'han relacionat amb les propietats òptiques i elèctriques dels dispositius.

En els dos primers tipus d'emissió (per punts i a la vora), la EL porta associades densitats de corrent molt altes, de l'ordre de $1 \ \mathrm{A\,cm^{-2}}$, i eficiències de conversió elèctrica-òptica baixes, de l'ordre de 10^{-8}. En canvi, l'emissió homogènia porta associades densitats de corrent més moderades, de l'ordre de $0.01 \ \mathrm{A\,cm^{-2}}$, i eficiències superiors, en el rang $10^{-7} \sim 10^{-5}$. Per tant, concloem que l'emissió homogènia és l'òptima pel que fa a l'eficiència. A més a més, hem proposat un model senzill que permet explicar l'aparició, i ocasionalment coexistència, dels diferents tipus d'emissió. En aquest model, l'emissió homogènia s'associa a una conducció uniforme a través del dispositiu, mentre que els punts lluminosos es produeixen en camins preferents de baixa resistivitat a través de la capa.

També hem explorat l'efecte d'una capa de nitrur de silici sobre la capa activa de SRO, i hem conclòs que és beneficiosa de cara a obtenir una conducció uniforme

que afavoreix l'emissió homogènia en els dispositius.

També hem relacionat els estats de conducció associats als diferents tipus d'emissió amb el comportament de la característica CV dels dispositius. Els resultats indiquen que l'emissió homogènia va associada a una característica CV típica d'una capacitat MIS, al contrari que l'emissió per punts. Això reforça la idea d'una conducció uniforme en un l'emissió homogènia i un dispositiu trencat en l'emissió per punts.

Hem estudiat els mecanismes d'injecció de càrrega en mostres fabricades per deposició química de vapor assistida per plasma (PECVD) i per implantació iònica de ions de silici. Hem conclòs que no hi ha cap mecanisme que pugui donar comptes de la injecció de càrrega en tots els règims de camp elèctric en el rang estudiat. L'efecte túnel Fowler-Nordheim o l'efecte túnel assistit per trampes juguen un paper important en les mostres PECVD. En mostres implantades, l'efecte túnel Fowler-Nordheim sembla dominar a camps baixos, però el mecanisme Poole-Frenkel guanya protagonisme a camps alts.

Durant la caracterització dels dispositius, hem trobat grans diferències entre els espectres de PL i els de EL. Qualitativament, aquestes diferències s'expliquen fàcilment per efectes d'interferències en la multicapa dels dispositius. En el capítol 3 hem presentat un programa d'ordinador basat en el mètode de Crawford que ens permet estudiar quantitativament els efectes de les interferències en l'emissió dels nostres dispositius. Hem pogut concloure que la PL i la EL són iguals malgrat les aparences. El nostre anàlisi també fa palès que és molt important determinar quantitativament els efectes de les interferències en les multicapes si es vol obtenir informació precisa dels processos físics que donen lloc a la luminescència.

Al capítol 4 ens hem ocupat del segon objectiu d'aquest treball. és a dir la fabricació d'un transceptor òptic. N'hem presentat el disseny, el procés de fabricació i la caracterització, i hem superat amb èxit les dues principals dificultats que comporta la integració monolítica d'emissor, guia d'ones i detector en un procés CMOS, que són l'obtenció del detector, la fabricació del qual està sotmesa als requeriments de càrrega tèrmica en la fabricació de l'emissor, i l'obtenció d'una trinxera d'òxid que sigui raonablement plana i uniforme.

Malauradament, el transceptor no ha funcionat correctament i no hem pogut mesurar acoblament entre emissor i detector. El principal culpable n'és l'emissor de SRN, que presenta una luminescència molt pobre. Cal fer més feina per controlar millor l'emissió de la capa de SRN al transceptor. Tanmateix, creiem que el disseny bàsic del transceptor és correcte, ja que l'acoblament elèctric entre l'emissor i el detector és molt baix.

Bibliography

[1] Stefano Ossicini, Lorenzo Pavesi, and Francesco Priolo. *Light Emitting Silicon for Microphotonics.* Springer-Verlag, 2003.

[2] Lorenzo Pavesi and David J. Lockwood, editors. *Silicon Photonics*, volume 94. Springer-Verlag, 2004.

[3] David J. Lockwood, editor. *Light Emission in Silicon, from Physics to Devices.* Academic Press, 1998.

[4] L. A. Nesbit. Annealing characteristics of Si-rich SiO_2 films. *Applied Physics Letters*, 46(1):38–40, 1985.

[5] L. Dal Negro, J. H. Yi, L. C. Kimerling, S. Hamel, A. Williamson, and Giulia Galli. Light emission from silicon-rich nitride nanostructures. *Applied Physics Letters*, 88(18):183103, 2006.

[6] P. Nguyen. Direct observation of quantum confinement of Si nanocrystals in Si-rich nitrides. *Physical Review B*, 85(8), 2012.

[7] O. Bisi, Stefano Ossicini, and L. Pavesi. Porous silicon: a quantum sponge structure for silicon based optoelectronics. *Surface science reports*, 38(1-3):1–126, April 2000.

[8] Margit Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Bläsing. Size-controlled highly luminescent silicon nanocrystals: A SiO/SiO2 superlattice approach. *Applied Physics Letters*, 80(4):661, 2002.

[9] Chang-Ching Tu, Liang Tang, Jiangdong Huang, Apostolos Voutsas, and Lih Y. Lin. Visible electroluminescence from hybrid colloidal silicon quantum dot-organic light-emitting diodes. *Applied Physics Letters*, 98(21):213102, 2011.

[10] Peter De Dobbelaere, Sherif Abdalla, Steffen Gloeckner, Michael Mack, Gianlorenzo Masini, Attila Mekis, Thierry Pinguet, Subal Sahni, Adithyaram

Narasimha, Drew Guckenberger, and Mark Harrison. Si Photonics Based High-Speed optical Transceivers. *European Conference and Exhibition on Optical Communication*, (1):We.1.E.5, 2012.

[11] Hyundai Park, MN Sysak, and HW Chen. Device and integration technology for silicon photonic transmitters. *Selected Topics in . . .* , 17(3):671–688, May 2011.

[12] Dries Van Thourhout, Thijs Spuesens, Shankar Kumar Selvaraja, Liu Liu, Günther Roelkens, Rajesh Kumar, Geert Morthier, Pedro Rojo-Romeo, Fabien Mandorlo, Philippe Regreny, Oded Raz, Christophe Kopp, and Laurent Grenouillet. Nanophotonic Devices for Optical Interconnect. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(5):1363–1375, September 2010.

[13] Constantin Bulucea. Recalculation of Irvin's resistivity curves for diffused layers in silicon using updated bulk resistivity data. *Solid-State Electronics*, 36(4):489–493, April 1993.

[14] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 1, pages 5–8. CRC Press, third edition, 2012.

[15] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 1, chapter 4, pages 237–245. CRC Press, third edition, 2012.

[16] Takashi Hori. *Gate dielectrics and MOS ULSIs*, volume 34 of *Springer series in electronics and photonics*, chapter 4, pages 149–156. Springer, 1997.

[17] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 7, pages 414–438. CRC Press, third edition.

[18] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 1, pages 15–17. CRC Press, third edition, 2012.

[19] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 3, pages 155–185. CRC Press, third edition, 2012.

[20] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 4, pages 215–318. CRC Press, third edition, 2012.

[21] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 1, chapter 4, pages 233–236. CRC Press, third edition, 2012.

[22] Gueorgui Horhe-Karlos Barreto. *Silicon Nanoclusters-based light sources*. PhD thesis, Universitat Autònoma de Barcelona, 2009.

[23] Mariano José Perálvarez Barrera. *Si nanocrystal MOS-LEDs: strategies to improve efficiency and device lifetime*. PhD thesis, Universitat de Barcelona, 2009.

[24] J Ziegler. SRIM-2003. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 219-220:1027–1036, 2004.

[25] www.synopsys.com.

[26] Max Born and Emil Wolf. *Principles of optics, electromagnetic theory of propagation, interference and diffraction of light.* Cambridge University Press, seventh edition, 1999.

[27] Ilsin An, Hans Arwin, Chi Chen, Robert W. Collins, Andre S. Ferlauto, James N. Hilfiker, Josef Humlícek, Eugene A. Irene, Jr. Gerald E. Jellison, Joungchel Lee, Frank A. Modine, Arnulf Röseler, Mathias Schubert, Harland G. Tompkins, and Juan A. Zapien. *Handbook of ellipsometry.* Springer, William Andrew Publishing, 2005.

[28] R. M. A. Azzam and N. M. Bashara. *Ellipsometry and polarized light.* North-Holland Publishing Company, 1977.

[29] John R. Taylor. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. August 1996.

[30] H. R. Verma. X-ray photoelectron spectroscopy. In *Atomic and nuclear analytical methods*, chapter 5, pages 213–241. Springer, 2007.

[31] Alfredo Morales-Sánchez, Jorge Barreto, Carlos Domínguez, Mariano Aceves-Mijares, and J. A. Luna-López. Optical characterization of silicon rich oxide films. *Sensors and Actuators A: Physical*, 142(1):12–18, March 2008.

[32] Fabio Iacona, G. Franzò, and Corrado Spinella. Correlation between luminescence and structural properties of Si nanocrystals. *Journal of Applied Physics*, 87(3):1295, 2000.

[33] Gueorgui Horhe-Karlos Barreto, José A. Rodríguez, M. Perálvarez, Alfredo Morales-Sánchez, Blas Garrido, and Carlos Domínguez. Photoluminescence characterization of silicon nanostructures embedded in silicon oxide. *Superlattices and Microstructures*, 43(5-6):588–593, May 2008.

[34] Edward S. Yang. *Microelectronic devices*, chapter 7, pages 192–196. McGraw-Hill series in electrical engineering. McGraw-Hill, 1988.

[35] S. M. Sze and Kwok K. Ng. *Physics of semiconductor devices*, chapter 3, pages 187–191. John Wiley & Sons, third edition, 2007.

[36] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 1, chapter 4, pages 229–233. CRC Press, third edition, 2012.

[37] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 1, pages 3–87. CRC Press, third edition, 2012.

[38] Marc J. Madou. *Fundamentals of microfabrication and nanotechnology*, volume 2, chapter 7, pages 398–405. CRC Press, third edition, 2012.

[39] D. Izquierdo, María C Garralaga, Íñigo Salinas, Jorge Barreto, and Carlos Domínguez. Silicon nanocrystals light emitting devices: characterization and coupling to SU-8 waveguides. In *SPIE Photonics Europe 2010*, page 7719M, Brussels, 2010.

[40] S. Lombardo, A. La Magna, Corrado Spinella, C. Gerardi, and F. Crupi. Degradation and hard breakdown transient of thin gate oxides in metal-SiO2-Si capacitors: Dependence on oxide thickness. *Journal of Applied Physics*, 86(11):6382, 1999.

[41] Adnan Mehonic, Sébastien Cueff, Maciej Wojdak, Stephen Hudziak, Olivier Jambois, Christophe Labbé, Blas Garrido, Richard Rizk, and Anthony J. Kenyon. Resistive switching in silicon suboxide films. *Journal of Applied Physics*, 111(7):074507, 2012.

[42] M Meretoja, R Punkkinen, H-P Hedman, L Heikkilä, M Hirvonen, T Suominen, and M Heinonen. Electroluminescence in Si/SiO2 Microstructures. *Physica Scripta*, T114:53–56, 2004.

[43] Alfredo Morales-Sánchez, Jorge Barreto, Carlos Domínguez, Mariano Aceves-Mijares, and J. A. Luna-López. The mechanism of electrical annihilation of conductive paths and charge trapping in silicon-rich oxides. *Nanotechnology*, 20(4):45201, January 2009.

[44] K F Schuegraf and Chenming Hu. Reliability of thin SiO2. *Semiconductor Science and Technology*, 9(5):989–1004, May 1994.

[45] Jorge Barreto, Alfredo Morales-Sánchez, M. Perálvarez, Blas Garrido, and Carlos Domínguez. Stoichiometry of silicon-rich dielectrics for silicon nanocluster formation. *Physica Status Solidi (C)*, 8(3):804–807, March 2011.

[46] Chih-Kuo Tseng, Ming-Chang M. Lee, Hau-Wei Hung, Jiun-Ru Huang, Kun-Yu Lee, Jia-Min Shieh, and Gong-Ru Lin. Silicon-nanocrystal resonant-cavity light emitting devices for color tailoring. *Journal of Applied Physics*, 111(7):074512, April 2012.

[47] Takashi Hori. *Gate dielectrics and MOS ULSIs*, volume 34 of *Springer series in electronics and photonics*, page 45. Springer, 1997.

[48] S. M. Sze and Kwok K. Ng. *Physics of semiconductor devices*, page 227. John Wiley & Sons, third edition, 2007.

[49] Rohana Perera, Akihiro Ikeda, Reiji Hattori, and Yukinori Kuroki. Trap assisted leakage current conduction in thin silicon oxynitride films grown by rapid

thermal oxidation combined microwave excited plasma nitridation. *Microelectronic Engineering*, 65(4):357–370, May 2003.

[50] J G Simmons. Conduction in thin dielectric films. *Journal of Physics D: Applied Physics*, 4(5):613–657, May 1971.

[51] E. H. Nicollian and J. R. Brews. *MOS Physics and Technology*. Wiley-Interscience, New York, 1982.

[52] D. K. Schroder. *Semiconductor Material and Device Characterization*. Wiley-IEEE Press, New Jersey, third edition, 2006.

[53] Konstantinos Misiakos and Dimitris Tsamakis. Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K. *Journal of Applied Physics*, 74(May 2010):3293, 1993.

[54] W. E. Beadle, J. C. C. Tsai, and R. D. Plummer. *Quick Reference Manual for Silicon Integrated Circuit Technology*. Wiley-Interscience, New York, 1985.

[55] S. M. Sze and Kwok K. Ng. *Physics of semiconductor devices*. John Wiley & Sons, third edition, 2007.

[56] M. Lenzlinger and E. H. Snow. Fowler-Nordheim tunneling into thermally grown Si02. *Journal of Applied Physics*, 40(1):278, 1969.

[57] Ravi Kumar Chanana. Determination of hole effective mass in SiO2 and SiC conduction band offset using Fowler-Nordheim tunneling characteristics across metal-oxide-semiconductor structures after applying oxide field corrections. *Journal of Applied Physics*, 109(10):104508, 2011.

[58] K. J. Hebert and E. A. Irene. Fowler-Nordheim tunneling current oscillations at metal/oxide/Si interfaces. *Journal of Applied Physics*, 82(1):291, 1997.

[59] Alessia Irrera, Fabio Iacona, I. Crupi, Calogero D. Presti, G. Franzò, Corrado Bongiorno, Delfo Sanfilippo, Gianfranco Di Stefano, Angelo Piana, P. G. Fallica, Andrea Canino, and Francesco Priolo. Electroluminescence and transport properties in amorphous silicon nanostructures. *Nanotechnology*, 17(5):1428–1436, March 2006.

[60] O. H. Crawford. Radiation from oscillating dipoles embedded in a layered system. *The Journal of Chemical Physics*, 89(10):6017–6027, 1988.

[61] Max Born and Emil Wolf. *Principles of optics, electromagnetic theory of propagation, interference and diffraction of light*. Cambridge University Press, seventh edition, 1999.

[62] www.python.org.

[63] www.scipy.org.

[64] code.google.com/p/numexpr.

[65] www.cyton.org.

[66] Carlos Domínguez, José A. Rodríguez, Francisco J. Muñoz, and Nadia Zine. Plasma enhanced CVD silicon oxide films for integrated optic applications. *Vacuum*, 52(4):395–400, April 1999.

[67] Nicholas J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, 2nd edition, 2002.

[68] Jon Orloff, Lynwood Swanson, and Mark Utlaut. *High resolution focused ion beams: FIB and its applications*. Springer, 2003.

[69] Ray F. Egerton. *Physical principles of electron microscopy: an introduction to TEM, SEM, and AEM*. Springer, 2005.

[70] M.S. Tyagi and R. Van Overstraeten. Minority carrier recombination in heavily-doped silicon. *Solid-State Electronics*, 26(6):577–597, 1983.

[71] K. Bücher, J. Bruns, and H. G. Wagemann. Absorption coefficient of silicon: An assessment of measurements and the simulation of temperature variation. *Journal of Applied Physics*, 75(2):1127, 1994.

[72] www.scipy.org.

[73] github.com.

[74] git-scm.com.